Smart Gloves for Hand Gesture Recognition

Sign Language to Speech Conversion System

Abhijith Bhaskaran K, Anoop G Nair, Deepak Ram K, Krishnan Ananthanarayanan, H R Nandi Vardhan
Department of Electronics and Communication Engineering
Amrita School of Engineering, Bengaluru
Amrita Vishwa Vidyapeetham
Amrita University
India
abhijithbhaskarank@gmail.com, anup21195@gmail.com, deepak7946@gmail.com, akrishnan394@gmail.com,
hr_nandivardhan@blr.amrita.edu

***Abstract*—People with speech impairment find it difficult to communicate in a society where most of the people do not understand sign language. The idea proposed in this paper is a smart glove which can convert sign language to speech output. The glove is embedded with flex sensors and an Inertial Measurement Unit (IMU) to recognize the gesture. A novel method of State Estimation has been developed to track the motion of hand in three dimensional spaces. The prototype was tested for its feasibility in converting Indian Sign Language to voice output. Though the glove is intended for sign language to speech conversion, it is a multipurpose glove and finds its applications in gaming, robotics and medical field.**

Keyw ords*—*Glove, Indian Sign language, Flex Sensors, Inertial Measurement Unit, State Estimation Method, Three Dimensional Space, gaming, robotics, medical field

I. Introduction

Indian Census in 2011 estimated that over 26 million people suffer from some kind of disability that is equivalent to 2.1% of the total population. The most common types of disabilities are disability in movement (that accounts to 20.9%), hearing impairment (19.5%), and speech impairment (7.68%) [1]. People with speech impairment use sign language to communicate with the society. The conventional idea for gesture recognition is to use a camera based system to track the hand gestures. The camera based system is comparatively less user friendly as it would be difficult to carry around. In addition, it would not be feasible to use it in crowded areas as it would pick up multiple gestures from different people who are in its viewing angle. Separation of such unwanted gestures is tedious and not a viable solution.

The sign language varies from region to region. The current prototype is concentrated on Indian Sign Language (ISL). There are variations within Indian Sign Languages, some of which are;

- Mumbai–Delhi Sign Language, the most influential
- Calcutta Sign Language
- Bangalore–Madras Sign Language (or Bangalore–Chennai–Hyderabad Sign Language)

The initial focus had restriction of the region, however the focus was broadened as the concept was refined to be more versatile with the addition of reconfigurable features. This paper discusses the novel concept of glove based system to address the concerns of camera based system and also promises to be more portable. The glove is used to recognize the gesture and convert it to speech which can be played out through the speakers.

II. Related Work

Few attempts have been made in the past to recognize the gestures made using hands but with limitations of recognition rate and time which includes:

1. Using CMOS camera
2. Leaf switches based glove
3. Copper plate based glove

In CMOS camera based glove, a CMOS camera transmits image data via UART serial port. The UART performs serial-to-parallel conversions on data received from a peripheral device (CMOS camera in this case) and parallel-to-serial conversion on data received from the CPU.

Disadvantage: Highly expensive, and high latency

Leaf switch based glove are similar to normal switches but these are designed in such a way that when pressure is applied on the switch, the two ends come into contact and the switch will be closed. The leaf switches are placed on the fingers of the glove such that the two terminals of the switch come into contact when the finger is bent.

Disadvantage: With increase in usage, the switch instead of being open when the finger is straight, it will be closed resulting in improper transmission of gesture

In copper plate based glove, a copper plate is fixed on the palm as ground. The copper strips indicate a voltage level of logic 1 in rest position. But when copper strips come in contact with the ground plate, the voltage associated with them is dropped and they indicate a voltage level of logic 0.

Disadvantage: The use of copper plate makes the glove bulky which makes it uncomfortable to use it for a longer time [3].

III. BACKGROUND

Much of our research has been focused on utilizing flex sensors to track the orientation of fingers and IMU to track the motion of hand in 3-D space. Hand-talk Gloves proposed by Ambika Gujrati et.al, introduces the idea of using flex sensors embedded on a glove to track the orientation of fingers. Flex sensors are variable resistors, whose resistance depends on its radius of curvature. The output of the sensor can be digitalized and processed by a microcontroller to voice out the word corresponding to the gesture [2].

To track the motion of hand in 3-D space an Inertial Measurement Unit is used. Attaching an IMU on the human limb allows us to track its movement in any random direction. The position can be found by double integrating the acceleration values obtained from the accelerometer [4]. However, they add noise and induce drift errors. Due to the drift, positional values diverge within the small intervals of time. Hence, it has to be compensated and accuracy of positional values has to be improved. Another method to track the movement of hand, using the IMU, is to find the angular coordinates (pitch, roll and yaw). Pitch and roll can be calculated using the accelerometer data and is susceptible to noise. Pitch and roll can also be calculated by integrating the gyroscope values. To minimize the error induced due to integration and to complement each other's weaknesses, a suitable sensor fusion algorithm can be used. Kalman Filter can be designed to obtain stable values of pitch and roll by combining the accelerometer and gyroscope data. A magnetometer can be used to obtain the yaw (or heading). When a magnetometer is tilted it introduces errors in the reading. A tilt compensation technique can be implemented to avoid the errors in heading due to the tilt [5].

The fusion of sensor and gyroscope data can be achieved by using filters. There are different filters that can be such as Kalman filter and Complementary filter. However, the implantation of the Kalman is complex and hence complementary filters can be used as it would give agreeing values and is simple to implement. Pitch and roll can be found out from accelerometer data using inverse trigonometric functions. To eliminate the effect of external noise in accelerometer readings, the values are passed through a low pass filter. To eliminate the drift error in integration of gyroscope readings, the integrated values are passed through a high pass filter. A stable output can be obtained by combining the output of both low pass and high pass filters [6].

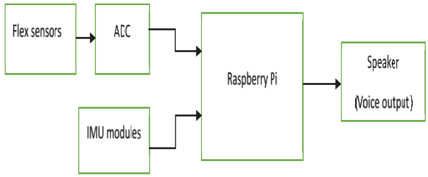


Fig. 1. *Hardware Block Diagram*

IV. METHODOLOGY

A. Process Flow

The gloves are embedded with five flex sensors to track finger orientation and an Inertial Measurement Unit (IMU) to track hand movement in 3-Dimensional space. 'One handed gestures' is the area of focus of the current design. Raspberry Pi is used as the processing unit. A systematic representation of the system is shown in Fig. 1.

The values obtained from flex sensor and IMU module are used to identify the position of hand in three dimensional spaces. The algorithm is divided into two sections viz., finger orientation detection and 3- D orientation detection. The flow chart of algorithm is as shown in Fig. 2.

As shown in Fig. 2, once a start has been initiated by the user, the Raspberry pi takes input from flex sensor and IMU module. Using novel method of state estimation as explained in section IV.C, it determines the motion path of the hand. After calculating the path and reading values from the voltage divider, for each finger, the program searches its database for a match. If it finds a matching pattern the text corresponding to the pattern is fetched from the database and loaded into the TTS engine. The TTS engine voices out the text. If the program couldn't find a matching pattern then it continues to the next iteration.

The input from flex sensors is fed to Raspberry Pi through the ADC. Below example shows detection of finger position using the method of range.

Example:

if (analog_value> θ_1):
range="1"

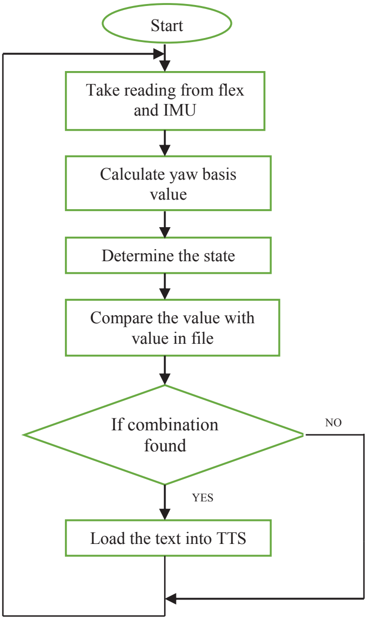
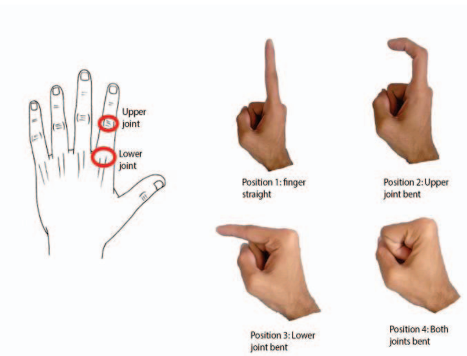


Fig. 2. *Flow chart of algorithm*

Where, *analog_value* represents the analog equivalent of the digital value from the ADC, θ_1 represents the reference voltage for range classification. 3.3v is given as the input to the voltage divider. The analog value is used to classify the finger gestures into one of the four ranges which correspond to the four possible orientations of a finger (Fig.3).



B.

To detect the finger orientation flex sensors are embedded behind the finger. As the finger is bent there will be a corresponding change in radius curvature of the flex sensor, which in turn changes its resistance. To convert the change in resistance to the corresponding change in voltage, a voltage divider circuit is utilized. The voltage values obtained from the voltage divider circuit is fed to MCP 3008 (ADC), which converts the analog voltage value to a 10 bit digital value. The digital value is given as input to Raspberry Pi for processing.

There are four gestures that can be made with a finger, as shown in Fig. 2. One 2.2” flex resistor is used for the thumb and 4.5” flex resistors are used for the other fingers. The fixed resistor values are calibrated such that the voltage will remain in the same range for a particular orientation irrespective of the finger (range of voltage varies for each position). The voltage range was found out by trial and error method. To distinguish between position 2 and position 3 the placement of flex resistors has to be precise. The placement is in such a way that for positions 2 and 3 the radius of curvature of the flex sensor is different (radius of curvature of the flex sensor for position 3 is more than that for position 2). This results in different voltage ranges for position 2 and 3.

C. *Mathematical Model of Flex Sensor*

Equation 2 represents a generic equation of resistance of a material with respect to it physical properties (resistivity, length and area).

$$\mathrm{R_{total}} = \rho * \frac{L}{A} \quad (1)$$

$$\mathrm{R_{total}} = \rho * \frac{L}{wt}$$

$$\mathrm{R_{total}} = \frac{\rho}{t} * \frac{L}{w}$$

$$\mathrm{R_{total}} = \mathrm{R_{sheet}} * \frac{L}{w} \quad (2)$$

Where, $\mathrm{R_{sheet}} = \frac{\rho}{t}$

The resistance for infinitesimally small segment of the strip

$$\Delta\mathrm{R} = \mathrm{R_{sheet}} * \frac{\delta x}{w} \quad (3)$$

Where, δx is the length of small strip.

Therefore,

$$\mathrm{R_{total}} = \mathrm{R_{sheet}} * \int_0^L \frac{\delta x}{w}$$

Where, L is the total length of the strip, when the flex resistor is not flexed

$$\mathrm{R_0} = \mathrm{R_{sheet}} * \left.\frac{x}{w}\right|_0^L$$

$$\mathrm{R_0} = \mathrm{R_{sheet}} * \frac{L}{w} \quad (4)$$

On flexing the sensor, the length of the flex resistor increases by ΔL. So the new resistance will be as follows.

$$R_{total} = R_{sheet} * \int_0^{L+\Delta L} \frac{\delta x}{w}$$
$$= R_{sheet} * \frac{x}{w}\Big|_0^{L+\Delta L}$$
$$= R_{sheet} * \frac{L+\Delta L}{w}$$
$$= R_{sheet} * \frac{L}{w} + R_{sheet} * \frac{\Delta L}{w}$$
$$\boldsymbol{R_{total} = R_0 + R_{sheet} * \frac{\Delta L}{w}} \quad (5)$$

Where, R_0 is the total resistance of the flex resistors when it is not flexed.

According to Ohm's law, at node V_{out} (Voltage measured across fixed resistor R)

$$\Delta V_{out} = \Delta I * R$$
$$\Delta I = \frac{V_{in}}{\Delta R_f + R} \quad (6)$$

Where, ΔR_f is the change in resistance value of flex resistor when it is flexed.

Therefore a change in the resistance of the circuit brings about a corresponding change in the output voltage V_{out} (voltage across the fixed resistor R, as in Figure 4.5).

$$\boldsymbol{\Delta V_{out} = \frac{V_{in} * R}{\Delta R_f + R}} \quad (7)$$

D. 3-D Orientation Detection

To determine the position of hand in 3 D space angular coordinates (pitch, roll and yaw) are used, which is calculated from the data obtained from the IMU module. A representation of pitch, roll and yaw is depicted in Fig. 4.

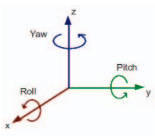


Fig. 4. *Diagrammatic representation of pitch, roll and yaw*

The pitch and roll are obtained using the equations given below

$$Pitch = \sin^{-1}\left(\frac{A_x}{\sqrt{Ax^2+Ay^2+Az^2}}\right) \quad (8)$$

$$Roll = \sin^{-1}\left(\frac{A_y}{\sqrt{Ax^2+Ay^2+Az^2}}\right) \quad (9)$$

The absolute value of pitch and roll can be used to determine pitch and roll of hand, however yaw value is dependent on the earth's orientation with magnetic field. Since the Yaw is highly dependent on the orientation with earth's magnetic field, when the user turns to face a different direction the yaw basis value has to be recalibrated.

The forces working on the object can be measured with an accelerometer. Even a small force working on the object will cause disturbance in the accelerometer reading. When working on an actuated system, the sensor measures all the forces that drive it, hence accelerometers are proven to be reliable only on the long term. So to address this issue a low pass filtering has to be done. The measurement obtained from gyroscope is not susceptible to external forces, however integration over time, induces a shift in the measurement i.e. zero is not reached even when system moves back to its original position. The gyroscope data is reliable only on the short term, as it starts to drift on the long term. To correct the error filters can be used. Commonly used filters are Kalman filter and Complimentary filter. We have used the complementary filter in our prototype.

Complementary filter is used to combine the pitch and roll values obtained from accelerometer and gyroscope [5]. The equations for obtaining pitch and roll are as follows.

$$Pitch = AA * (Pitch + gyro_x rate * LP) + (1 - AA) * Pitch^* \quad (10)$$

$$Roll = AA * (Roll + gyro_y rate * LP) + (1 - AA) * Roll^* \quad (11)$$

AA is the filter coefficient, *LP* is the loop time of the program, $gyro_x rate$ and $gyro_y rate$ are rate of change of angle (gyroscope data) along x-axis and y-axis respectively, $Pitch^*$ and $Roll^*$ are the angular coordinates calculated using accelerometer. The loop time has to be precise, because an error in the loop time will result in an error in the integration of gyroscope value.

The gyroscope data is integrated every time-slice with the present angle parameter. After this it is combined with the low-pass data from the accelerometer. The filter constants has to sum up to 1, so 0.8 and 0.2 were chosen but it can be changed for better tuning.

The value of yaw is found using magnetometer with respect to the horizontal plane. Hence when IMU module is tilted the heading or yaw value becomes erroneous. Tilt compensation technique has been used to avoid this error. Tilt compensation process maps the magnetometer data to the horizontal plane and provides the accurate heading calculations regardless of the position of the magnetometer [4]. The roll and pitch are utilized to correct the tilt error. Therefore, the equations for calculating yaw values are as follows.

$$MAG_x = m_x cos(Pitch) + m_z sin(Pitch) \quad (12)$$

$$MAG_y = m_y cos(Roll) + m_x sin(Roll) sin(Pitch) - m_z sin(Roll) cos(Pitch) \quad (13)$$

$$Yaw = tan (MAG_y / MAG_x) - m_z sin(Roll) cos(Pitch) \quad (14)$$

Where, m_x, m_y and m_z are the magnetometer values along x, y and z axis respectively, MAG_X and MAG_Y are the corrected magnetometer values along X and Y axis.

E. State Estimation Algorithm

A novel algorithm of state estimation has been used to identify the orientation of hand in three dimensional space. A state is defined by a set of pitch roll and yaw values and is represented by a number. The work area (area around the body where the hand can move) has been divided into two states. The work space constraints for a human hand is depicted in Fig. 5. In the horizontal plane the hand can reach up to a maximum deviation of 220° (the value was found out by experimental method). In the vertical plane the hand can reach up to a maximum deviation of 180° (the value was found out by experimental method).

In the State Estimation method the state is defined based on the position of hand within this restricted workspace. A two dimensional representation of states in State Estimation Method is shown in Figure 6.

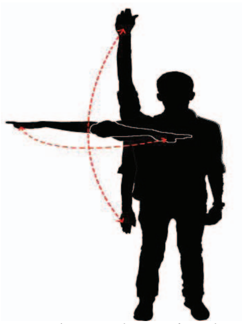


Fig. 5. *Workspace of Hand*

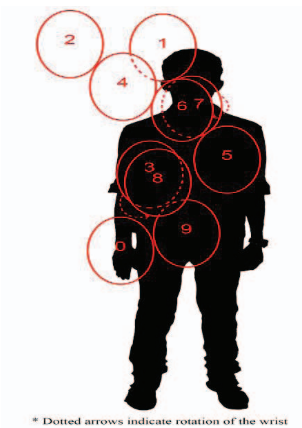


Fig. 6. *Two -dimensional representation of states*

The prototype utilized three angular coordinates to define states in a three dimensional space. The area occupied by each state may vary, which determines the tolerance for that particular position.

In the program, four consecutive states have been recorded. For a gesture in which the hand is not in motion (static gestures) all the states will show same value as the hand is in the same state for a specific amount of time. For a gesture in which the hand is in motion (dynamic gestures) the hand will pass though different states. The state pattern information is utilized by the processor to make an estimate of the motion path/gesture.

Example for a static gesture:

TABLE I. SAMPLE PATTERN FOR STATIC GESTURE

State 1 (current state) (S_n)	State 2 (S_{n-1})	State 3 (S_{n-2})	State 4 (S_{n-3})
0	0	0	0

Example for a dynamic gesture:

TABLE II. SAMPLE PATTERN FOR DYNAMIC GESTURE

State 1 (current state) (S_n)	State 2 (S_{n-1})	State 3 (S_{n-2})	State 4 (S_{n-3})
2	2	1	1

Table I suggests that the user is making a gesture in which the hand in stationary at state 0. Table II suggests that the user has made a gesture in which his/her hand moves from state 1 to state 2. The processor searches its database to figure out the action/word corresponding to this pattern.

So based on the states, that the system (i.e. the hand) passes through, the corresponding gesture is detected and voiced out through the speaker. This state pattern detection algorithm can be further enhanced to predict the next state based on the previous states. More the number of states, better will be accuracy of prediction.

V. RESULTS

TABLE III. VOLTAGE VALUES CORRESPONDING TO EACH FINGER POSITION

Voltage Range	**Finger Position**
> 39000	1
< 39000 and > 37500	2
< 37500 and > 35500	3
< 35500	4

State Pattern [S_n, S_{n-1}, S_{n-2}, S_{n-3}]	**Finger positions [f_1, f_2, f_3, f_4, f_5]**	**Word**
3, 3, 3, 3	4, 4, 4, 4, 1	Good
5, 5, 5, 5	1, 1, 1, 1, 1	Please
4, 4, 2, 2	1, 1, 1, 1, 1	Bye
8, 8, 7, 7	1, 1, 1, 1, 2	Thank You

TABLE IV. STATE PATTERNS AND CORRESPONDING WORD

Table III shows the voltage values corresponding to each finger position. Table IV shows the gestures and corresponding words voiced out. In Table IV f_1, f_2, f_3, f_4 and f_5 represents the little finger, the ring finger, the middle finger, the index finger and thumb respectively. S_n to S_{n-3} represents the four consecutive states recorded, where S_n is the current state.

The human motion of a hand in 3D space can potentially be tracked and recorded using flex sensors complemented with IMU modules. The flex sensor distinguishes different orientations of fingers and IMU modules, one placed on the fore arm is capable of tracking the movement of the arm. This technology will aid the speech and hearing impaired people to communicate with the external world by voicing out the gestures.

This system could be further be enhanced by placing one more Berry IMU on the upper arm. This would make the whole system more dynamic and robust by increasing the range of gestures that can be detected. Also, the current algorithm for yaw basis correction is based on an assumption that the person keeps his hand close to his body when he changes the direction he faces, which need not be the case practically. Hence, we have defined a method using four IMU modules (two on each hand, when the system develops further for two hands) to tackle this issue.

There is also scope for improvement in the system by incorporating contact sensors. Strategic placement of the contact sensors on the glove, can enhance the gesture recognition capability of the system.

VI. CONCLUSION

Our work till date has involved five flex sensors, an IMU module, a Raspberry Pi, a voltage divider circuit and an ADC. The ADC with voltage divider interfaces the Raspberry Pi to the flex sensors. The flex sensors placed on each fingers enable us to identify the orientation of the finger. The Berry IMU placed on the fore arm tracks the movement of the hand in the 3D space. The values from the accelerometer and the flex sensors are compared with our database of values to detect the gesture. A few more applications where the gloves could be used in the future are listed below:

Gaming industry: Hand gestures play a vital role in the gaming industry, especially in first person shooting games. The player can control the character in the game using his hand and this could give a real life experience of the game. Also, virtual reality is gaining grounds in the gaming industry. Combining virtual reality with the gloves with a haptic feedback can give the gamer a real life gaming experience.

Controlling a robotic arm using the gloves: The gloves could be used to control a robotic arm. The applications for this system are wide. With incorporation of haptic feedback, the glove-robotic arm interface could also be used in bomb diffusion.

Remote medical surgery: In this the surgeon need not be at the physical location to perform the surgery. He could control a robotic arm remotely to perform the surgery and a haptic feedback could give him the feel of actually performing the surgery. But this would require the gesture recognition to be very precise and the transfer of data from the hand to the robotic arm should be without even a tiny glitch.

Helping hand for people with Cerebral Palsy: Cerebral palsy usually appears in early childhood and involves a group of permanent movement disorders. The symptoms vary with people and often includes poor coordination, stiff muscles, weak muscles, and tremors. Also, the problems with sensation, vision, and hearing, swallowing and speaking have been identified as other symptoms. This problem can be solved to a great extent by providing them with a provision to communicate with just a single finger. The frequently used words by such people can be put across to people with just a small movement in the finger and using our state estimation technique it could predict the letters / words.

REFERENCES

[1] "Census of India 2011," http://www.censusindia.gov.in/2011census/population_enumeration.html .

[2] Kanika Rastogi, Pankaj Bhardwaj "A Review Paper on Smart Glove - Converts Gestures into Speech and Text" *International Journal on Recent and Innovation Trends in Computing and Communication*, *Volume: 4 Issue: 5 ISSN*: 2321-8169, pp. 92 - 94

[3] Ambika Gujrati, Kartigya Singh, Khushboo, Lovika Sora and Mrs. Ambikapathy, "Hand-talk Gloves with flex Sensors: A Review," *International Journal of Engineering Science Invention ISSN (Online),* vol. 2, no. 4, pp. 43-46, 2013.

[4] Xiaoping Yun, Eric R. Bachmann, Hyatt Moore IV and James Calusdian, "Self-Contained Position Tracking of Human Movement Using Small Inertial/Magnetic Sensor Modules," in *IEEE International Conference on Robotics and Automation*, Roma, Italy, 2007.

[5] Fatemeh Abyarjoo, Armando Barreto , Jonathan Cofino and Francisco R., "Implementing a Sensor Fusion Algorithm for 3D Orientation Detection with Inertial/Magnetic Sensors," in *Innovations and Advances in Computing, Informatics, System Sciences, Networking and Engineering*, Springer International Publishing, 2014, pp. 305-310.

[6] Rohan Urdhwareshe, Md. Zaki Bakshi, Pranav Naiknavare and Sumit Naik, "Design and Implementation of IMU Sensor Fusion and PID Control in Quadrotor," *IPASJ International Journal of Electronics and Communication (IIJEC),* vol. 2, no. 9, 2014.