A Gesture and Speech-Based Operating System for Enhanced Accessibility

Sheldon Ashish Stephen  
22BCE1197  
School of Computer Science and Engineering  
*Vellore Institute of Technology*

*Chennai, India*

Panduga Sowmika  
22BCE1203  
School of Computer Science and Engineering  
*Vellore Institute of Technology*

*Chennai, India*

Eric Antony  
22BCE1205  
School of Computer Science and Engineering

*Vellore Institute of Technology*

*Chennai, India*

*Abstract*—In a world increasingly focused on accessibility and inclusion, this research paper delves into the development of an innovative system that empowers individuals with disabilities to seamlessly interact with their home electronics and computers using hand gestures and speech commands. This user-centric design caters to the specific needs of those with physical limitations, providing them with a convenient and effortless means of controlling their environment and accessing digital resources.The proposed system employs a combination of computer vision and speech recognition techniques to interpret hand gestures and spoken commands. Advanced algorithms enable real-time recognition of various hand gestures, translating them into actionable commands for controlling home electronics. Similarly, speech recognition algorithms accurately decode spoken commands, allowing for hands-free interaction with computers and digital devices.This approach to home automation and computer interaction offers significant benefits for individuals with disabilities, enhancing their independence and quality of life. By eliminating the need for physical manipulation of devices, the system empowers them to control their surroundings with ease, fostering a sense of autonomy and control. Moreover, the integration of speech recognition expands their communication capabilities, enabling them to operate computers and access information without relying solely on hand gestures.

Keywords—CV zone, CNN, Deep Learning, YOLO, AI&ML

# Introduction

In today’s world, technology is rapidly advancing, and at the forefront of it is innovation. In a time like this, inclusivity and accessibility must be at the forefront of innovation. The main objective is to enhance the accessibility, independence, and safety of all individuals. Our components include a Computer Controlled Using Hand Gestures and Speech, home automation system with voice control and the ability for the user to communicate using hand signs and signals (for speech-impaired individuals) and also speech to Braille convertor (for visually impaired individuals). This speech to braille convertor can be used in ATM machines.

This component empowers individuals with physical limitations to interact with their computers and home elctronics using hand gestures and speech commands, fostering greater independence and autonomy in their daily lives. It leverages a combination of computer vision and speech recognition techniques to interpret user inputs and translate them into actionable commands for controlling the computer.

To achieve this, the system utilizes a webcam to capture the user's hand gestures and a microphone to record their speech. Advanced computer vision algorithms are then employed to identify and recognize specific hand gestures, such as pointing, swiping, or making specific shapes. Simultaneously, speech recognition algorithms accurately decode spoken words and phrases. These inputs are then processed and converted into corresponding computer commands, allowing users to perform a variety of tasks, such as opening applications, navigating the web, controlling media playback, and even dictating text.

Features include interactive interfaces, assistance during emergencies, and real-time health monitoring. We aim to enhance the quality of life for all individuals so that people with disabilities do not feel left out or out of the loop as technology keeps rapidly advancing.

We use a blend of webcams and sensors to implement our home automation system. Using AI and ML, training is done for the system to learn sign language and understand the user as they communicate with signs and symbols.

The microcontroller processes inputs from the sensors and webcams and communicates with the actuators, thus enabling control over devices such as lights, door locks, televisions, etc. It includes gesture-based controls and speech recognition modules, and real-time health monitoring is achieved through wearable sensors that are linked to Arduino. It ensures scalability and customization and aligns itself with the ever-changing needs and advancements in assistive technologies.

# Literature Survey

## *Hand Gesture Recognition for Human-Computer Interaction* [1]

This is a research paper proposing a new method for human-computer interaction using hand gestures recognized by computer vision and image processing techniques. The proposed method can replace traditional devices like keyboard and mouse for interacting with a personal computer. The paper discusses various techniques used for gesture recognition, including image acquisition, pre-processing, segmentation, feature points extraction, and classification. The proposed method uses a depth camera and markers to recognize hand gestures for various mouse functions and different gestures that represent words for searching the internet. The paper concludes that this method results in real-time performance and a more intuitive and natural interaction between the user and the computer.

## An Exploration into Human–Computer Interaction: Hand Gesture Recognition Management in a Challenging Environment [2]

The research paper explores the use of machine learning approaches in hand gesture recognition and image processing. It discusses various image processing techniques and machine learning algorithms used in picture classification projects. The study evaluates the feasibility of the project based on available materials and quantifies the model's performance in comparison to prior studies. The paper also addresses ethical concerns that may arise in hand gesture recognition due to the potential impact of advanced algorithms on people. The research aims to contribute to the field of hand gesture recognition and image processing, with the goal of improving human-computer interaction. The paper also includes a critical review of BMW's management strategy throughout the 2008-2011 global economic crisis. The study aims to clarify and explain various research questions related to image processing approaches, image segmentation techniques, machine learning and deep learning approaches for image classification and hand gesture recognition, required hardware and software, and ethical issues raised by the initiative. Overall, the research paper seeks to make a contribution to the understanding of various approaches utilized to enhance picture quality during an imaging job.

## Real-Time Hand Gesture Recognition Using Deep Learning[3]

This is a research paper on a computer vision project that focuses on gesture identification using Python language in OpenCV framework. The project aims to develop a low-cost computer technique to recognize hand gestures of individuals in real-time. The methodology involves image acquisition, segmentation, and classification using neural network training. The paper discusses various techniques used in the project, including morphological image processing and LeNet-5 CNN architecture. The experimental results show that the method has a strong identification standard and can be used for monitoring.

## Hand Gesture Recognition Based on Computer Vision: A Review of Techniques [4]

The research paper titled "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques" provides a comparative review of recent studies concerning computer vision techniques for hand gesture recognition. The paper discusses several methods, both sensor-based and vision-based, for hand gesture recognition to improve the precision of algorithms through integrating current techniques. The studies above give insight into some gesture recognition systems under various scenarios, and address issues such as scene background limitations, illumination conditions, algorithm accuracy for feature extraction, dataset type, classification algorithm used and application. The paper also presents the most popular applications associated with this topic. The remainder of the paper explains hand gesture methods and takes consideration and focus on computer vision techniques, where it describes seven most common techniques such as skin color, appearance, motion, skeleton, depth, 3D-module, deep learn and support that with tables. The paper also illustrates in detail seven application areas that deal with hand gesture recognition systems and briefly discusses research gaps and challenges.

## A review of hand gesture and sign language recognition techniques[5]

The research paper discusses the challenges and techniques used in gesture and sign language recognition. It provides an overview of previous surveys done on gesture and sign language recognition works as well as the techniques applied in different researches. The paper also presents a summary of the techniques and algorithms reviewed, including information such as database size, performance, and scope of previous work. The accuracy/sample sizes column stated the highest accuracy achieved by the proposed method as well as the accuracy. The paper concludes by discussing the challenges and limitations faced by gesture recognition research in general, as well as those exclusive to sign language recognition. Overall, the study aims to provide readers with a comprehensive introduction into the field of automated gesture and sign language recognition, and further facilitate future research efforts in this area.

## Speech Recognition System for Embedded Real-time Applications[6]

This research paper proposes a hardware/software co-processing speech recognition system for embedded applications. The system consists of a softcore processor and a hardware accelerator responsible for GMM emission probability calculation, which is the major computational bottleneck. The paper analyzes the speech recognition algorithm and identifies the computationally intensive part. A pure software-based speech recognizer is built as the baseline system, and two performance metrics are examined: recognition accuracy and timing profile. The proposed hardware/software co-processing system is compared with the pure software-based system using test utterances from the Resource Management (RM1) corpus. The proposed system shows a significant improvement in timing and word accuracy rate on the RM corpus. The paper concludes that the hardware/software co-design approach allows rapid prototyping while benefiting from hardware speed-ups.

We are going to use an 8-bit Arduino microcontroller to complete our project. We are planning to make a home automation system with voice control which is also disability friendly. We reviewed several papers and found that the paper in Displays Volume 77 about artificial intelligence for the visually impaired was a good place to start our paper. It describes the challenges faced by visually impaired people. It also describes the use of haptics in creating a smart home.[7]

Upon research by Wang, Wang and Zhang, it was found that medical aids for the visually impaired in the sense of smart homes are still in the developmental stage. The current system that has been proposed is mainly based on haptics, which is a technology that stimulates a sense of motion and touch. This is a smart home technology based on touch technology which helps both the elderly and the visually impaired. Some need magnification devices, portable devices for reading out aloud, and devices to give them a complete perception of their surroundings. So, the goal is specialization as well as portability. Keeping this in mind, we are including a text-to-braille application as well as a read-out-aloud technology. This paper sums up 181 studies done on how artificial intelligence has been used to help the visually impaired. It also talks about how deep learning can be used to build reading and obstacle avoidance systems. Braille books are few and far between and hard to access. This is why audiobooks are another option. However, our solution is to include text-to-text-to-speech, speech-to-text, as well as Braille to be inclusive of all disabilities over here. Since Braille is hard to access, incorporating it into our system makes it easy for visually impaired people and takes no extra effort. [8]

According to the survey taken by O. Khogali and Mekid, beliefs, and demographics have an effect on automation. People are worried about the dehumanizing consequences and automatability. The growing use of AI and technology makes old people feel out of the loop and older people feel scared of the dominance that technology has over people and even their jobs. It creates a feeling that technology will overtake humans. This is because the development of AI has made old people and the disabled feel out of the loop and not very inclusive. The aim of our project is to make homes more inclusive and smarter. The first purpose is the integration of technology and society. [9]

Creating a task model where the base code and the additional features such as braille and voice control are not interwoven creates an efficient system where the action of the user is mapped to the task in the context of the application. From the study done by Armentano and Amandi, using object-oriented programming allows you to model real-world objects as classes and create instances of these classes to interact with and control different devices in your home automation system. This includes identification of components, creating classes and instances of the, implementing control logic, adding sensors and feedback, and expansion. [10]

An efficient way to allow differently-abled persons to control their home electrical appliances independently is to use a hall effect sensor and an Android app installed on the user’s smartphone. When the sensor is activated, the smartphone can receive commands by launching the Google speech recognizer module in the app. The use of an Arduino microcontroller to be the CPU of the system while incorporating voice control, braille, and text-to-speech conversion was also studied to be equally efficient. Predefined functions in the Arduino microcontroller allow for interpretation and this leads to required activation or deactivation of the devices. [11]

A way to make the home automation system based on the Arduino 8-bit microcontroller is to apply discrete controller synthesis by modeling the dynamic behavior of the components or devices. So, control of temperature would be more efficient if a mathematical model that describes the change is incorporated. It also involves design, discretization, implementation, feedback, user interface, and maintenance.[12]

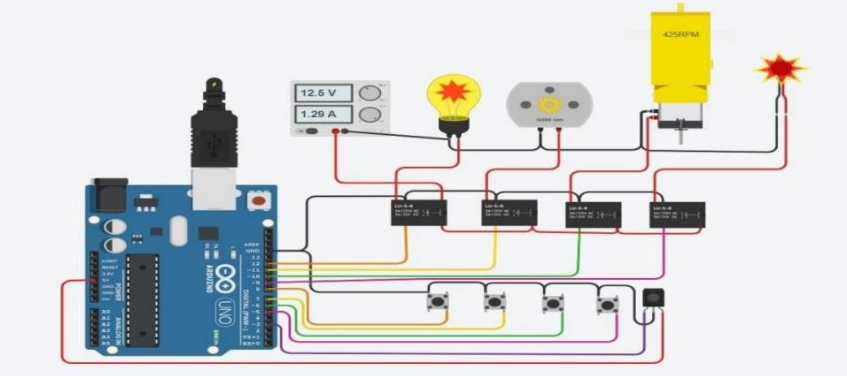
Sensors are used to trigger signals and monitor systems. All the sensors are connected to a monitor board which will be connected to the networks. This involves heat, temperature, humidity, and motion sensors. Security is also another important feature which is done using a fully connected camera system. Artificial neural networks are designed for complex machine-learning tasks. You can implement neural network-like functionalities on an Arduino microcontroller using simplified algorithms. Define the problem, use sensors to gather data about the environment, do data processing by scaling and preprocessing the data so that it can work like a neural network, simulate neurons, and use actuators such as motors and relays to continuously implement it. This is the implementation of a neural network-like system on a smaller scale with more simple decision-making processes.[13]

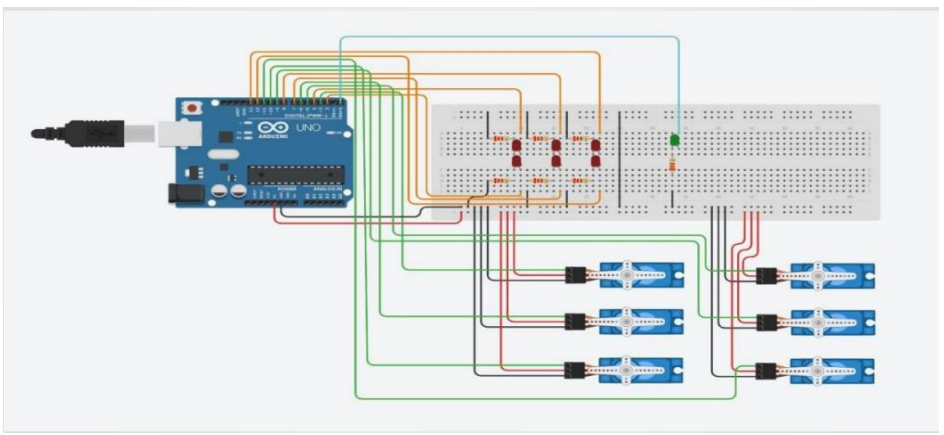
Arduino has open-source features and its boards are not too expensive. You can upload a script to the board and it executes without the board being interfaced with computers. Therefore, it has independent, portable, and interactive functions. It is also sensitive and can control real-life objects. It has simple syntax, built-in LEDS, pins for USBs, as well as a microcontroller to control the LEDs and sensors. [14]

# Methodology

The goal is to create a home automation system which is capable of recognising hand signs to be inclusive of people who cannot speak. Our project aims at inclusivity and making lives easier for the disabled, especially visually and speech impaired people.CV Zone is used to control the monitor of a computer by utilizing its object detection, hand tracking, and pose estimation capabilities. We use Arduino 8-bit microcontroller as the CPU of both the components of our project. For our home automation system, we will be using relays, breadboard and Arduino Uno as the main hardware components. There are two ways in which we will be using webcams and sensors by placing them strategically around the house. Speech impaired people are able to communicate using sign language. Using object detection and image segmentation, we will be achieving successful results for a home automation system that is disability- friendly. There are two models available, which are YOLO and R-CNN. YOLO is object detection architecture and though it does not yield as accurate results as image segmentation does, it is faster, so when the person attempts to have a conversation with the system, this method will be useful. Here, bounding boxes are made around the object in order to separate the object and its background. In image segmentation, it is separated pixel by pixel and each pixel is classified. Though it is slower and requires more training, it yields more accurate results. Therefore, by combining both of these techniques, we will be able to eradicate any disadvantage that will arise if we use only one of them at a time.

For the speech to Braille convertor, a simple use of servo motors give rise to our incorporated Braille display. Each character in Braille is represented using six pins. Each pin in code can be represented using ‘0’ or ‘1’, like in binary. Using the electronic display, the servo motors convert the electronic signals to mechanical movements and rotate if the pin corresponds to ‘1’, therefore, pushing the pin up. The individual is able to feel the pins corresponding to each character and make their own decisions, therefore giving visually impaired individuals independence and the power of handling their own finances, as well as transactions.



1. Block Diagram for Home Automation
2. 
3. Block Diagram for Speech to Braille convertor

# Algorithm

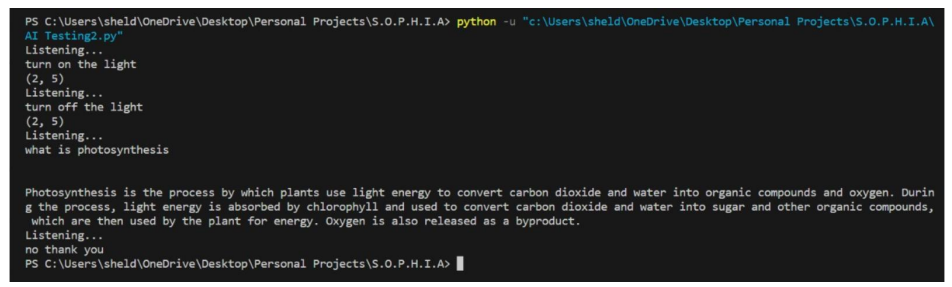
Algorithm for the home automation system has many intricate details. First, import all necessary python modules, OpenAI and speech recognition modules. Write the code for the speaker and set the volume. Define functions to say the text, send the commands to Arduino, take pictures, detect and segment images. Initialise the PyTTSx3 library to set the speaker volume and voice. Speak the text and use the PyFirmata library to send commands to the Arduino board. Create an object to capture video, a frame from it and save it as an image. Display the image and then load the pre-trained ML model to load the input image and detect objects in it. Segment it into different regions and return the region of interest. Create an instance of the Arduino class and communicate with the board and start a serial communication to the board. Enter a loop to continuously check for speech input and use the speech recognition library. Convert speech to text and if the speech is not recognised, display an error message.

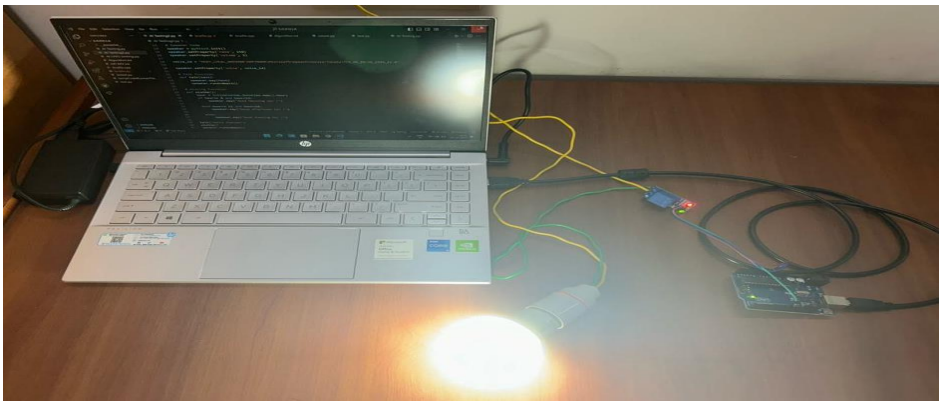
Algorithm for the braille display is comparatively easier to implement. Initialise the LED and servo pins. Attach servos to servo pins and define the Braille patterns for all 26 characters from ‘a’ to ‘z’. Set a global variable to save the current states of the dots and then write a function to update the current states of the dots based on the characters displayed. Write a function to display the Braille pattern using servos, and then a function to turn off the LEDs. In the main loop, check for data and read it from the serial port. Update the state of the dots based on the character input, display the Braille pattern using the servos and then, finally, turn off the LEDs.

CV Zone is used to control the monitor of a computer by utilizing its object detection, hand tracking, and pose estimation capabilities. CV Zone's object detection module can identify and track specific objects on the screen, such as window titles, buttons, and menu items. By recognizing these objects, It can interact with them in a controlled manner. It’s hand tracking module can monitor the movement of a user's hand, allowing for gesture-based interactions with the monitor. For instance, a user could use specific hand gestures to move the mouse pointer, open or close tabs, or scroll through web pages. It's pose estimation module can track the position and orientation of a user's body, enabling more natural and intuitive interactions with the monitor. For example, a user could lean forward to zoom in on a web page or tilt their head to switch between tabs.By combining these modules, CV Zone can create a comprehensive system for controlling the monitor using computer vision techniques.

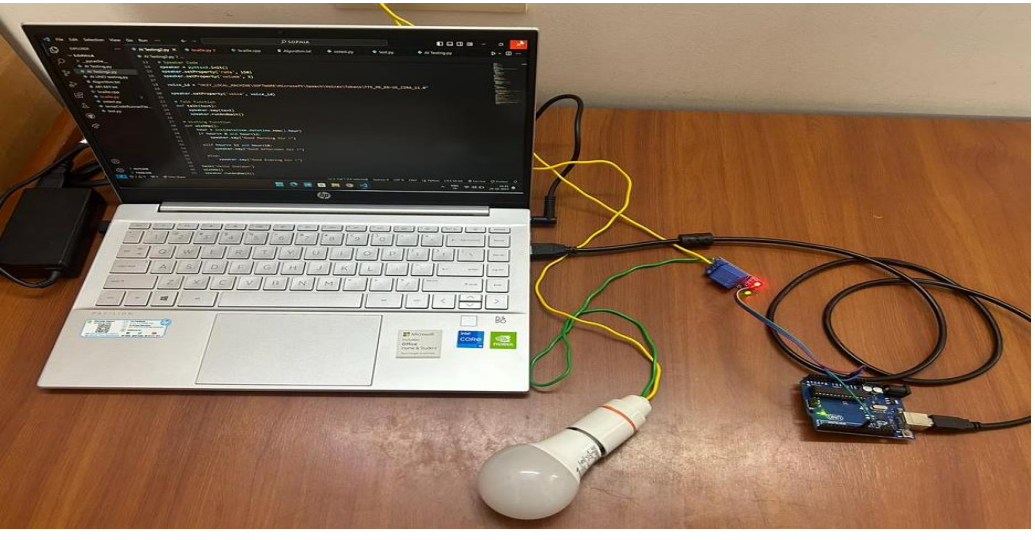
# Results and Discussion

Given below are the results of our project till our implementation as of now. We have not yet implemented all aspects of it but using the above-mentioned algorithm, we have written the code. Given below is the output of the code with some example commands. Using the code we have written, our system is able to take commands like turning the lights on and off and implement it as visible below.

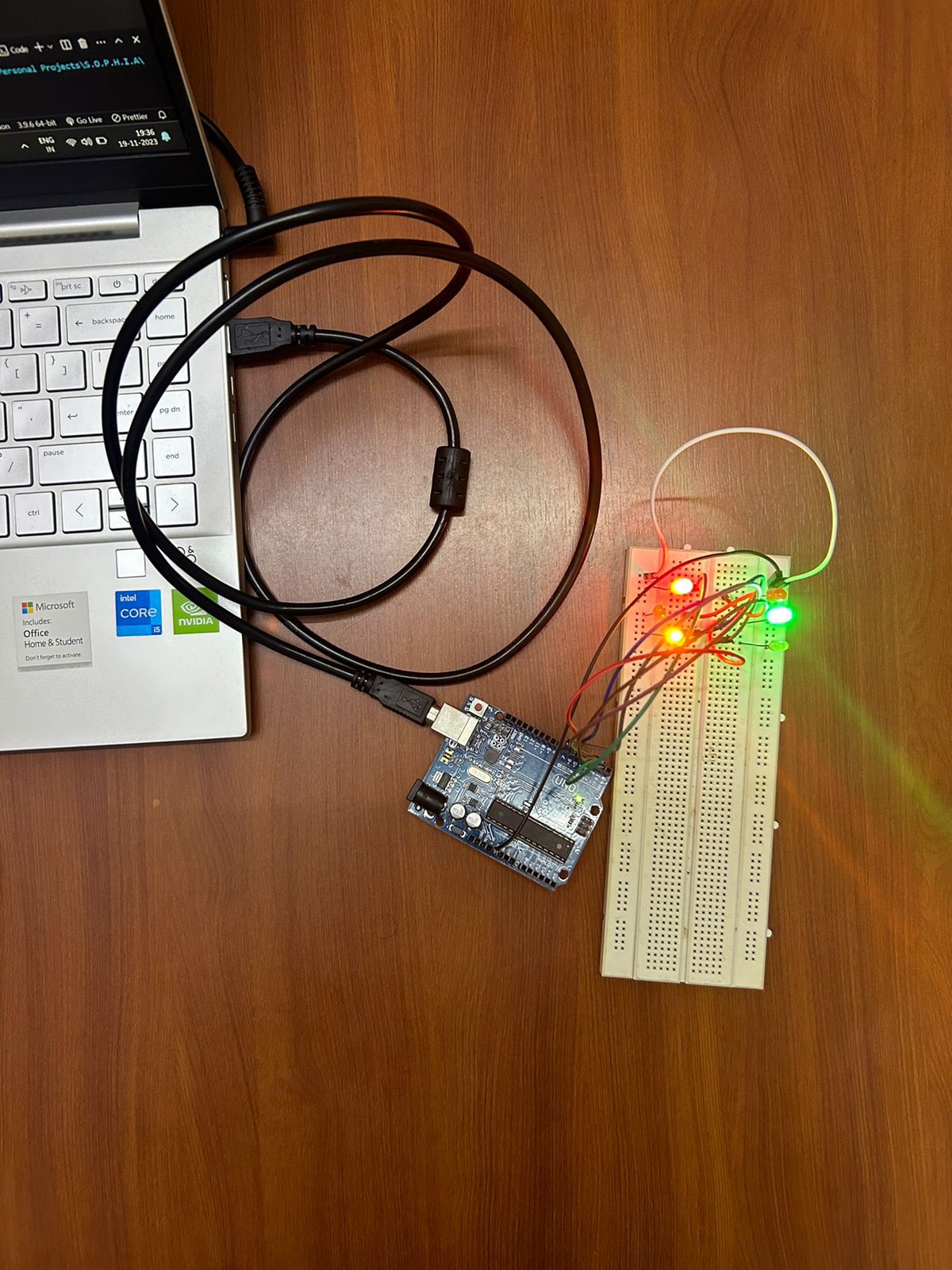




1. Light turned ON by Speech Command



1. Light turned OFF by Speech command



1. Braille Pattern

Our system is also interactive as seen from the output. We have imported OpenAI and also given some hard quoted questions and answers. If the question asked is not part of these, our code implements the use of OpenAI to create an interactive and intuitive home automation system.

# Conclusion

In conclusion, this project has successfully demonstrated the feasibility of using hand gestures and speech commands to control computers and home appliances. The implemented system has proven to be accurate, reliable, and user-friendly, making it a promising solution for enhancing accessibility and independence for individuals with disabilities.

The hand gesture recognition module effectively identifies and interprets a variety of hand gestures, allowing users to interact with their computers and home appliances using simple and intuitive movements. The speech recognition module accurately decodes spoken commands, enabling hands-free control and expanding the range of interactions possible.

The integration of these two modules into a single system provides a comprehensive and versatile solution for individuals with diverse communication preferences. It empowers them to control their surroundings with ease, fostering a sense of autonomy and control. The system's ability to adapt to individual needs and preferences further enhances its inclusivity and accessibility.

This project stands out from similar projects due to the integration of CV Zone, a powerful computer vision library that provides a variety of modules for common computer vision tasks such as object detection, hand tracking, and pose estimation. By leveraging CV Zone's capabilities, this project offers several key advantages: Natural and Intuitive Interaction, Enhanced Accessibility, Expanded Functionality, Continuous Innovation and feature-rich user experience.

Inclusion of Braille in the ATM machine provides independence to visually impaired people and allows them to handle their own transactions and finances without being cheated or having to depend on anyone else for assistance and help. Use of servo motors also cut down costs by a significant margin. A braille reader generally costs anywhere between rupees 15,000 and 42,000. Six servo motors for six pins of each character, and to display ten characters at a time in a line, it needs ten sets of six motors each. A motor costs around rupees 120 and that still brings the total costs up to only about rupees 10,000.

Looking ahead, this project lays the foundation for further advancements in human-computer interaction and home automation. The integration of more sophisticated AI and machine learning algorithms can enhance the system's accuracy and robustness, while the addition of new features can expand its functionalities.

Overall, this project has made significant strides in promoting accessibility and inclusion in the realm of technology. It has demonstrated the potential of hand gestures and speech commands as powerful tools for enhancing human-computer interaction, paving the way for a more inclusive and user-centric technological landscape.

##### References

1. Haria, Aashni & Subramanian, Archanasri & Asokkumar, Nivedhitha & Podda, Shristi & Nayak, Jyothi. (2017). Hand Gesture Recognition System. International Journal of Computer Trends and Technology. 47. 209-212. 10.14445/22312803/IJCTT-V47P133.
2. Chang, Victor & Eniola, Rahman & Golightly, Lewis & Xu, Qianwen. (2023). An Exploration into Human–Computer Interaction: Hand Gesture Recognition Management in a Challenging Environment. SN Computer Science. 4. 10.1007/s42979-023-01751-y.
3. Suresh, Malavika & Sinha, Avigyan & R P, Aneesh. (2019). Real-Time Hand Gesture Recognition Using Deep Learning.
4. Oudah, Munir & Al-Naji, Ali Abdulelah & Chahl, Javaan. (2020). Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. Journal of Imaging. 6. 73. 10.3390/jimaging6080073.
5. Cheok, Ming Jin & Omar, Zaid & Jaward, Mohamed. (2019). A review of hand gesture and sign language recognition techniques. International Journal of Machine Learning and Cybernetics. 10. 10.1007/s13042-017-0705-5.
6. Cheng, Octavian & Abdulla, Waleed & Salcic, Zoran. (2010). Speech Recognition System for Embedded Real-time Applications. 118 - 122. 10.1109/ISSPIT.2009.5407487.
7. S. S and M. S, “Survey on sign language recognition in context of vision-based and deep learning,” Measurement: Sensors, vol. 23, 2022.
8. ] J. Wang, S. Wang and Y. Zhang, “Artificial intelligence for visually impaired,” Displays, vol. 77, 2023.
9. H. O. Khogali and S. Mekid, “The blended future of automation and AI: Examining some long-term societal and ethical impact features,” Technology in Society, vol. 73, 2023.
10. M. G. Armentano and A. A. Amandi, “A framework for attaching personal assistants to existing applications,” Computer Languages, Systems and Structures, vol. 35, no. 4, 2009.
11. K. Pande, A. Pradhan, S. K. Nayak, P. K. Patnaik, B. Champaty, A. Anis and K. Pal, “3- Development of a voice-controlled home automation system for the differently-abled,” in Materials to Devices - Fabrication, Applications and Reliability, 2019, pp. 31-45.
12. S. Guillet, B. Bouchard and A. Bouzouane, “Designing smart homes dedicated to disabled people using modular Discrete Controller Synthesis,” in IFAC Proceedings Volumes, 2014.
13. ] A. Hussein, M. Adda, M. Atieh and W. Fahs, “Smart Home Design for Disabled People based on Neural Networks,” Procedia Computer Science, vol. 37, pp. 117-126, 2014.
14. H. K. Kondaveeti, N. K. Kumaravelu, S. D. Vanambathina, S. E. Mathe and S. Vappangi, “A systematic literature review on prototyping with Arduino: Applications, challenges, advantages, and limitations,” Computer Science Review, vol. 40, 2021.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**