Capstone Proposal
SignaSpek

Alexander Mark ID: 400207526
Gurkarn Johal ID: 400170680
Rana Raaiq Rauf ID: 400240123
Jaimik Patel ID: 400176106
Parth Mistry ID: 400193090

Table of Contents

Executive Summary

This project aims to build a feasible way of sound awareness for people with hearing impairment. The current solution for people with severe hearing impairment involves using service dogs to identify essential sounds and notify the owner. Well trained dogs cost anywhere from 15,000 to 25,000 dollars and approximately 1000 a year dollars to maintain. Our solution aims to provide an affordable alternative to expensive service dogs. This prototype is divided into 3 main deliverables: Bronze, Silver, and Gold. Bronze allows audio processing to convert speech into real-time captions printed on an oled screen placed in front of the eye which will be the foundation of this project. Additional features involve calculating audio source directions and displaying them on the screen using arrows and transcription language translation for Silver and Gold levels respectively. SignaSpek will be an affordable product to aid hard of hearing clients and ideally replace expensive service dogs. This will be a low cost prototype costing approximately $200 to build within a 6 month timeframe starting with the Bronze deliverable.

Background and Motivation

Individuals with hearing disabilities have a hard time hearing and noticing alarms, doorbells, and other important prompts in the individuals lives. With service dogs rising in prices from 20 to 25 thousand dollars and years of service from 8 to 10 years,[1] the use of an alternative can provide options for individuals in harsh financial situations. Speech to Text on lens can provide a quick and easy solution for hearing impaired individuals, by giving real time information of the noises they hear around them whether that be speech or sounds?
Other products that provide aid to the hearing impaired include three distinct solutions. Assistive listening devices (ALDs), which help amplify the sounds a user may want to hear, especially where there's a lot of background noise. Augmentative and alternative communication (AAC) devices which help users who may have communication disorders and have a hard time expressing themselves. These devices could be a simple picture board or a computer program that synthesizes speech from text. Lastly, Alerting devices which can connect to a doorbell or telephone, and inform the user with loud sounds, lights and motion to alert them of an event taking place. Our solution gives the user the information they need by combining these three options and turning it into a better product.

A similar project done by others include the XRAI Glass. As described by their website, "XRAI Glass uses off-the-shelf augmented reality 'AR' glasses that are tethered to a smartphone app with AI driven software that gives the user a new kind of personal assistant, one that remembers what you have forgotten." This product gives text to speech for the user similar to our product but with a reliance on AI and your phone.
The Apple Vision Pro, google glasses, microsoft vision, are also other somewhat similar products which can provide real time information on glasses as you go about your day, they all don't provide speech to text and don't give alerts to the sounds around you but they have a similar idea.

Objectives

Our objective is to create a lightweight pair of glasses able to recognize speech from various languages and transcribe the speech to text in real time to a transparent display located on the lens. This product can be used by people who are hard of hearing or deaf allowing real time speech to text translation or by people traveling abroad who want to understand what is being said to them. There will be additional features for people who are hard of hearing such as sound recognition of key noises for example fire alarms, doorbells etc. which are then displayed as well as direction indicators to determine the direction of the sound source.

Step-by-Step Technical Approach

The technical approach is split up into 3 sections, each corresponding to its category of deliverables. The individual steps to approach this problem are listed below:

Bronze Level

1. Research into what ML & Python libraries are needed to transcribe audio to text
2. Research hardware needed to fulfill project demands (i.e microcontroller, batteries, oled screen,etc)
   a. Look into if OLED screens are suitable to use as display close to the eye, look into image projection if not
   b. Make sure microcontroller can handle application load
   c. Have rough design of glass frame to hold all required equipment necessary, and use design constraint to choose the appropriate hardware
3. Order parts
4. Develop a conceptual working demo of speech to text.
5. Test application for bugs:
   a. Test for accuracy in transcribing speech. Look for better alternatives if accuracy is too low
   b. Test for background noise filtering
6. Develop application to display/project text to screen
7. Unit application that displays text to screen
8. Once parts have arrived, port application onto microcontroller and test for any functional bugs
9. Fabricate glass frame and assemble device

Silver Level

10. Research into existing libraries for detecting different sounds
    a. If no library exist, then research into how different sounds can be distinguished
11. Develop and test a simple program that can distinguish a simple alarm bell from noise.
    a. Expand to other types of sounds if program runs well

12. Develop program to display different images onto glasses depending on sound source
13. Extend program to deduce the direction of sounds using multiple microphones
14. Develop and test final source detection program integrating onto the device
15. Determine what languages to additionally support to as input
16. Research for existing transcribing libraries that support multilingual input, otherwise figure out way to extend existing program to accept multiple languages

Gold Level

17. Research into existing libraries that convert text to sign language
    a. If not, look into how text can be converted to sign language
18. Develop and/or integrate sign language library into device

Resources

• Expected Materials:Raspberry pi zero w, 2 Microphones, Glasses, Oled screen,Haptic feedback, Circuitry Components (wires, breadboards, small components)
• Knowledge and skills available: Expertise in python programming (possibly C), various python libraries to transcribe speech, process audio, etc
• Bread boards and other circuitry available
• Time: ~7 months
• Manpower: 5
• 3D printers and microcontrollers available in Thode Makerspace
• Team member has 3D printer available
• Instructors:Dr. Shahram Shirani, ITB-A225 ext. 27943
• TAs' are online 2:30 to 5:30 , Monday to Friday

Schedule of activities and milestones

September	October	November	December
Weekly Meetings (3-hours each)	Finish research for all software, hardware components	Build program to transcribe audio	Build and assemble prototype glasses.Port transcribe program onto glasses

January	February	March	April
Complete Bronze level prototype and start researching for Silver level	Implementing Language translation and sound detection	Implement gold level features	Completion and Capstone presentation

Risks and alternative plan

There are several risks involved with this product. Designing effective glasses involves being able to see through the lenses, wear them comfortably, and see the text on the screen. If any of these attributes are not satisfied then the glasses cannot be used as we intend them to be used. There is a risk that the transparent display is too opaque and blocks vision for the user, an alternative could be only placing the display over a single eye and making it movable. If the glasses are too bulky or if there is not enough space for hardware, they will be uncomfortable to wear and an alternative could be changing the form to goggles. Goggles go around the entire head and can allow for more space to add hardware. Another alternative is to offload processing, perhaps through the use of wifi. The Final risk is the resolution of the display. The resolution may not be ideal at the distance the lens will be from the eyes on glasses. An alternative could be projecting the text rather than using an LED display or finding a way to move it further while still being a part of the glasses.

Deliverables

The Deliverable outcome of this project is split into 3 tiers: Bronze, Silver and Gold. The minimum tier achievement to consider this project complete will be Bronze. Each level after that will involve major upgrades to the product to expand customer base and improve the problem solution. The features for each level of deliverable is outlined below:

1. Bronze Tier:
 I. Glasses with an OLED screen extension for visual text display
 II. Realtime transcription of audio onto the OLED screen done with a microprocessor and ML libraries
 III. Glasses, microprocessor, OLED screen and peripherals mounted safe and securely to head piece

2. Silver Tier:
 I. Includes all features of Bronze Tier
 II. Specific sound detection of key sounds such as doorbells, sirens, and alarms
 III. Arrows will be displayed on the OLED screen to direct the customer to the source of the noise
 IV. The audio transcription will now feature translation from different languages to English

3. Gold Tier:
 I. Have a display of sign language with transcribing audio
 II. Have a multi language selection feature
 III. Have other on screen display of weather, time, and phone notifications

References

Research Based:
Hud on glasses projects- https://forums.adafruit.com/viewtopic.php?t=21734
Possible design alternative - https://www.digitaltrends.com/cars/head-up-display-for-any-helmet/
End to end voice recognition - https://www.youtube.com/watch?v=X12N2Rn-q5o
Offline real time audio - https://www.youtube.com/watch?v=JMQV6teC4KQ
Speech to text docs - https://picovoice.ai/docs/cheetah/
Service Dog Prices - https://www.goamplify.com/blog/moneymanagement/service-dog-costs/

Product Links:
LCD transparent screen -
https://www.aliexpress.com/item/1005003981451741.html?src=google&aff_fcid=e11d79d3f6be4bdd9c21c140265324b2-1697147731337-00738-UneMJZVf&aff_fsk=UneMJZVf&aff_platform=aaf&sk=UneMJZVf&aff_trace_key=e11d79d3f6be4bdd9c21c140265324b2-1697147731337-00738-UneMJZVf&terminal_id=47b769a58bc5419997c2795453fe0637&afSmartRedirect=y

Citation

[1]"Service Dogs For PTSD and other Mental Health issues | Kingston 4 Paws Service Dogs."
https://kingstonservicedogs.ca/service-dogs-for-ptsd-and-other-mental-health-issues/#:~:text=Costs (accessed Oct. 16, 2023).
[2] A. Khoury. "Mount this head-up display to your helmet to project any phone app".
Digitaltrends. https://www.digitaltrends.com/cars/head-up-display-for-any-helmet/