

Hand Gesture Recognition and Conversion to Speech for Speech Impaired

E. Annapoorna^{1*}, *Bonam John Nikhil*², *Burra Kashyap*², *Jula Abhishek*², *Tarun Sri Sai Vadlapatla*²

¹UG Student, Department of CSBS, GRIET, Hyderabad, Telangana, India

²Associate Professor, Department of CSE, GRIET, Hyderabad, Telangana, India

Abstract. The primary method of interaction between individuals is communication. In the case of speech-impaired people, the communication between them and normal people is not understandable by the latter as the former communicates through gestures and sign language. In the case of speech-impaired. Today, the Internet of Things (IoT) can be used for gesture recognition and to convert gestures into audible speech. In this project, a prototype is created such that the gesture or the sign language that is communicated is recognized by the flex sensors connected to the fingertips of a plain glove. The ESP32 microcontroller recognizes the gesture and converts it into the text format and this text format is updated onto the Google Firebase and is converted to speech form thus producing the text onto the OLED display sensor and the converted speech for communication in the mobile application Kodular Companion.

1 Introduction

The Internet of Things, or IoT, is a network of interconnected computing devices, mechanical and digital machines, things, animals, or people with the ability to transfer data over a network without having to interact with each other or with computers. Users can interact with computers using either method without the need for touch, controllers, or other devices. While markers, gloves, or sensors are used in some hand-tracking and gesture-recognition systems, the ideal system generally does not require the user to touch anything. Gesture recognition has a lot of potential for making live experiences that are interactive and fun. The term "gesture" can be used to describe any nonverbal communication intended to convey a particular message in its broadest sense. A gesture is any large or small physical movement that can be interpreted by a motion sensor in the field of gesture recognition.

2 Literature Survey

Oudah, Ali al naji and Javaan chahl [1] presented a hand gestures are a form of body language that can be expressed through the position of the fingers, the center of the palm, and the shape the hand forms. There are two types of hand gestures: static and dynamic. The static gesture,

* Corresponding author: annapoorna1675@grietcollege.com

as its name suggests, relates to the stable shape of the hand, whereas the dynamic gesture is made up of a sequence of hand motions like waving.. The camera vision-based sensor is a widely used, appropriate, and practical approach because it makes it possible for people and computers to connect without having to make any physical contact. One can use monocular, fisheye, TOF, and IR cameras in various combinations. The computer based vision sensors are easily available. The accuracy of the IR sensor is quite remarkable.

“Design of human machine interactive system based on hand gesture recognition” by Xiaofei ji, Zhibo wang in 2019 [2]. In recent years, technology for gesture detection has drawn increasing attention. Hand gesture interaction features are currently lacking in many of the most popular human-computer interaction systems. Preprocessing, hand gesture detection and recognition, tracking, and interaction make up the bulk of the interactive system. In order to implement the function of human-computer interaction, the tracking and interaction module tracks user gestures using Kalman tracking and controls virtual hardware in accordance with the results of gesture recognition. The system has been tested in challenging conditions, demonstrating its resilience to varying illumination and intricate backgrounds. The processing performance is increased to more than 50 frames per second on average by using CPU-GPU parallel computing. The suggested system processes images and videos using open source Python tools and the Kinect camera. The parallel computing technique has a quick response time. It is resilient to varying backgrounds. The Limitation is High cost of sensors and Complex algorithms consume a lot of processing time.

“Hand gestures based on instrumented glove approach” by Poonam sonwalker, tanujha sakhare, Ashwini patil, sonal kale [3]. The location and motion of the hands can be recorded using the wearable glove-based sensors. Additionally, using sensors included into the gloves, they can quickly supply the 12 precise coordinates of the palm and finger placements, orientations, and configurations. The user cannot easily engage with the computer using this method because it necessitates a physical connection between them. Advantages are Ease of interaction between user, computer and The glove haptic feedback is accurate. The issue is the price of these devices is quite high.

To keep a line of communication open with the rest of the population, the Hand Gesture Recognition and Voice Conversion (HGRVC) technology locates and tracks the hand motions of the dumb and deaf [4]. Using a web camera, hand gesture detection is possible. The images are subsequently processed by pre-processing to become standard size. The goal of this project is to upload the images to the database, which will match the images, and turn them into text. The detection process involves watching hands move. The method produces text-based output that aids in bridging the communication gap between the deaf-mute and the general public. The predefined images in the database helps the process of recognition of the gesture easier. It is cost-effective. The problem is it relies on too many machine learning algorithms and Longer processing time.

Ahmed Skaik, Mohamed Alsheakhali, Mohammed Aldahdouh, Mahmoud Alhelou [5] Using a hand gesture detection system, you can identify dynamic movements when one gesture is made against a complicated background. This gesture recognition system doesn't use any markers or instrumented gloves, in contrast to earlier systems. The newly proposed barehanded method only accepts 2D video input. This method entails locating the hand, following the path of the moving hand, and analysing changes in hand position. Advantages are Low cost of equipment and Less use of instruments. The issue is Lot of dynamic movements cannot be recognized.

SH jeng, CL huang M Zhou and Z Tian [6] :With the help of a glove marked with various colors, a camera is used in this technique to follow the hand's movements. In order to interact with 3D objects, this technique has been utilized. It allows for some processing, such as zooming, moving, sketching, and writing using a virtual keyboard with good flexibility. Because of the colors on the glove, the camera sensor can follow and locate the

palm and fingers, allowing for the extraction of a geometric model of the hand's shape. It is Simple to use and low price compared to sensor data glove, and the limitation is less accuracy, It limits the ability to react spontaneously.

3 Proposed method

3.1 Objectives

- To reduce the problem of miscommunication between the normal people and the speech impaired people.
- To develop a user friendly reliable automatic gesture recognition system that recognises the hand gestures made by the speech impaired people.
- To use the Flex sensors to detect the Hand gesture made and the ESP32 microcontroller to convert the gesture to the appropriate text.
- To use the OLED sensor to display the name of the gesture made by the user.
- To use the Kodular Companion application to speak out the gesture made by the user.
- To use the Google Firebase Database to update the gesture values from time to time for the required results.
- To convey the message produced by the speech impaired person and convey it to the normal person in a simple and unambiguous manner.

3.2 Architecture diagram

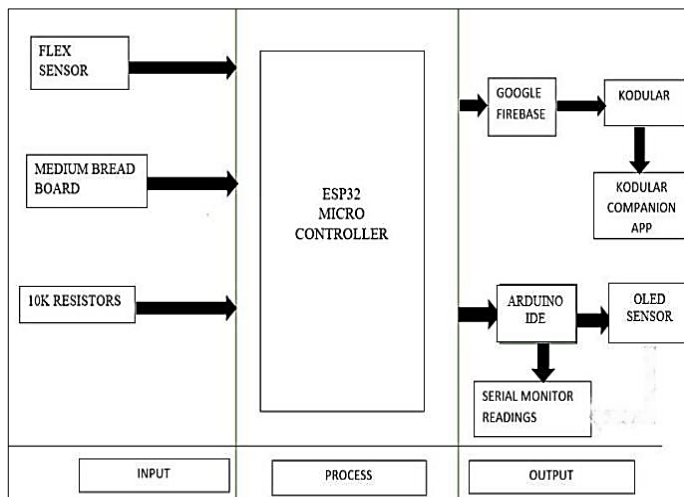


Fig. 1 Architecture diagram.

3.3 Modules and description

3.3.1 Module 1: Hardware setup

In this module the entire hardware setup is described and it tells us about the connections between the various components of the project. We connect the components as follows:

- The first connection is connecting the ESP32 microcontroller to the medium bread board by inserting the microcontroller into it carefully.
- The next connection is the connections for the ground(GND) and the VCC(3.3V), they are soldered and connected to the microcontroller and the bread board.
- The next connection is the connection for the 5 flex sensors that are used for the hand gesture recognition they are connected to the microcontroller through the following pins: VN, D34 , D35, D33, D32,
- These 5 flex sensors are again soldered to the connecting wires carefully through soldering. The resistors are also connected in series according to the flex sensors. The OLED sensor is finally soldered to the following pins: D21, D22
- The flex sensors are attached to the glove through simple tape or rubber bands and the medium bread board is stuck on the fist.

3.3.2 Module 2: Hand gesture recognition

In this module, after both the Hardware and the Software are connected we start the actual implementation of the project It starts by including and downloading the libraries required for the project. These are the following libraries that have to be installed from the Library manager in the Arduino IDE. The libraries are as follows:

- Adafruit BUSIO by Adafruit(Version 1.14.1)
- ESP8266 and ESP32 OLED Driver for SSD1306 displays by ThingPulse, Fabrice Weinberg(Version 4.3.0)
- Firebase Arduino Client Library for ESP8266 and ESP32 by Mobizt(Version 4.2.6)
- Firebase ESP32 Client by Mobizt
- Firmware zip file from the internet.

3.3.3 Module 3: Connecting to Firebase

In this module the Arduino IDE is connected to the Google firebase. This is a simple step but first we need to create a firebase and set it up for our project. Follow the following steps to create a firebase for the project:

Step 1: Enter Google Firebase in the Google Search Engine. Step 2: Click on the first Google search result.

Step 3: Click on Get Started.

Step 4: Click on the New Project and type a project name.

Step 5: Click on Run in test mode and click Done.

Step 6: The firebase is created and ready to use.

Then to connect our project to firebase we need to include the firebase credentials in our code. For doing so we need to include the firebase URL and its secret key in our code. To do so we need to follow the below steps:

Step 1: Click on the Settings icon

Step 2 : Click on Project Settings

Step 3; Click on Service Accounts

Step 4: Click on Database secrets

After this both the Firebase URL and the secret key are found below. Copy them and paste these database credentials in the code.

3.3.4 Module 4: Kodular app creation

In this module, the kodular application is introduced which is used for the speech production part in the project. The kodular also has a companion known as the kodular companion application which should be installed in the mobile phone. A free, user-friendly web toolkit for developing mobile apps is called Kodular. It primarily offers an online drag-and-drop interface for creating Android applications that anyone can use to develop any kind of application without writing a single code line. The kodular application has 2 different pages namely designer and blocks. The designer page is the pagewhere we drag and drop components into the Android phone and the blocks page is the page wherewe build the building blocks of the application. First a project is created in kodular. When the website is opened, a project name is given and the new project is ready. For the project the designerpage consists of 3 components namely:

- Clock
- Firebase_Database
- Text to Speech

In this component the time interval is set as 3000, which means that the time difference between each speech production is 3 seconds. The time enabled and the time always fires are also enabled in the properties section. In this component we connect both the firebase and the kodular by including both Firebase Token and the Firebase URL. Once these attributes are filled with the right credentials both the firebase and the kodular and connected. Both the Firebase URL and the Firebase Token are found in the service accounts section in the project settings of the Google Firebase Project. The text-to-speech component is the component that actually converts our text to the required speech. This component's properties are set as follows:

- Country: USA
- Language: English
- Pitch rate: 1.0
- Speechrate:1.0

4 Experimental results

4.1 Serial monitor output

The Serial Monitor displays the flex sensor values that are caused due to the change in resistance of the flex sensor. The main principle of the gesture recognition is change in resistance and thus the change in resistance values are the values that are visible in the serial monitor. The Serial Monitor is available on the top right corner of the Arduino IDE and when the code is uploaded into the ESP32 microcontroller, the Serial Monitor icon is clicked and the display box appears showing us the flex values for the five fingers and the text to be displayed in the firebase and the OLED sensor. All the flex sensors give us the required values that range above 2500 for an upright finger and above 3000 value for a completely bent finger. Thus the precoded gestures constitute of the gestures starrng from values 2500 to 3000 only and they have been put in a sequence such that the gesture with the maximum values of 3000 have been put first and the gestures with the minimum number of values of 3000 or the maximum values of 2500 have been put last in the code. The flex values in the serial monitor are observed constantly changing with minimal difference that is because of the main principle of the project which is change in resistance. In the above image the observation shows the serial monitor readings which contain the name of the gesture and the five flex sensor values which vary for every second. This is due to the principle of the project

which is change in resistance. In the above instance the text recognized is “Hii” and the change in the values is observed every second but the gesture name remains the same until we change it to another gesture. In this gesture all the fingers are upright and thus this represents the gesture as “Hii”. The second observation is observed with the text of the gesture displayed as “Four” and the flex sensor values are varied from values greater than 2500 to values greater than 3000. The keen observation made here is that when the finger is bent the value goes beyond 3000 and when the finger is upright the value is greater than 2500.

4.2 Realtime database

In the Realtime database we observe the change in the text according to the change in the gesture and it is observed in such a way that a “Key” is created below the URL in the realtime database and thus key consists of the appropriate text and when the text is updated the section is highlighted and it changes the text upon the change in the gesture. The basic outline is like Key: “Value” where the value is replaced with the text that our gesture is recognised.

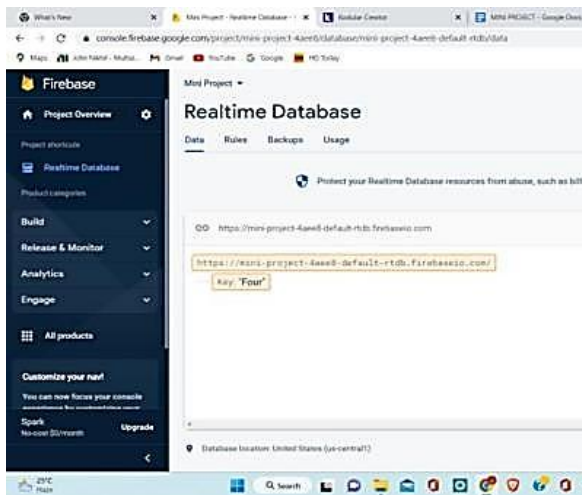


Fig. 2 Realtime database output.

4.3 OLED sensor

The OLED sensor displays the text output in the sensor. The OLED sensor is soldered to the ESP32 microcontroller. The OLEDUpdate() function in the code updates the text in the sensor and the OLED.Init() function is used to display the text in the sensor in the first place. Through both these functions the output is observed in the form of text.

4.4 Kodular companion

Kodular Companion is a mobile application used to test the output produced by Kodular. This is used for the speech production part in the project. The text is converted to speech in the Kodular based on the app built in the blocks section and thus when we include the firebase credentials both the firebase and the kodular are connected and the speech is produced from the Kodular Companion application. The speech aspect is pretty much simple as it is just a voice produced from the application that pronounces the gesture recognized text. The process

for connecting both the Kodular Companion application and Kodular. It is done by the following steps:

Step 1: Click on Test on the designer page. Click on Connect to Companion.

Step 2: Scan the QR code or type the secret code, both these components are now connected and ready to test.

4.5 Entire model results

The entire model results represents the overall result observed from after the completion of the project. From a direct view, the observation made is that both the expected output and the actual output do match and the objectives of the project are met by the prototype model. The entire setup produces output from all the three modules which include Kodular Companion output, Google Firebase output and the OLED sensor output. The below images are the screenshots of the entire model representing the output from all the modules that fulfill the objectives/requirements of the project.

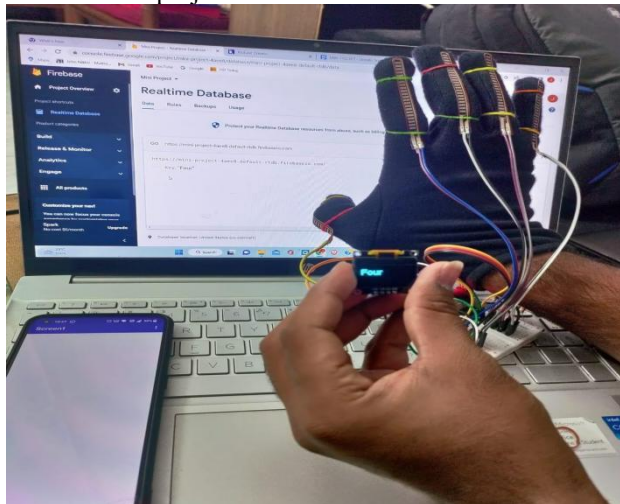


Fig. 3. Components output.

5 Conclusion

The project mainly emphasizes on the speech impaired people having the disability of miscommunication between them and the normal people. It mainly focuses on gesture recognition as that is the main means of communication for the speech impaired people. They communicate through gestures and these gestures are only understood by other speech impaired people and thus it cannot be understood by normal people and this is where the project comes into picture. The gesture made by the hand movement of the glove causes a change in the resistance around the flex sensors attached to the glove and the ESP32 where these are soldered to the microcontroller. The microcontroller is fitted onto the medium bread board and thus the resistors are also fixed in the bread board to calculate the change in the values. In the ESP32 microcontroller the gesture is converted to text and the text is displayed in the Realtime firebase and the text is also displayed in the OLED sensor that is also soldered to the microcontroller. Through the help of the app built by the Kodular we connect the Kodular companion application from the mobile phone and we obtain the speech production from that application .This enables any user to understand the gesture made, visualizes the

text in both the firebase and the OLED sensor and thus hears the pronunciation of the text from the Kodular Companion application.

References

1. M. Oudah, A. A. Naji, Javaan Chah, J. Imaging. **6**, 8, (2020)
2. X. Ji, Z. Wang and X. Zhang, *Design of Human Machine Interactive System based on Hand Gesture recognition*, in the Proceedings of the 6th International Conference on Information Science and Control Engineering (ICISCE2019), (2019)
3. P. Sonwalker, T. Sakhare, A. Patil, S. Kale, Int. J. of Eng. Tren. and Tech. **6**, 9, (2019)
4. M. Alsheakhali, *Hand gesture recognition system*, in the Conference Proceedings of International Conference on Information and Communications Security (ICICS2011), 2011
5. X. Ji, Z. Wang and X. Zhang, *Design of Human Machine Interactive System based on Hand Gesture recognition*, in the Proceedings of the 6th International Conference on Information Science and Control Engineering (ICISCE2019), Shanghai, China, (2019)