EyeCU - Assistive Eyewear

Team 36 - Nikhil Mehta, Irfan Suzali, Abishek Venkit
ECE445 Project Proposal - Spring 2020
TA: Shuai Tang

1 Introduction

1.1 Objective

Many of us take full vision for granted in our everyday lives, and enjoy the benefits of being able to read, identify objects, and navigate around the world without concern. However, it's not that simple for the 7 million people in the US [1] , and millions more around the world, that live with visual impairments. Although many visually impaired people live perfectly functional, independent lives, there are still some situations that are not fully accessible. For our project, we will be focusing on the specific problem of reading text. The lack of ability to read text is an extremely common problem for people with high visual impairment or blindness, and can prevent a person from being fully independent. For example, if a person with visual impairments goes to a grocery store to purchase milk, they will be unable to identify which milk carton contains whole milk, 2%, or soy without assistance from another person, or an app on your smartphone. This is just one of many examples in which the ability to read text quickly and seamlessly would be extremely beneficial, and we envision a better way to accomplish this task.

Our solution is an eyewear device that has the ability to capture an image using a built-in camera, identify and read text from the image, and finally convert that text to speech and play it in your ear. This system will be wearable, always ready at the push of a button, and a user-friendly way to assist with reading text. Although the scope of our project is confined to reading text for now, the potential is much greater, and this system can be extended to run many other computer vision detection tasks. However, even in its most simple state, we believe that the value will still be substantial.

1.2 Background

Partial inspiration for this project came from an existing app called "Be My Eyes" [2]. This app already serves almost 200,000 people with visual impairments, by connecting them with a sighted volunteer over a video call. This allows the person with low vision to ask questions and receive feedback from a sighted volunteer. Although this system is simple and gained some attention, it doesn't accomplish the goal of full integration and independence because it relies on a network of sighted volunteers for the service to function. There may also be a delay between requesting a volunteer and receiving one, and requires you to navigate to the app before getting assistance. We can improve upon this by providing the same benefits, but in a ready-to-use, simple package, and without the need for a volunteer network.

To accomplish our text recognition task, we require a computer vision model to recognize and identify words in an image, which is known as Optical Character Recognition, or OCR. We're confident in our ability to do this because this has been a huge area of development for the past several decades, and there are many open source resources that can accomplish this at a very high accuracy rate, including OpenCV Tesseract [9], Google OCR [7], and Kraken OCR. These are well-established and documented, and will help us accomplish our task effectively.

Next, in order to dictate the text to the visually impaired user, we must convert that text string into a speech signal using a Text-to-Speech system, or TTS. TTS is another highly developed area, commonly used in voice assistants and other text dictation tools.

1.3 High-Level Requirements

- The speech produced by the device must have a word accuracy of above 90% on clear, unobstructed text, and above 75% for surface distances up to a meter away, on a variety of surface textures, text styles, and text clarity.
- The glasses must weigh between 30-70 grams, and be comfortable enough for regular wear.
- The device should be able to identify individual characters at a confidence level of 95% or higher for clear, unobstructed text.
- The device must be reliable (powered and functional) for 30+ translations (2.5 uses per hour for a 12 hour period).

2 Design

For design purposes, our block diagram has been designed with modularity in mind so that testing of different modules could be done independent of other components. We have identified 4 main modules essential for our project to function as proposed. The control unit will control the I/O of the project, sending the correct signals to modules when needed. There are two peripheral modules, the camera module and speaker module, which will handle the visual and audio components of the project. Within the camera module, the Raspberry Pi is responsible for the computer vision task. The power supply module will power all hardware.

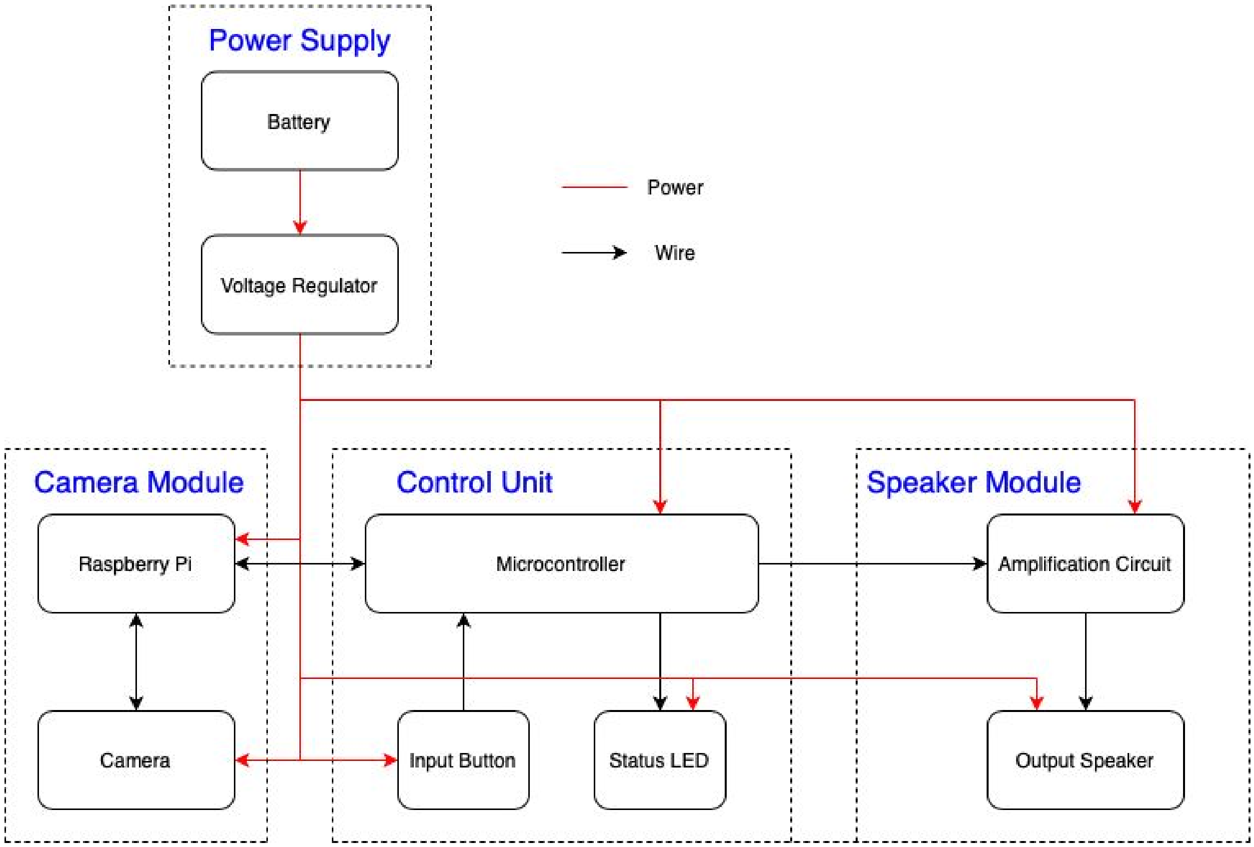
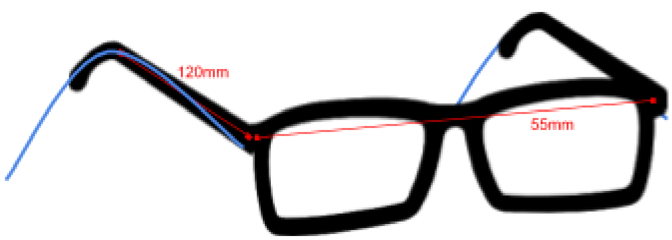


Figure 1. Block Diagram

2.1 Physical Design

The physical design is comprised of two main components, the wearable sunglasses and pocket module. The sunglasses will have the camera and PCB board built into the sunglasses, one on each side to distribute weight for stability reasons. Two wires (which will run down the sides of the glasses) will connect the pcb and the camera to the pocket module where the power supply and image processing component is held.



(Wearable Glasses) (Pocket Module)

Figure 2. Physical Design

2.2 Camera Module

The camera module will allow the device to take images and process them to generate text. The camera will take images and transfer this data to the Raspberry Pi. The CPU on the Raspberry Pi will do the image processing, and send a text string out to the control unit.

2.2.1 Raspberry Pi

The device will use a Raspberry Pi 4 for image processing using computer vision (openCV) [5]. The Pi will take image input directly from the camera, process this image on the CPU, and send data out. *No other portion of the Raspberry Pi will be used.*
Requirement: Must read the image data, process it, and output string text in under 2 minutes (for demo purposes).
Requirement: Must be able to read pixel data over CSI interface.
Requirement: Must be reliable when powered by 5V, 2.4A battery, and have a max current draw of 2A.

2.2.2. Camera

The camera used will be a Raspberry Pi compatible camera (5 MP, 1080p resolution). The camera must respond to a button click from the control unit, and take an image of the scene in front of the user with minimal delay [8].
Requirement: Must take clear, 360p-1080p resolution images (will downsample to improve speed) immediately upon button press.
Requirement: Must communicate over CSI interface and CSI bus, allowing for fast pixel data transfer.

2.3 Power Supply

The power supply is required to allow usage of all components in the device. This includes the Raspberry Pi, the microcontroller, the camera, the LED, and the speaker.

2.3.1 Battery

The battery will be a rechargeable, usb charger (compatible with smartphones and other devices). It must provide sustained power to all components for a 12 hour usage period.
Requirement: Must supply 5V at 2.4A (max) for 30 translations (2.5 uses per hour for a 12 hour period) to all components of the device.
Requirement: Temperature must be below 45C in order to maintain safety of the user and device circuitry.

2.3.2 Voltage Regulator

This circuit must supply 3.3V to the microcontroller and pcb and 5V to Raspberry Pi, and must handle 5V peak input from the battery and 2A peak current draw from the Raspberry Pi.
Requirement: Must provide 3.3V +/- %5 from the 5V battery pack.
Requirement: Temperature must be below 125C for stability and safety.

2.4 Control Unit

The control unit needs to process string data coming from the Raspberry Pi, handle button presses from the user, control the status LED, and output audio to the speaker module. The button presses will be processed by the microcontroller, which needs to communicate with the Raspberry Pi to take a picture and receive string data after computation. Simultaneously, the status LED will be activated. Finally, the text is converted to speech and played by the speaker.

2.4.1 Microcontroller

The microcontroller will be a PIC32, and will communicate with the Raspberry Pi using UART. It will run the Software Automatic Mouth (SAM) Text-to-Speech (TTS) program, and interface with the status LED and speaker [4].

Requirements: Must communicate over UART at speeds of 4.5Mbps
Requirements: Must use 3.3V +/- 5% for power, and draw less than 2A current.

2.4.2 Input Button

The input button will be a standard push button, and will signal the start of the picture → speech action.
Requirements: Must be easy to press, and not require too much force (which will destabilize the glasses)
Requirements: Must be reliable (requires a single press)

2.4.3 Status LED

The status LED will indicate when the computation process is occuring by flashing on and off. This will be used for debugging purposes, and to notify bystanders of an image capture.
Requirements: Must be visible from 2-3 meters away.
Requirements: Must draw 10mA current.

2.5 Speaker Module

The speaker module will amplify the audio signal sent from the microcontroller, and output this signal on a single tweeter. (This implementation is ideal for demo purposes, but a refined product may replace the tweeter with an earpiece).

2.5.1 Amplification Circuit

The amplification circuit will include an voltage offset circuit (to supply +/- 2.5V to the opAmp) and a opAmp to amplify the signal.
Requirements: Must amplify audio signal enough so the user can hear it clearly.
Requirements: Must remain below 45C in order to maintain safety of the user and device circuitry.

2.5.2 Output Speaker

The output speaker will be a single, small, lightweight tweeter. This will allow the user to clearly hear the audio, while also allowing others to evaluate the device functionality.
Requirements: Must be below 70 dBA (a-weighted decibels) for user safety.
Requirements: Must be lightweight, less than 10 grams, to maintain wearability and comfort

2.6 Risk Analysis

The success of the project is determined mainly by our ability to convert images to text and text to speech. Thus, it is important that our camera module be able to efficiently and successfully detect text within an image with a certain degree of accuracy. Once processed, we would then need to process the string of text to an audio form that could be outputted with enough clarity for the user to understand. Since our project is centered around the user, we must ensure that the results achieved meet a certain benchmark to ensure usability. Doing so will be the greatest challenge of the project.

3 Ethics and Safety

We have an obligation to our profession to uphold the highest level of ethical and professional conduct. We stand to follow and commit ourselves to the guidelines stated by the IEEE Code of Ethics. Therefore, it is our duty to outline the ethical and safety concerns that may arise with our proposed project.

We acknowledge that there is a certain degree of error that can arise from recognizing text from images given different situations. The core of the project depends on users being able to trust our system to identify text with a certain degree of accuracy. According to #3 of the IEEE Code of Ethics [6], it is our duty to be honest of the estimates provided from the available data provided to us. To uphold this, we plan on allowing only text identified to a certain degree of certainty to be outputted through the system. Users should be able to trust our system in giving valuable and actionable insight so that they may make decisions accordingly.

Safety of the user is of utmost importance especially since there are significant hardware components situated on the body of the user. #1 of the IEEE Code of Ethics [6] outlines the importance of safety and health of users and the public. Components of our system have a possibility of overheating and exploding.

Our main concerns for safety stem from the external battery and a possibility of a short circuit. The external battery runs the risk of overheating due to an excessive current or voltage passing through it [3]. We intend to design our PCB with overcurrent and overvoltage precautionary by implementing a resettable fuse and zener diode. Human error during the design and building of the circuit is unavoidable. It is possible that whilst designing the PCB board, a reverse polarity may have been introduced. We plan to introduce diodes to ensure the proper setting of polarities. To avoid a short circuit from external factors such as exposure to liquids, we can encase electronics in durable hardware components.

References

[1] "Blindness Statistics," *National Federation of the Blind*. [Online]. Available: https://www.nfb.org/resources/blindness-statistics. [Accessed: 13-Feb-2020].

[2] "Bringing sight to blind and low-vision people," *Be My Eyes*. [Online]. Available: https://www.bemyeyes.com/. [Accessed: 13-Feb-2020].

[3] "DC Power Supply Protection from Unexpected Damage and Inconvenience," *Altium*, 01-May-2018. [Online]. Available: https://resources.altium.com/pcb-design-blog/dc-power-supply-protection-from-unexpected-damage-and-inconvenience#. [Accessed: 13-Feb-2020].

[4] "English text to speech on a PIC microcontroller," *ToughDev*, 05-Jul-2019. [Online]. Available: http://www.toughdev.com/content/2014/09/english-text-to-speech-on-a-pic-microcontroller/. [Accessed: 13-Feb-2020].

[5] L. Hattersley, "Raspberry Pi 4 vs Raspberry Pi 3B ," *The MagPi magazine*, 13-Nov-2019. [Online]. Available: https://magpi.raspberrypi.org/articles/raspberry-pi-4-vs-raspberry-pi-3b-plus. [Accessed: 13-Feb-2020].

[6] "IEEE Code of Ethics," *IEEE*. [Online]. Available: https://www.ieee.org/about/corporate/governance/p7-8.html. [Accessed: 13-Feb-2020].

[7] S. Panigrahi, "Google's Optical Character Recognition (OCR) software works for 248 languages," *Opensource.com*. [Online]. Available: https://opensource.com/life/15/9/open-source-extract-text-images. [Accessed: 13-Feb-2020].

[8] "Raspberry Pi FAQs," *Raspberry Pi Documentation*. [Online]. Available: https://www.raspberrypi.org/documentation/faqs/. [Accessed: 13-Feb-2020].

[9] A. Rosebrock, "OpenCV OCR and text recognition with Tesseract," *Pyimagesearch*. .