Smart Glass Using IoT and Machine Learning Technologies to aid the blind, dumb and deaf.

Siddhant Salvi¹, Sudipta Pahar², Yuvraj Kadale³

¹Bachelors of Engineering in Electronics and Telecommunication, D.J. Sanghvi College of engineering, Mumbai, India.
²Bachelors of engineering in Electronics and Telecommunication, KIIT DU Bhubhaneshwar, Odisha, India.
³Bachelors of engineering in Electronics and Telecommunication, NIT Hamirpur, Himachal Pradesh, India.

siddhantsalvi99@gmail.com

Abstract: One of the earliest written records of a sign language is from the fifth century BC, in Plato's Cratylus, where Socrates says: "If we hadn't a voice or a tongue, and wanted to express things to one another, wouldn't we try to make signs by moving our hands, head, and the rest of our body, just as dumb people do at present?"[Source: Wikipedia]. Technology is far more developed (advanced) than time then. Now, we can detect what the deaf and mute person is saying just by recording the same and comparing it with predefined datasets, revealing what the individual is trying to accomplish. In this paper, we have proposed a solution by which we can help disabled people, people who are Deaf, Dumb, or Blind by providing them a new technological eye, ear, and brain. Machine learning algorithms are used for object detection to provide the eye for the blind. Speech to text conversion to aid the deaf and text to speech conversion to aid the dumb while communicating. The amalgamation of all these technologies with IOT technologies will help resolve the issues related to these differently abled people.

Keywords—Machine-learning; image processing; radio communication; internet of things; embedded; beagle bone; beagle bone black; Bluetooth communication; IP communication; ADC; DAC; CNN; ASR; API; AR; Augmented Reality; VR; Virtual Reality; YOLO; low cost; smart glass; AR glasses; Mobile Computing; Cellular Networks; Network Management.

1. Introduction

We approached the problem using some pre-defined algorithms and modified those according to our needs. The design implementation was done according to the hardware requirements and many there are many hardware and software requirements apply to be used in this paper.

The list of the hardware requirements is:

Transparent OLED panel: T-Oled(Transparent O-Led panel) for the help of deaf and dumb individuals for viewing the speech-to-text conversion displayed over the HUD of the glasses.
Normal OLED: Instead of T-Oled panel, we can use 128x32 normal Oled panel and a mirror to reflect the text coming on the Oled screen directly to the user’s glass in front of the eye so that the user doesn’t have to install another hardware dangling in front of the eye.

IP Cameras (2): IP Cameras are used because we aren’t using any server storage in the glass itself and so we need to send the raw footage to another mobile device, for its processing and then converting it to our output requirement.

Proximity sensors: This is used for impact-resistant technology and if any obstacle comes within the available range, it will warn the blind individual about it and the individual needs to check his route.

Ultrasonic sensors: This sensor will be used for calculating the distance to any incoming obstacle or vehicle. Since proximity sensors can’t be used for larger distances, the ultrasonic sensors will control the situation from there.

Microphone array: This device will be essential for the Deaf and Blind individuals. It will receive the surround sound from the place of existing of the individual and process it according to the input sound, like to detect any incoming vehicle from the rear and the front camera for some predefined tasks.

Speaker: To output the text to speech near the ear of the individual (useful in case of the blind individual, details described in the lit. survey paragraph). The speaker will act as an earpiece.

ADC: (Analog to Digital Converter) ADC is an electronic integrated circuit used to change the Analog signals over to digital or binary form comprising of 1s and 0s. The vast majority of these converters take a voltage input as 0 to 10V, -5V to +5V, and so on and correspondingly produces digital output as a binary number. We are using ADCs to convert the received analog signal to digital to store in the microcontroller before sending it to for further processing. The circuit diagram is displayed in the figure 1 below.

DAC: A DAC (Digital to analog converter) is used to modify the stored sound in digital format to analog before sending it to the app for further processing. The circuit diagram is as shown in the figure 2 above.

Microcontroller: For the development purpose, we are using Beagle bone Black. Before starting a full-fledged production line, we will convert the modules to SMDs and bare processors will be used instead of development boards. For a fast demo of the product, we are using PC as our device with added...
peripherals like Node MCU/Arduino via a serial port for connection to a local server for sending and receiving required data[1][2].

2. Problem Statement

The paper rotates around the way that the quantity of hard of hearing quiets on the planet is generally determined to be from 700,000 to 900,000," and of these 63 percent, are supposed to be brought into the world hard of hearing, the others losing their hearing by various mishaps and the assessed number of individuals outwardly debilitated on the planet is 285 million, 39 million visually impaired and 246 million having a low vision; 65 % of individuals outwardly weakened and 82% of all visually impaired are 50 years and more established.[source: Wikipedia, verified]

If this technology helps at most 10% of the above data, it would be a huge success for both the technology and the social world.

The most common problem faced by the disabled (deaf/ mute/ blind) people in daily social life are:

1. **Overhelp-full friends or family members**: Being overhelp-full makes them feel helpless and it removes any possibility for them to learn to do things themselves. They can accomplish tasks at a much slower speed, but they can.

2. **Transportation**: Going and discovering new places is always a challenge, especially if the person resides in a rural area.

3. **Being Independent**: Most disabled people are hesitant to go outdoors until they are forced to or bound to. They are always tending to lose an intersection if they have by hearted the route to a park or nearby intersection and so are always in a need of a caretaker to reroute them to the desired path.

4. **Finding Objects**: Things move by themselves(sarcasm), if not done by some other person, and forgetting afterward. This creates a huge problem for the disabled people to get to the thing which he knew was on the kitchen countertop and inadvertently or unknowingly used spices instead of sugar while baking cookies to have the tastiest cookies baking adventure.

5. **People Talking Slowly Throws Off Lip Reading**: One of the primary thing’s individuals do when they understand somebody is hard of hearing is change to a much slower type of discourse. This is regularly done as people acknowledge it will help in lip perusing, contemplating dynamically purposeful elocution. The truth, in any case, is this makes life progressively hard for anyone endeavoring to lip read. The requirement for an amplifier who utilizes lip-perusing relies upon people to talk typically. This is how they sorted out some way to see the shapes individuals make when they talk, so they can connect with others on an undeniably pleasant level. Changing to a slower movement of talk in like manner changes how your mouth moves, and overall makes it harder to fathom what is being said.

6. **Communicating in the dark**: For most people, communicating in the dark is relatively easy to do. For those with hearing troubles, night time, and dim spaces, for example, bars or concerts represent their one of kind issues. All things considered, the hard of hearing depend solely on visual contrast, similar to lip-perusing or gesture-based communication, to speak with others. Without satisfactory light, it can immediately turn out to be practically difficult to talk viably with others. It's simply too dull to even consider seeing whatever could be deciphered adequately. Indeed, even faintly lit rooms can present huge issues for the nearly deaf.

7. **Not having a common sign language**: As odd as it would appear to the new, gesture-based communication is certifiably not a general lingo. Various nations have their norms, and even the contrasts between British Sign Language and American Sign Language are huge. Add that to the way that local territories have their varieties, much like accents in communicated in dialects, and you get a ton of normal misconceptions.[3][4]

We are trying to introduce an all-in-one smart glass technology where; we will overcome and solve some of the problems being described in the previous paragraph. For the blind person a two-way camera at the rear of glass to detect any vehicle while on road and on front cam for detecting common
objects. For deaf equipped with glass smart text and will be equipped with a microphone. To detect surroundings speech to convert it to text enabling him to read and for the dumb person, both the technologies will be helpful to communicate or understand the common world, the social world. The upcoming parts will describe how we approached the solution and how the final product affected the lives of those disabled individuals, in a good or a bad way, mostly good.

3. Implementation

The algorithms used to process the data are ASR, Google speech to text API and YOLO, and some more. The reason we implemented our ideas towards these algorithms are:

1. **Python:** Our main programming language for the system is a python in place of embedded C just because python has a greater number of predefined libraries which our coding shorter than C, where we have to define each library from scratch. And there is also a larger community for python out there in the wild rather than embedded C programming.

2. **OpenCV:** OpenCV (Open Source Computer Vision) is a computer vision technique developed for the purpose of image processing mainly. It consists of various computer vision and machine learning software libraries. It consists of plethora of algorithms mainly for the used for the purpose of object detection, pattern recognition, face recognition etc. It is mainly used to do all the operations related to Images. We are using OpenCV here to read the incoming feeds from the cameras installed at the front and back of the Smart Glass and process them into useful and required outputs.

3. **CNN:** A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which is used for the purpose of image classification, image segmentation and mainly for the purpose of image processing. It consists of five layers and with the help of these layers CNN algorithm filters the input image. (This technique is used for the hand symbol/ sign language detection algorithm.)

4. **ASR:** Automatic speech recognition (ASR) is the use of computer hardware and software-based techniques to identify and process human voice. This algorithm is used to identify the words a person has spoken and convert it to other forms or to authenticate or check the identity of the person speaking into the system. This technique will be used by the deaf individual for converting any incoming texts or speeches into a deaf-friendly or written format which will be projected into the HUD (Heads Up Display) O-Led (Organic Light-emitting diode) display. This display will be connected with the frame of the smart glass system and hanging proudly in front of one of the eyes of the individual (primarily the individual is considered only deaf or in some cases dumb. In case the individual is blind, this technique would be useless).[7]

5. **Google speech to text API:** Google Speech-to-Text is a method which enables programmers to convert the speech samples into text by applying required neural network models in an easy-to-use API. The speech to text conversion is a result of natural language processing techniques applied to the input speech sample. With the help of NLP, the input speech sample is refined and based on the developed machine model, it is converted to text accordingly. This process of speech to text conversion is highly efficient and accurate. We can enable voice command-and-control, transcribe audio from call centers, and more. It can process real-time streaming or prerecorded audio, using Google’s machine learning technology. This can be used in place of ASR technology because mainly it is Google and secondly, it is modified and can be customized to make more user friendly. Google's text to speech is more developer-friendly. The main cons of this API are this is not open Sourced like the former one, so developers most of the time go for ASR (Automatic Speech Recognition) rather than Google’s speech to text API.

6. **Google text to Speech API:** A Text-to-speech API is an technique which makes use of the natural language processing techniques to convert the text samples into speech signals, by analysing and processing the text and at that point utilizing Digital Signal Processing (DSP)
innovation to change over the processed content into synthesized speech representation of the content. This API will be useful if the person is blind, just projecting the text onto the HUD won’t solve the problem. For him, special algorithms, which converted the processed text to speech again is required to be outputted through the speaker.

7. **YOLO:** You only look once (YOLO) is a real-time object detection system. YOLOv3 is incredibly quick and precise. Besides, we can without much hectic, trade off among speed and exactness basically by changing the size of the model, no retraining required! Earlier recognition frameworks repurpose classifiers or localizers to perform location. They apply the model to a picture at different areas and scales. High scoring areas of the picture are viewed as identifications. YOLO utilizes an entirely unexpected methodology. YOLO applies a single neural network to the entire image and divides the image into different parts. It also assigns different parts, unique weighted values by predicting the probability of different parts of the image. Because of this methodology it is highly efficient and faster as compared to some other object detection models. The model has a few focal points over classifier-based frameworks. It takes a glance at the entire picture at test time so its forecasts are educated by worldwide setting in the picture. [Source: https://pjreddie.com/darknet/yolo/]. YOLO is used in the rear camera of the smart glass to convert incoming objects from the back of the person to speak, which can, in turn, be converted to text using the former algorithms and be projected into the HUD of the person.

![Figure 3: YOLO for Object detection](image-url)
8. Web API: A server-side web API is a programmatic interface consisting of one or more publicly exposed endpoints to a defined request-response message system, typically expressed in JSON or XML, which is exposed via the web—most commonly utilizing an HTTP-based web server. Web APIs will be used to access the database for image comparison for the hand symbols/signals and also the front camera object detection/identification using the YOLO algorithm.

9. Image Search and Image Recognition API: The Image Search API allows the developer to easily integrate image search capabilities within the app or website or any device using real-time methods. They provide JSON responses with high-quality thumbnails, URLs, and relevant metadata. The Image Search API will be certainly useful in cases where the blind individual comes across some unknown object previously not present in the database and so that the device can notify about the obstacle, in certain conditions where the device might fail in recognizing certain things, the world-wide-web might come to rescue.

10. HTML5 Geolocation feature along with the Maps JavaScript API: Displays the geographic location of a user or device on a Google map, using the browser’s HTML5 Geolocation feature along with the Maps JavaScript API. (A point to be taken into consideration here and i.e. this system will only work if the user has confirmed with sharing his location data. In this case, the system will by-default turn on the services for the blind individual to know the place he currently is or he can through voice input ask the system to give his current location at any moment).

Now, the explanation of the hardware and software is being described.

4. Theoretical Approach

The theoretical approach is being described here:

- Deaf Individual: For the particular deaf individual, we are implementing a two-way camera system, one at the back of the head and one at the front, paired with microphones. The camera at the back will detect any incoming obstacles like vehicles from the back of the individual and send them to the microcontroller for processing. The microphones will detect any sound from
the surroundings, mainly speech, and send them to the microcontroller using wireless techniques. After the microcontroller has all the input data, it will process all those according to the set limits and boundaries and send them as text messages to the HUD of the smart glass to be seen and act accordingly. In a real-world scenario, let there be a car coming from the back of the deaf individual. Since he won’t be able to hear that, the camera at the back will capture it and send to the microcontroller. In response to that, the microcontroller will send an alert message to be projected on the HUD warning that a car is incoming. The person can see the warning and act on it. Another case might be that a person is calling the deaf person. The microphones will detect the call and send them as speech to the microcontroller and microcontroller will similarly send a message to the HUD of the smart glass.

- **Dumb Individual:** For the case of a dumb individual, we proceeded completely app-based rather than any hardware-based. For the dumb individual, we have coded an app (application similar to the Android app or IOS app), which will detect(capture) the hand movements or signals from the dumb individual and convert it to speech or text. This app will be on the receiver side and not with the transmitter side. The receiver (normal person, who is considered physically fit) will aim the video camera at the acting hand symbols of the dumb person. The video cam of our application will input the video, process it, and output the required data on the mobile screen or as a speech output.

- **Blind Individual:** For the blind individual, cameras fitted at the front and back of the device comes to the rescue. The front camera does all the jobs as the back camera like detection of incoming obstacles and all and also additionally it will perform identification of obstacles like identifying a TV, table and converting them to speak through the microcontroller, to send to the speaker to be outputted or scanning a QR code, wherever useful. As a shortcoming, it may happen sometimes that the camera can’t specify what kind of object is the camera looking at. Then, the camera will capture the image, do an image search and output the top result in the category and in case it doesn’t happen, it can only describe the name of the same shaped objects and derive a path accordingly for the blind individual to walk through. There are fitted proximity sensors on the front frame of the smart glass body which will detect any obstacle within the preset range and will warn the individual about the same. Besides, the blind individual will have voice assistant technology if he requires any time, like make a call or know the current time or if he needs to nag anyone for some help.

- **The app:** All the raw data, feed, data from the microphone will be received to the android app. The video feed will be sent to the app through wireless communication through IP address, i.e., the app’s backend code will access the IP address of the ESP module to access the video feed and work on it as described in the Methodology paragraph. The receiving of analog sound directly through the microphone will incur a lot of noise both while recording and also while transmitting. So, we will record the sound through a microphone, convert it to digital using ADCs and then store it in the microcontroller before sending it to the app, for both noise reduction and also synchronizing with the ESP clock signal for sending. After it is stored in the microcontroller, again during sending it is converted to analog by DACs for the app to receive and process it in the backend. The app will also receive both the distance data from the proximity and ultrasonic sensors for notifying the user about any incoming obstacle the computer thinks the user should be aware of. For reducing the processing time of the app and the backend codes, there will two subscriptions of the app, or we may update our system for 3 different types for people having a total disability, having a blind and deaf disability at the same time. One subscription will be there for managing the blind individual, which will only receive the video feeds from the cameras with some added benefits of the microphone for voice assistant and other subscriptions will be working for the deaf and dumb individual, for processing the incoming sounds and speeches. There is no requirement for the deaf and dumb individuals to have extra cameras on their eyes because they can primarily see and have no problem with that. For the blind individual, we have still all the hardware as described, but the functionality like
listening to any incoming speeches will be disabled and the microphone will only work on any predefined calls and triggers, to reduce the workload on the processor and mobile device. The subscription selection will be one time while starting and pairing the device with the app for the first time. For the third type of subscription, we don’t have any way rather than utilize all the hardware mentioned and all processing to be done to help the individual. [5,6]

How the product will affect the individual?

- **For Blinds:**
  - The product will help the blinds to detect objects and will allow them to interact.
  - They will be able to read independently.
  - They can manage money and things on their own.
  - Can navigate around avoiding the obstacles
- **For dumb:**
  - Can communicate with others through sign language and others can listen to sign as speech, through the android app as described in the previous section
- **For deaf:**
  - They can read what others are saying and will be able to respond accordingly.
  - Input can be processed and converted to text on HUD.

The traffic light detection using OpenCV can be performed for the Blind Individual. The object detector algorithm will be modified such that detects colours and whenever it sees red colour in an object similar looking to traffic light post, it will warn. Accordingly, it will work for other colours like green, yellow. The working will go like this: The system proposed in the flowchart below.

![Flowchart](image.png)

**Figure 5:** Flowchart For algorithm to determine traffic light color to aid the blind.
shortcomings at the beginning for the detection of vehicles coming from the sides. To minimize the shortcoming, we could’ve attached more cameras at both the sides of the smart glasses, but that would increase the hardware cost and also increase the software processing time, for processing the incoming (count 4) numbers of raw footage at any moment and will cause a final delay in outputting the required processed data, which might cause a massacre in a real-world scenario where instant processing power is required.

So, instead of using more than 2 cameras at the front and back combined, we might work on the placements of the cameras at the front and back, like attaching wide-angle cameras at the opposite diagonals will capture a rather wide range of video, than directly attaching at 90˚ to the front tangent. (The given solution is only the thoughts of the author and may or may not be seen at the end product. The main thing to consider is the problem statement is real and the developers are working on removing the shortcoming at the earliest.)

5. System Working

Detailed methodology about how the product will be useful for the blind individual:

- Sending video to server: Since ESP32 is also a development board, we can directly upload codes to the board to send videos to the server. By inserting the network credentials to the code, the serial monitor will give the board’s IP address, which will show us the real-time video captured from the cam (before uploading the code we have to make sure we are using an FTDI programmer and selecting correct board make and model while uploading).

- Accessing the IP cam in python: We need to extract the URL from the IP address for working with the video stream in OpenCV. We need to make sure that the URL directly leads us to the video stream and not to any HTML page where we can view the video stream. Just because OpenCV can only read the video stream if the filename/URL has the right suffix. Adding? type=some.mpeg sometimes might be helpful. (camera=cv2.VideoCapture("http://192.168.43.1:8080/video")) example

- The main processing starts here. At first, after importing all the required libraries, we have to create an instance of Video Object Detection class. Then, after setting the model type to YOLOV3, which corresponds to a YOLO model library, which contains a huge set (at least 80) of different kinds of objects for comparison. Then after setting the model path to the model file,
we load the model into the instance of the VideoObjectDetection class that we created. Then we called the detectObjectsFromVideo function and parse the required arguments (we need to state the camera input into one of the arguments) to the function. After detecting the object's name, we will convert them to image to speech and send them to the connected speaker module (earpiece) notifying them about the incoming obstacle from behind or front.

Figure 7: Flowchart of Object detection for the blind

- We will also check the proximity sensors and alert the person about any obstacle from the sides and aware of him of her.
- Since there is no direct method to convert image to speech, we need to extract the object type from the model described above and send them to another algorithm, known as text to speech.
- Text to speech works in a similar platform as the video object detection. The text we received is to be stored in a variable and using gTTS algorithm convert them to sound and then play them on the speaker (earpiece) connected to the smart glass.

We have also made arrangements for a version helpful for the deaf and dumb people to take advantage of our model:

- For the deaf and dumb individuals, we have connected microphones on the system to collect any incoming sound from the surroundings and send them to the HUD of the smart glass as a text message.
- For converting speech to text, we require certain predefined modules. We need to install speech-recognition, pyaudio, pyttsx3. We need to process the received sound in such a way that we allow the program a second or two to adjust the energy threshold of recording so it is adjusted according to the external noise level.
- After the required sound is received, we run speech to text translation. (This requires an active internet connection) We need to initialize the libraries using init() function with its arguments. Finally, to run the speech we use runAndWait(). All the say() texts won’t be run unless the interpreter encounters runAndWait().
- After we received the required text output, we can just output the same to the HUD on the smart glass.

As an add on we will implement certain other API’s like the Image search and Image recognition API’s and HTML5 Geolocation feature along with the Maps JavaScript API to notify the user about the current location he is standing and also if the YOLO model fails to identify some of the objects it notices, it can run a normal Google algorithm and notify about the top search results to describe the object.

6. Conclusion
This research shows how different IOT technologies along with smart machine learning algorithms can be used to resolve the problems that has persisted over the years in case of blind, dumb and deaf. Using
automatic speech recognition would help give the voice to the dumb, help the deaf understand what the people in the vicinity are talking. Live Image transmission along with on board object detection would help guiding the blind while walking, in short giving him the eye to see things around. The implementation of smart glasses would help resolves problems of numerous blinds, dumb and deaf person who lack the confidence while communicating and even guiding them. This low-cost effective technology would be efficacious and provide the blind the eyes, the deaf his ears and the dumb his tongue. Smart Implementation using sensor and machine learning are evolving to help the mankind and this is one of the examples that could help change the scenarios pertaining to blind, dumb and deaf.

7. References

[1] R. Rastogi, S. Mittal and S. Agarwal, "A novel approach for communication among Blind, Deaf and Dumb people," 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2015, pp. 605-610.

[2] B. Rajapandian, V. Harini, D. Raksha and V. Sangeetha, "A novel approach as an AID for blind, deaf and dumb people," 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS), Chennai, 2017, pp. 403-408.

[3] G. G. Nath and V. S. Anu, "Embedded sign language interpreter system for deaf and dumb people," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2017, pp. 1-5.

[4] Indian Sign Language Recognition System for Differently-ablePeople, Goyal, Kanika Singh, Amitoj, Indian Sign Language, Open CV, Image Processing. Chitkara University Publications

[5] Estrada Jiménez L.A., Benalcázar M.E., Sotomayor N. (2017) Gesture Recognition and Machine Learning Applied to Sign Language Translation. In: Torres I., Bustamante J., Sierra D. (eds) VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th -28th, 2016. IFMBE Proceedings, vol 60. Springer, Singapore.

[6] Jafri, R., Ali, S.A.: Exploring the potential of eyewear-based wearable display devices for use by the visually impaired. In: International Conference on User Science and Engineering, Shah Alam, 2–5 September 2014.

[7] Velázquez, R.: Wearable assistive devices for the blind. In: Lay-Ekuakille, A., Mukhopadhyay, S.C. (eds.) Wearable and Autonomous Systems. LNEE, vol. 75, pp. 331– 349. Springer, Heidelberg (2010).

[8] AJ Obaid, S Sharma, Recent Trends and Development of Heuristic Artificial Intelligence Approach in Mechanical System and Engineering Product Design, Saudi Journal of Engineering and Technology 5 (2), 86-93, 2020.