International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

Hand Gesture to Speech and Text Conversion Device

K. P. Vijayakumar, Ananthu Nair, Nishant Tomar

Abstract: Sign language and facial expressions are the major means of communication for the speech flawed people. General people can understand the facial expression to an extent but cannot understand the sign language. Dumb people are unable to express their thoughts to normal humans. To reduce this gap of communication, this paper presents an electronic system which will help the mute people to exchange their ideas with the normal person in emergency situations. The system consists of a glove that can be worn by the subject which will convert the hand gestures to speech and text. The message displayed will also help deaf people to understand their thoughts. This prototype involves raspberry pi 3 as the micro-controller along with the flex sensors, accelerometer sensor. The resistance of the flex sensor changes due to the bending moment of the fingers of the subject. The accelerometer measures the angular displacement of the wrist along the y-axis. The microcontroller takes the input from the two sensors and matches it with the pre-programmed values and plays the respective message. The system makes use of python and its libraries for microcontroller programming.

Keywords: Flex sensor, Accelerometer, pyttsx3, Raspberry Pi, Analog to Digital Converter (ADC).

I. INTRODUCTION

A huge portion of the global population has the inability of speaking either partially or completely. In India, around 2.78 percent of the total population are speech flawed [1] and a very small fraction of these are good in communicating with hand gestures. The possibility for a normal person knowing the sign language is very less. So to reduce the communication gap, the research in the field of gesture to speech (G2S) system becomes more important. In recent years many researchers focused on hand gesture detection and developed many techniques in the field of robotics and artificial intelligence [2]. This project uses a similar approach but tries to implement the idea distinctly and came up with an important application in the domain of IoT. The device helps a dumb person to communicate with the normal person as well as a deaf person. Various methods implemented for the gesture to speech conversion by researchers all over the world. The motivating factors of the paper come from the idea that (i) a system that can interpret many messages by using the minimum number of sensors thus making the system less complex to use. (ii) This paper provides a method for designing a faster system using sensors. (iii) The system should be free from thermal injuries and must be shockproof. To fill the communication gap between the normal people, dumb people, and deaf people, a system is designed to convert hand gestures into the audio message as well as a text message. The objective of the proposed system is to interpret many messages using less number of sensors. Thus the system will become lightweight and faster. Another main objective is to provide two types of output, text and audio output, to help the dumb people to communicate even with a deaf person. The organization of the paper is as follows: Section 2 gives the discussion of all the techniques and algorithms that have been used in existing systems. The system modules such as flex sensors and accelerometer are thoroughly discussed in section 3. The results and discussions are done in section 4 which includes the parameters which define and test the quality and accuracy of the system. Finally, the conclusion, advantages, and future works are discussed in section 5.

II. RELATED WORK

The system [3] mainly consists of an Arduino microcontroller and flex sensors. Flex sensors are used for sensing the gestures. The output of the flex Sensor is processed by the Arduino. The output from the microcontroller is then transmitted via a Bluetooth module. An android device which is connected to the device makes use of the MIT app inventor to convert gesture to speech. The system [4] reviews the sensors that are human wearable and can measure the different reflections and activities of the human body. Many features like weight and sensitivity levels are reviewed. Wearable sensors provide portability to the system. The system according to [5] gives the conversion of hand gesture to speech system which has been developed using the images of hand gestures of the dumb person (subject) captured by the camera. The images are segmented by using the skin region detection algorithm. According to this algorithm, the skin region remains white and every other part of the image becomes black based on the R/G ratio. The feature extraction technique is used to classify several types of hand gestures. The classified values are matched with the pre-recorded soundtracks in the database using MATLAB. A hand glove [6] is designed by using the flex sensor and advanced virtual RISC microcontroller which converts the analog signal as input sent by the flex sensor attached on glove fingers.

Revised Manuscript Received on March 30, 2020.
***** Correspondence Author
Dr. K P Vijayakumar, Assitant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai.
Ananthu Nair, Undergraduate Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai.
Nishant Tomar, Undergraduate Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai.

As the subject makes a gesture using the fingers, the change in resistance is sent to the microcontroller which converts the signal to an 8-bit binary code, according to which the respective messages are played using the speaker system. The system [7] converts the hand gesture to the speech system as well as displays the respective message. The system uses the 5 flex sensors, one on each finger fixed on a glove along with the Arduino-nano and speaker amplifier.
The flex sensor measures the change in resistance of the flex sensors due to the bending moment of the fingers.
These values are converted into digital parameter which is then matched with the values pre-entered in the memory. The respective voice message gets played and also the same message is displayed using the LCD screen.
The main method used in this system [8] is image processing. The camera captures the subjects' hand images. Those images are processed using different methods like color splitting and feature extraction. Every image plays the respective pre-fed sound using the hardware. This system is the vision-based method of changing gestures to the audio system.
The system [9] is embedded with flex sensors that measure the resistance across the fingers. Google text to speech library is used for text to speech conversion. This device needs to have an active internet connection for the conversion of text to speech. The system given in [10] uses Mandarin-Tibetan bilingual speech synthesizer with the support of a deep neural network model and an SVM to identify different facial expressions and hand movements that will enable the device to include emotion to output audio. This system [11] makes use of a low-cost packaging material called Velo-stat. Its electrical resistance decreases when pressure is applied to it. This technique cane utilized to measure the bending of fingers to identify the gesture. This system [12] uses an accelerometer to measure the orientation of the wrist and gyroscope to measure the angular velocity with a hidden Markov model mainly for the regional language.
The inference derived from the related work is (i) the limited number of voices/gestures due to the limited number of sensors which allow only four messages [13] to be interpreted and if more sensors are used then there are chances that the system will slow down and processing time increases gradually. (ii) Another popular technique to capture gestures is image processing [14]. The problem with image processing is that it requires additional overhead to process the image which makes the system time consuming and complex. Also, the camera has to be necessarily used for image capturing. Advanced computational resources are required to implement an efficient image processing system. Sensitivity depends on precision, various topologies like novel algorithm acquisition needs to be analyzed. (iii) Another disadvantage from the design perspective is that it has the risk of thermal injuries and also the application cost is high. (iv)There is a need for fabric data collecting glove for getting high accuracy. It includes the problem of bending sensor repeatability. It requires the full hand gestures and also the thumb rotation sensor. The sensitivity level of the sensor is low.
Therefore the challenges for the proposed system would be to design an electronic glove that should be shockproof and producing minimal thermal effects that should not seriously affect the subject. The portability of the system is a major challenge as the system should be lightweight. The main goal is to obtain the maximum messages with optimal processing time.

Table 1: Literature Review

Authors	Processing Device	Input Devices	Output (Text/Speech)	Technique/No. of Messages
H. S. Kala, et al	Arduino UNO	Flex sensor	Text	Gesture conversion/ Not Specified
S.C. Mukhopadhyay	Not Specified	ECG sensor, Accelerometer	Text	Image Processing/Not Specified
R.R. Itkarkar, et al	Personal Computer (PC)	Camera	Speech	MATLAB and Image Processing/Not specified
Ahmed, et al	AVR ATMEGA	Flex sensor	Speech	Gesture conversion/4
Safayet Ahmed, et al	Arduino NANO	Flex sensor	Text and Speech	Gesture Conversion/Not Specified
S.K. Imam Basha, et al	PC	Webcam	Text and Speech	Image Processing/Not Specified
Vaibhav Mehra, et al	Arduino UNO	Flex Sensor, Accelerometer	Text and Speech	Gesture Conversion/Not Specified
Yang H., et al	PC	Camera	Speech	SVM and Neural Network/ Not Specified
Preetham C., et al	Not Specified	Velo-stat	Not Specified	Sensing Fabric/ Not Specified
Aiswarya V., et al	Raspberry pi	Full Body Image	Speech	Image Processing/Not Specified

III. METHODOLOGY

The system comprises of a method to convert the hand gestures to text and audio messages. The prototype built has modules like raspberry pi 3 microcontroller, bend sensitive flex sensor, analog to digital converter and accelerometer. The architecture diagram is shown in fig. 1. Working of different modules of the system is discussed in further subsections

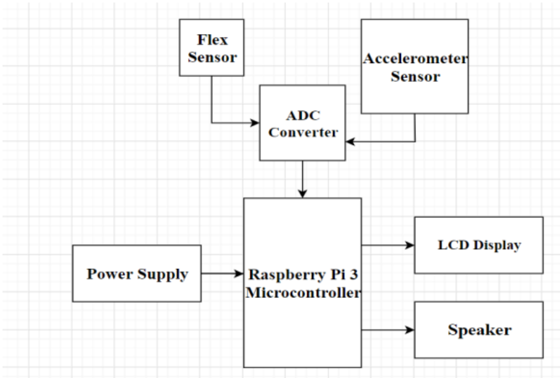

Fig 1: Architecture Diagram

A. Flex sensor

Flex sensors are bend sensitive sensors that measure the electrical resistance due to bending as shown in fig.2.
The bending is directly proportional to the bending value.
These sensors are in the form of strip ranging from 1 inch to 5 inches long. Their resistance value varies from 10 KΩ to 50 KΩ. These sensors are very thin and lightweight so it is comfortable for the subject.

This paper uses 4 flex sensors, 1 on each finger using a hand glove. These sensors are used to detect the bending of fingers during different hand gestures. As the flex sensors bend, there is a change in resistance and for the specific gesture, there is a specific output using voltage divider formula (i) as given below:

$$Vout = (R1 * Vin)/(R1 + R2) \qquad \text{(i)}$$

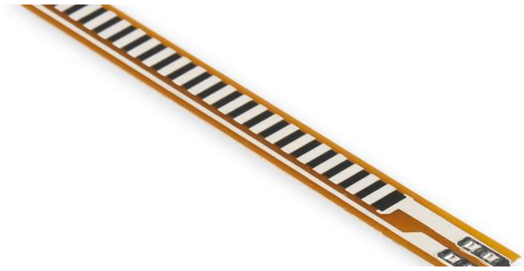

Fig 2: Flex Sensor

B. Accelerometer

The accelerometer is used to sense the angular displacement along the 3 axis x, y, z as shown in fig.3. The accelerometer is attached to the back of the hand to measure the angular displacement while making different gestures. The capacitance of the accelerometer varies according to the displacement. The change in capacitance is also given in analog form by the accelerometer.

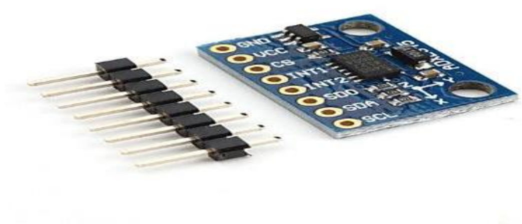

Fig 3: Accelerometer

C. Analog to Digital Converter (ADC)

The ADC used is 3208. The sensors give the analog values as the output but raspberry pi needs to have the digital values as the input. Also, the raspberry pi does not possess an internal analog to digital converter so there is a need for an external analog to digital converter that converts analog values provided by the sensors to the digital values as shown in fig.4 needed for the raspberry pi.

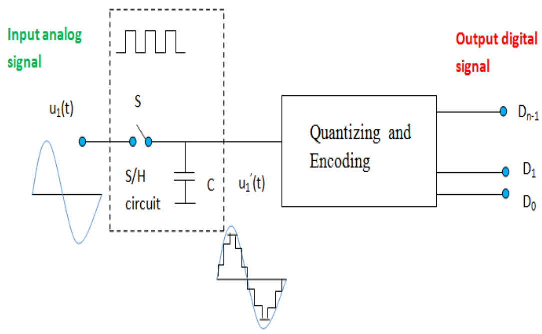

Fig 4: Working of ADC

D. Raspberry PI 3B

D.1. Features:

- Raspberry pi 3 B is the foremost version of the 3rd generation raspberry pi as shown in fig.5.
- It has 64 bit, quad-core. 1.2 GHz Broadcom BCN2837 processor.
- It has 1 GigaByte of Random Access Memory (RAM) for faster processing speed and 40 pins extended GPIO.
- It works on a micro USB power source up to a maximum of 2.5A.
- It also has a CSI camera port to support a raspberry pi camera if needed.
- The power can be given by simply connecting it to a PC with the help of a USB cable or start with a battery.

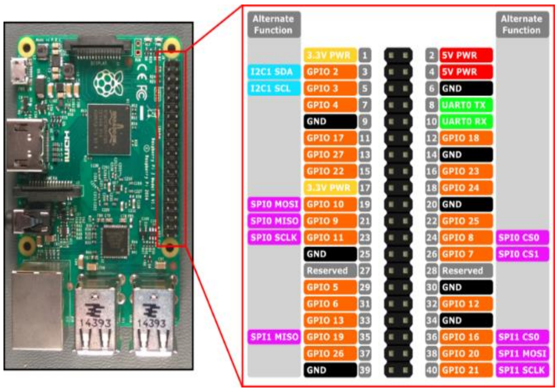

Fig 5: Raspberry PI 3B

D.1. Working:

A micro SD card having RASPBIAN operating system is mounted on the raspberry pi. The digital values obtained from the sensors are matched to pre-programmed values given in the python program and the respective messages are displayed and played through a speaker.

Different python libraries such as pyttsx3, RPI.GPIO, Time, OS are used.Pyttsx3 library is used to convert text to speech. Different voice modulation changes can be incorporated by varying the attribute values in this library. The set of messages can easily be changed by changing the text in the code.

E. Text and speech

A 16*2 display is used to show the output message from the raspberry pi board and an external set of speakers can be used to listen to the output message.

F. Process Flow

The flow chart of the process is simple and represented in fig.6. (i) The gesture is made by the user using the glove. (ii) The flex sensor and accelerometer give their respective output according to the gesture made. (iii) The analog values are converted to digital values by analog to digital converter (ADC). (iv) Raspberry pi matches those values with the pre-programmed values and if the value is matched the output message is played through speakers and displayed on an LCD screen. Else the user has to make the gesture again to obtain the output.

`

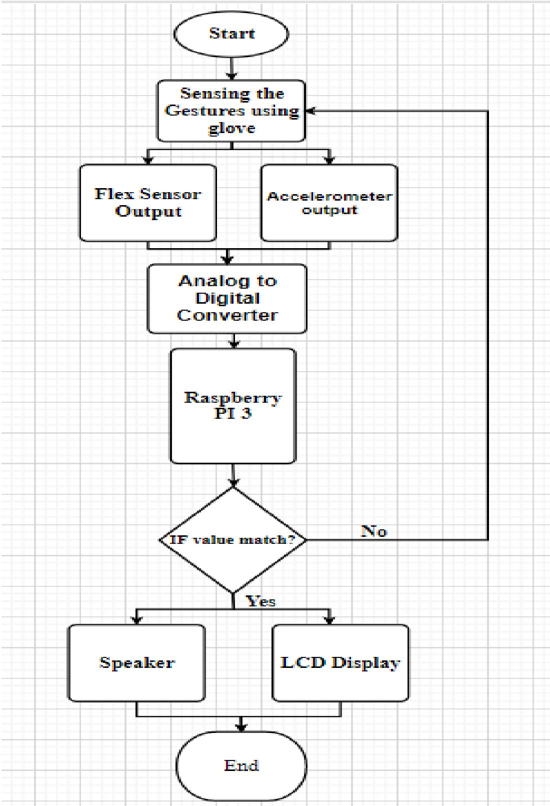

Fig 6: Flow Diagram

IV. RESULT AND DISCUSSION

A. System Setup

Following steps are used for setting up the system:

- The device works on the Raspbian operating system, therefore burn the Raspbian ISO image file to a micro SD card which is then mounted onto the raspberry pi board.
- Install VNC viewer which will work as a virtual application to access the Raspbian interface.
- Install and use the FING app to find the IP address of the Raspberry pi.
- Enter the obtained IP address in the VNC viewer and set up the password for system protection.
- Install python IDLE for writing the python code, which programs the raspberry pi.
- Open the new file and write the necessary python code for accelerometer, ADC, Raspberry pi and conversion of gestures.
- The power to pi can be given using a USB cable using an external power source.

B. Process Flow

Four flex sensors are attached to the index, middle, ring and pinky fingers. They are named as f1, f2, f3, and f4 respectively. The threshold value for the flex sensor is taken as 30. If the digital value of the flex sensor is above 30 the response is positive else negative. The accelerometer works in two positions horizontal and vertical which allows converting two gestures using the same finger positions. The output of the accelerometer in a horizontal position is less than 430 (<430) and in a vertical position, the output is greater than 430 (>430). The value 430 is used as a threshold value for the accelerometer. There are two sets of hand gestures having 7 gestures each. The details of each gesture are given in table 2. According to table 1, less than 30 (< 30) means no bending of the finger, while greater than 30 (> 30) means bending of the finger during gesture making.

Table 2: Flex values

Gesture	Accelerometer position	Index (f1)	Middle (f2)	Ring (f3)	Pinky (f4)
1	Horizontal	>30	<30	<30	<30
2	Horizontal	<30	>30	<30	<30
3	Horizontal	<30	<30	>30	<30
4	Horizontal	<30	<30	<30	>30
5	Horizontal	>30	>30	>30	>30
6	Horizontal	>30	>30	<30	<30
7	Horizontal	<30	<30	>30	>30
8	Vertical	>30	<30	<30	<30
9	Vertical	<30	>30	<30	<30
10	Vertical	<30	<30	>30	<30
11	Vertical	<30	<30	<30	>30
12	Vertical	>30	>30	>30	>30
13	Vertical	>30	>30	<30	<30
14	Vertical	<30	<30	>30	>30

The different gestures are shown in fig. 7. The figure has 7 gestures which are used to interpret 14 messages using two different modes of the accelerometer (horizontal and vertical).

Fig 7: Hand gestures

B.1. Accuracy Factor

Figures [7-9] depicts the accuracy vs gesture graph where the x-axis denotes the different gestures and the y-axis represents accuracy on a scale of 0 to 100. Accuracy for 7 gestures is shown when the accelerometer is in the horizontal position as shown in fig. 8. The accuracy varies from 79 percent to 93 percent. Gesture 7 is the least accurate while gesture 5 being the most accurate. The average accuracy for this set is 86 percent.

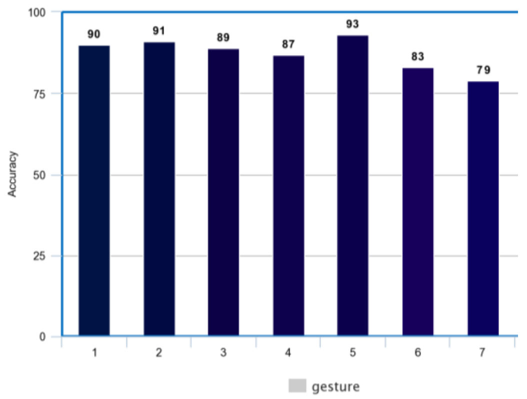

Fig 8: Gestures vs Accuracy (Horizontal)

Accuracy for the remaining 7 gestures is shown in fig. 9 when the accelerometer is in the vertical position. The accuracy varies from 71 percent to 89 percent. The accuracy of the vertical gestures is less than the horizontal gestures due to the difficulty in making the gesture while the hand is in the vertical position. The average accuracy of this set of gestures is 82 percent.

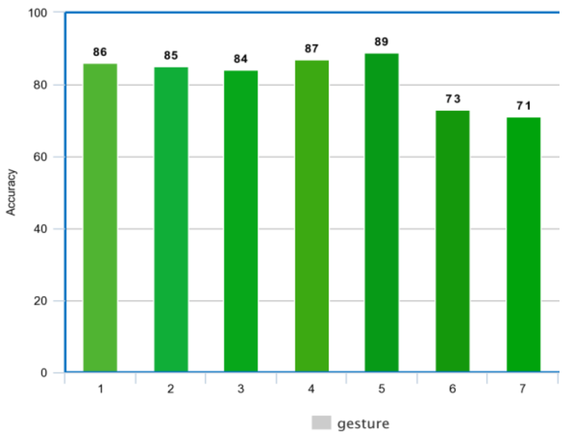

Fig 9: Gestures vs Accuracy (Vertical)

The combined accuracy of all 14 gestures is shown in fig. 10 while both the horizontal and vertical modes are active. There is a noticeable decrease in accuracy as there are more chances of mixing up messages and chances of inappropriate gestures. The overall average accuracy of the system is around 80 percent.

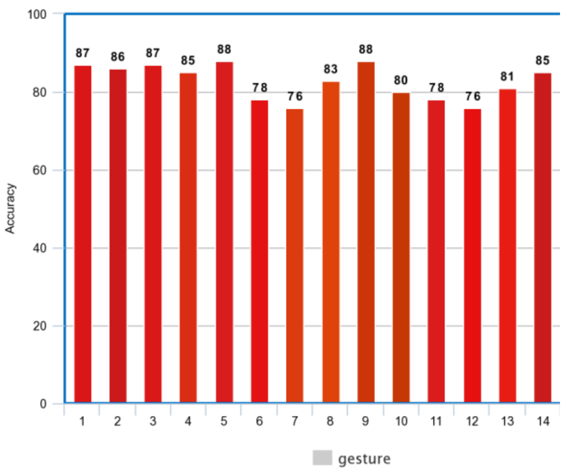

Fig. 10: Gestures vs Accuracy (Horizontal+ Vertical)

V.CONCLUSION

The way of living of the dumb people can be made better by the proposed system which can make them communicate effectively with the normal people as well as other hearing disabled people. This system is effective and gives fast respective response for the given respective 14 gestures using the 4 flex sensors. The messages can be modified according to the needs of the situation and subject (disabled person). The message is played as well as displayed through an LCD screen.

The future work includes the use of more number of flex sensors and gyroscope to increase the number of messages and their precision. The system can be made portable by using the battery or a small solar panel can be used as a power source. The inclusion of multilingual can provide flexibility to the system.

REFERENCES

1. V.Padmanabhan, M.Sornalatha, Hand gesture recognition and voice conversion system for dumb people, International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014 427 ISSN 2229-5518.
2. Z. Lei, Z. H. Gan, M. Jiang, and K. Dong, "Artificial robot navigation based on gesture and speech recognition," Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Wuhan, 2014, pp. 323-327.
3. H. S. K., Rai S, S., Pal, S., Sulthana K, U., & Chakma, S. Development of Device for Gesture to Speech Conversion for the Mute Community. 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), IEEE explore(2018).
4. S. C. Mukhopadhyay, "Wearable Sensors for Human Activity Monitoring: A Review," IEEE Sensors Journal, vol. 15, no. 3, pp. 1321-1330, March 2015.doi: 10.1109/JSEN.2014.2370945.
5. R. R. Itkarkar and A. V. Nandi, "Hand gesture to speech conversion using MATLAB, " 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, 2013, pp. 1-4.doi: 10.1109/ICCCNT.2013.6726505.
6. Ahmed, Syed Faiz & Ali, Syed & Munawwar, Saqib. (2010). "Electronic Speaking Glove for Speechless Patients, A Tongue to a Dumb". Sustainable Utilization and Development in Engineering and Technology, IEEE Conference on. 56 - 60. 10.1109.
7. Safayet Ahmed, Rafiqul Islam, Md.Saniat Rahman Zishan, Mohammed Rabiul Hasan, Md.Nahian Islam, "Electronic speaking system for speech impaired people: Speak up", Electrical Engineering and Information Communication Technology (ICEEICT) 2015 International Conference on, pp. 1-4, 2015.
8. S.K. Imam Basha, S. Ramasubba Reddy, "Speaking system to mute people using hand gestures", International Research Journal of Engineering and Technology, Volume: 05 Issue: 09, Sep 2018.
9. Vaibhav Mehra, Aakash Choudhury, Rishu Ranjan Choubey.Gesture To Speech Conversion using Flex sensors, MPU6050 and Python.International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6, August 2019.
10. Yang, H., An, X., Pei, D. and Liu, Y., 2014, September. Towards realizing gesture-to-speech conversion with an HMM-based bilingual speech synthesis system. In 2014 International Conference on Orange Technologies (pp. 97-100). IEEE.
11. Preetham, C., Ramakrishnan, G., Kumar, S., Tamse, A. and Krishnapura, N., 2013, April. Hand talk-implementation of a gesture recognizing glove. In 2013 Texas Instruments India Educators' Conference (pp. 328-331). IEEE.
12. Aiswarya, V., Raju, N.N., Joy, S.S.J., Nagarajan, T. and Vijayalakshmi, P., 2018, March. Hidden Markov Model-Based Sign Language to Speech Conversion System in TAMIL. In 2018 Fourth International Conference on Biosignals, Images and Instrumentation (ICBSII) (pp. 206-212). IEEE.
13. P. Vamsi Praveen, K. Satya Prasad,Electronic Voice to Deaf & Dumb People, Using Flex Sensor International Journal of Innovative Research in Computer and Communication Engineering Vol. 4, Issue 8, August 2016, ISSN(Online): 2320-9801

14. K. Manikandan, Ayush Patidar, Pallav Walia, Aneek Barman Roy,Hand Gesture Detection and Conversion to Speech and Text, International Journal of Pure and Applied Mathematics, Volume 120 No. 6 2018, 1347-1362, ISSN: 1314-3395

AUTHORS PROFILE

Dr. K P Vijayakumar, Assitant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai. He has interest in database systems, wireless sensor networks, IoT and big data analytics.

Ananthu Nair, Undergraduate Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai. He has keen interest in database systems and IoT.

Nishant Tomar, Undergraduate Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai. He has keen interest in machine learning and data science.