SIGN LANGUAGE TO SPEECH CONVERSION

Aarthi M
Department of ECE
SSN College of Engineering
Chennai, India
aarthimuthusami@gmail.com

Vijayalakshmi P
Department of ECE
SSN College of Engineering
Chennai, India
vijayalakshmip@ssn.edu.in

*Abstract—****Human beings interact with each other to convey their ideas, thoughts, and experiences to the people around them. But this is not the case for deaf-mute people. Sign language paves the way for deaf-mute people to communicate. Through sign language, communication is possible for a deaf-mute person without the means of acoustic sounds. The aim behind this work is to develop a system for recognizing the sign language, which provides communication between people with speech impairment and normal people, thereby reducing the communication gap between them. Compared to other gestures (arm, face, head and body), hand gesture plays an important role, as it expresses the user's views in less time. In the current work flex sensor-based gesture recognition module is developed to recognize English alphabets and few words and a Text-to-Speech synthesizer based on HMM is built to convert the corresponding text.***

Keywords - ASL, flex sensor, Atmega328, Tactile sensor, Accelerometer, Gesture recognition module, Text-to-speech synthesis module.

I. INTRODUCTION

Deaf-mute people need to communicate with normal people for their daily routine. The deaf-mute people throughout the world use sign language to communicate with other people. However, it is possible only for those who have undergone special training to understand the language. Sign language uses hand gestures and other means of non-verbal behaviors to convey their intended meaning [9]. It involves combining hand shapes, orientation and hand movements, arms or body movement, and facial expressions simultaneously, to fluidly express speaker's thoughts. The idea is to create a sign language to speech conversion system, using which the information gestured by a deaf-mute person can be effectively conveyed to a normal person. The main aim of this work is to design and implement a system to translate finger spelling (sign) to speech, using recognition and synthesis techniques. The modules to be present in the proposed system are,

a. Finger spelling (gesture) recognition module
b. Text-to-Speech synthesis module

Other applications of hand-gesture recognition systems include character-recognition, gesture recognition to remotely control a television set, home automation, robotic arm controller and gesture recognition for wheel chair control, games [9]. The overview of a sensor based system for gesture recognition is described in section II. Sign language to speech conversion system using flex sensor and Atmega328 microcontroller is described in section III. Hardware implementation of sensor-based gesture recognition system and the process of text-to-speech synthesis are discussed in section IV & V respectively. Performance of the sensor-based gesture recognition system is analyzed and the results are discussed in section VI.

II. OVERVIEW OF SENSOR BASED SYSTEM

Gesture recognition systems include two main categories namely, (i) Vision-based system and (ii) Sensor-based system.

A. Vision-based gesture recognition system

In a vision-based gesture recognition system, a camera is used for capturing the image/video of the gesture and it is shown in Fig. 1. The captured content is sent to the image processing unit where it is processed through image processing techniques. Features are extracted and the extracted features are trained using static images for which the corresponding gestures are recognized using various image recognition algorithms [1].

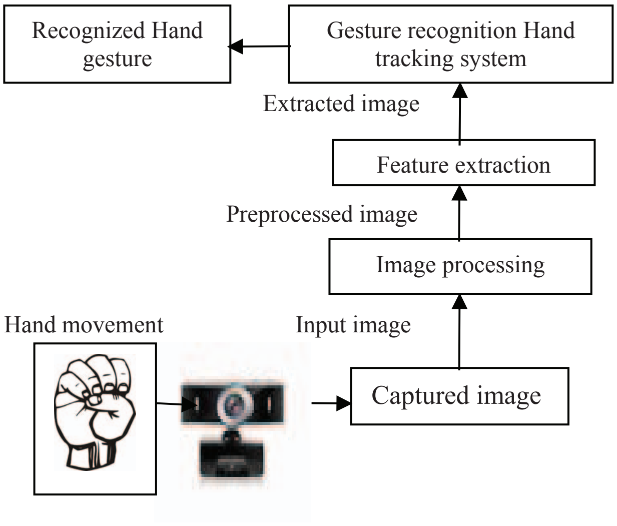


Fig.1 Vision based system (Redrawn from [24])

978-1-4673-9802-2/16/$31.00© 2016 IEEE

Compared to vision-based system sensor-based gesture recognition system (flex sensor) is sensitive and more accurate as it gives full degree of freedom for hand movement. The sensor-based approach is advantageous over vision-based system, since it requires only a motion sensor rather than a camera that makes it as a portable device with low cost. It also provides fast response in recognizing the gestures which in turn reduce the computational time in real time applications. In real time, recognition rate of 99% can be achieved using flex sensor-based system.

B. Proposed sensor-based gesture recognition system

The proposed system is a sensor based gesture recognition system which uses flex sensors for sensing the hand movements. The flex sensor is interfaced with the digital ports of Atmega328 microcontroller. The output from the microcontroller is the recognized text which is fed as input to the speech synthesizer. Arduino microcontroller processes the data for each particular gesture made. The system is trained for different voltage values for each letter. Gestures performed by multiple users have been tested for all the letters in ASL. Fig. 2 shows the block diagram of proposed sensor based system for sign language to speech conversion.

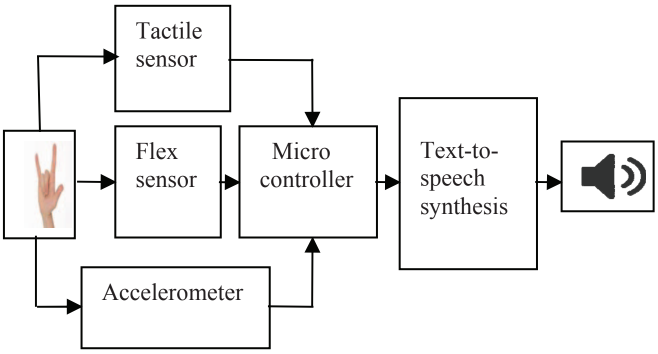


Fig.2 Block diagram of proposed sensor based system

In the proposed system, flex sensors are used to measure the degree to which the fingers are bent. Accelerometer within the gesture recognition system is used as a tilt sensing element, which in turn finds the degree to which the finger is tilted. Tactile sensor is used to sense the physical interaction between the fingers. The outputs from the sensor systems are sent to the Arduino microcontroller unit. In Arduino microcontroller unit, data derived from the sensor output is then compared with the pre-defined values. The corresponding gestures (matched gestures) are sent to the text-to-speech conversion module in the form of text. The output of text-to-speech synthesis system is heard via a speaker. The main features of this system includes it’s applicability in day-to-day life, portability and it’s low cost.

III. SIGN LANGUAGE TO SPEECH CONVERSION SYSTEM

This sign language recognition system will provide communication between normal people and the people with speech impairment. In this flex sensor-based gesture recognition system, Atmega 328 processor along with five flex sensors recognizes the sign language performed by user. Arduino microcontroller is advantageous over other platforms as it is of low cost and available as open source software. Accelerometer is used to measure the orientations of hand movements. Tactile sensor measures the force applied on one finger by other finger.

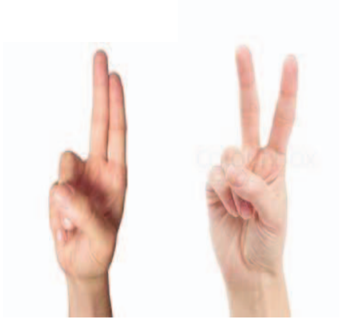


Fig.3 Gesture for letters U & V [16]

The letters such as M, N and T have similar gestures and also the letters U and V show similarity in their gestures, which is shown in Fig. 3. To overcome the difficulty in recognizing these letters tactile sensors are used. In this work tactile sensor is used to improve accuracy in recognizing these letters.

A. Flex sensor

Flex sensor changes its resistance value depending upon the amount of bend applied on the sensor. By measuring the resistance, we determine how much the sensor is being bent. An unflexed sensor has 10 to 30K ohm resistance and when it is bent, the resistance value increases to 100K ohm [2]. One side of the sensor is printed with a polymer ink that has conductive particles embedded in it. Fig. 4 and Fig. 5 depicts the structure and the flexible nature of flex sensor respectively.

Fig.4 Flex sensor characteristics [23]

The sensor based system is designed using three 4.5 inch and two 2.2 inch flex sensors. The 20kΩ resistor along with the flex sensor forms a voltage divider circuit which divides the input voltage by a ratio determined by the variable (resistance of flex sensor) and the fixed resistors (20kΩ).

The output voltage is determined using the following equation,

$$V_{OUT} = \frac{V_{IN} \times R_1}{(R_1 + R_2)}$$

R1 - flex sensor resistance
R2 - Input resistance

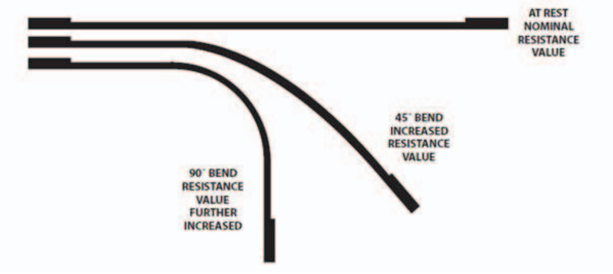


Fig.5 Flex sensor [18]

B. Atmega 328

Arduino is an electronics prototyping platform based on user-friendly software and easy-to-use hardware and it is available as an open-source. Arduino runs on Mac, Windows, and Linux [2].Atmega 328 is a microcontroller unit present in the Arduino board. Atmega 328 (Shown in Fig. 6) has 14 digital I/O pins out of which 6 can be used for Pulse width modulation outputs, 6 for analog inputs. It also consists of 16 MHz crystal oscillator, a port for USB connection, a port for power jack, an ICSP header, and a reset button [22]. It can be connected to a computer using an USB cable or by using an AC-to-DC adapter or by using a 9V battery. ATmega328 has 32 KB of flash memory module for storing the programming, out of which 2 KB is used by the bootloader. It has 2 KB of SRAM and 1 KB of EEPROM which can be read and write with the EPROM memory [8]. Atmega328 microcontroller can be programmed using the Arduino programming language.

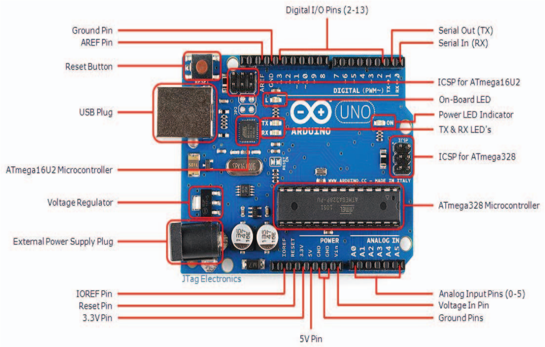


Fig.6 Atmega 328 [25]

C. Tactile sensor

A tactile sensor is a robust polymer thick film device whose resistance changes when a force is applied. This sensor can measure force between 1kN to 100kN [17]. They are also known as force-sensing resistors. Its resistance decreases with increase in force applied on the surface of the sensor. It requires a simple interface. Compared to other sensors, the advantages of tactile sensor are their size, low cost and good resistance. This sensor is used in human touch control, industrial and robotic applications.

Fig.7 Tactile Sensor [17]

Tactile sensor resistance changes as more pressure or force is applied. When there is no pressure applied, the sensor looks like an open circuit and as the pressure increases, the resistance goes down. Fig. 7 shows the tactile sensor used in the current work.

D. 3-axis Accelerometer

Accelerometer used within the gesture recognition system is employed as a tilt sensing element, used for finding the hand movement and orientation [14] and it is shown in Fig. 8. It measures the static as well as dynamic acceleration. The sensor has a g-select input, which in turn switches the measurement range of accelerometer between ± 1.5g and ±6g. Accelerometer has a signal conditioning unit with a single pole low pass filter. Provision for temperature compensation, self-testing, and 0g-detect (for detecting the linear free fall) is also present [21].

Fig.8 Accelerometer [21]

IV. HARDWARE IMPLEMENTATION

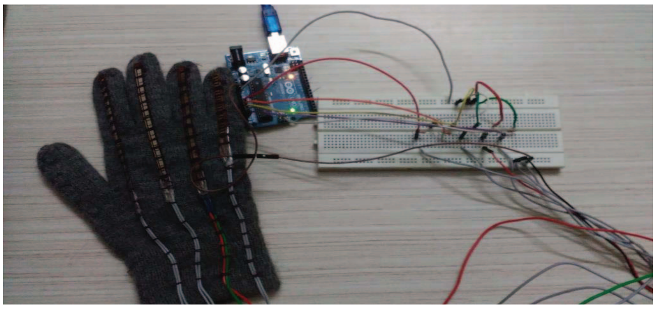


Fig.9 Setup for gesture recognition

The experimental set-up for gesture recognition is shown in Fig. 9 and the steps involved in sign language to speech conversion are described as follows:

Step1: The flex sensors are mounted on the glove and they are fitted along the length of each of the fingers.

Step2: Depending upon the bend of hand movement different signals corresponding to x-axis, y-axis and z-axis are generated.

Step3: Flex sensors outputs the data stream depending on the degree and amount of bend produced, when a sign is gestured [14].

Step4: The output data stream from the flex sensor, tactile sensor and the accelerometer are fed to the Arduino microcontroller, where it is processed and then converted to its corresponding digital values.

Step5: The microcontroller unit will compare these readings with the pre-defined threshold values and the corresponding gestures are recognized and the corresponding text is displayed.

Step6: The text output obtained from the sensor based system is sent to the text-to-speech synthesis module.

Step7: The TTS system converts the text output into speech and the synthesized speech is played through a speaker.

V. HMM BASED TEXT-TO-SPEECH SYNTHESIS SYSTEM

A statistical parametric approach for speech synthesis based on hidden Markov models (HMMs) is the most popular speech synthesis technique [12]. In HMM-based speech synthesis, the speech parameters such as the spectrum of speech signal, its fundamental frequency (F0), and the duration information for the phonemes are statistically modelled and the speech is synthesized using HMMs based on the maximum likelihood criteria. The model is parametric because it describes the speech using parameters and it is statistical because the parameters are described using means and variances. To eliminate the spectral discontinuity in the spectrum of synthesized speech along with static coefficients, delta and acceleration coefficients are also computed. These three types of parameters are stacked together into a single observation vector for the model. For the current work the HMM based models are initially trained for a female speaker's speech data.

A.HMM based speech synthesis system

In this work, Hidden Markov Model (HMM)-based text-to-speech synthesis system (HTS) is used to synthesize speech in English language. The HTS system involves two phases namely (i) Training phase and the (ii) Synthesis phase.

In the training phase, the spectral parameters such as Mel generalized cepstral coefficients (mgc) and its excitation parameter, namely the log fundamental frequency (lf0) are extracted from the collected speech data. The speech unit chosen for the model is pentaphone that includes two left and two right contexts.

Using this speech data and their corresponding orthographic transcription, time aligned phonetic transcription is derived based on a phoneset in the Festival platform [19]. In order to train these context-dependent phoneme models, question set is formulated.

In synthesis phase, given a test sentence in text format the corresponding context-dependent label files are generated. According to the label sequence, a sentence-level HMM is generated by concatenating context-dependent HMMs. Then by using speech parameter generation algorithm a sequence of speech parameters such as the spectral and excitation parameters is determined in such a way that it maximizes its output probability. Finally, speech is synthesized directly from the generated spectral and excitation parameters using a source system synthesis filter namely, mel-log spectral approximation filter [13].

B. Training Phase

The training phase involves the pre-requisites such as (i) normalized text data and the corresponding training speech data (ii) time aligned phonetic transcription of speech data (iii) phone set (iv) question set

a. Speech corpus

Text data of 600 sentences is collected from children short stories for English language and the corresponding speech data is collected. The speech is recorded at a sampling rate of 16 KHz. The collected English text data contains a total of 19,896 words out of which 1,141 words are unique.

b. Phonetic transcription

The time-aligned phonetic transcription is derived using forced Viterbi alignment algorithm. In the current work the segmentation is performed in the phoneme level.

c. Common phoneset

For each phoneme in the language, their phonetic and linguistic features are defined as a phone set, which are defined in terms of their place and manner of articulation, vowel/consonant, etc. The lexicons, letter to sound rules and waveform synthesizers requires definition of phone set for their operation. s

d. Question set

Question set plays a major role in tree-based clustering to handle unseen data and to cluster similar speech units. It is used to perform state-tying which enables in reducing number of parameters and hence computations. A question set contains questions relevant to the linguistic and phonetic classifications. At each node of the tree, a question is chosen, if there is a gain in the likelihood. 60 questions are formulated and the phonemes are grouped based on them.

The HTS system is trained using parameters (system and source) extracted from one hour of speech data, orthographic phonetic transcriptions, phone set and question set. Finally a sign language to speech conversion system is developed by interfacing the sensor-based system with the speech synthesizer.

VI. RESULTS AND ANALYSIS

TABLE I. Flex sensor voltage value and their corresponding alphabet

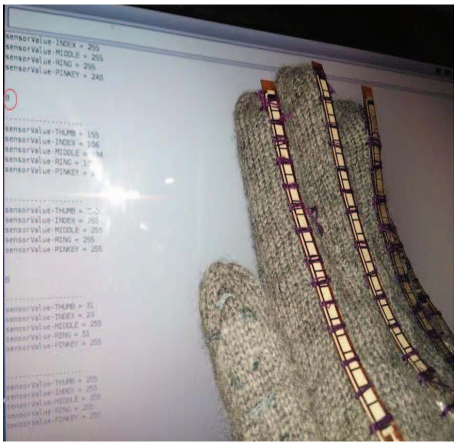
Index value	Middle value	Ring value	Pinky value	Angle (degrees)	Recognized alphabet
>100	>1	1-100	>150	90	A
255	255	>200	1	0	B
<150	<150	>50	<100	30	C
1	>100	>150	1	60	D

Table I shows the voltage values combination for displaying the alphabets A, B, C and D. Likewise different combinations of voltage values are used to recognize the remaining alphabets. Table II shows the voltage values for which the corresponding word is displayed.

TABLE II. Flex sensor voltage value and their corresponding word

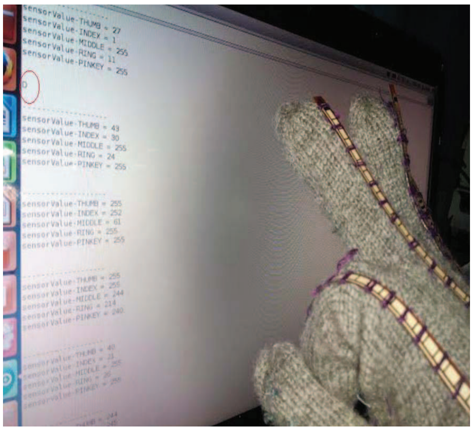
Index value	Middle value	Ring value	Pinky value	Recognized word
1to50	200	>250	255	Bag
255	>220	1	<200	Welcome
1	<100	255	1to100	Beg
>100	<150	1	>50	Egg
1	<240	255	<100	Bad

Fig.10 and Fig. 11 shows the alphabet "B" and alphabet "D" displayed on the serial monitor of the Atmega328 microcontroller when it is gestured by the user. Similarly the remaining alphabets can be gestured and the corresponding text is displayed. Fig. 12 shows the word "welcome" when it is gestured by the user. Likewise commonly uttered words can be gestured and the corresponding text is displayed.



```
sensorValue-INDEX = 255
sensorValue-MIDDLE = 255
sensorValue-RING = 255
sensorValue-PINKEY = 240

B
-------------------
sensorValue-THUMB = 155
sensorValue-INDEX = 106
sensorValue-MIDDLE = [illegible]
sensorValue-RING = [illegible]
sensorValue-PINKEY = [illegible]

-------------------
sensorValue-THUMB = [illegible]
sensorValue-INDEX = 255
sensorValue-MIDDLE = 255
sensorValue-RING = 255
sensorValue-PINKEY = 255

B
-------------------
sensorValue-THUMB = 31
sensorValue-INDEX = 23
sensorValue-MIDDLE = 255
sensorValue-RING = 51
sensorValue-PINKEY = 255

-------------------
sensorValue-THUMB = 255
sensorValue-INDEX = 255
sensorValue-MIDDLE = 255
sensorValue-RING = 255
sensorValue-PINKEY = 255
```

Fig.10 Gesture for 'B' and the corresponding text is displayed



```
-------------------
sensorValue-THUMB = 27
sensorValue-INDEX = 1
sensorValue-MIDDLE = 255
sensorValue-RING = 11
sensorValue-PINKEY = 255

D
-------------------
sensorValue-THUMB = 43
sensorValue-INDEX = 30
sensorValue-MIDDLE = 255
sensorValue-RING = 24
sensorValue-PINKEY = 255

-------------------
sensorValue-THUMB = 255
sensorValue-INDEX = 252
sensorValue-MIDDLE = 61
sensorValue-RING = 255
sensorValue-PINKEY = 255

-------------------
sensorValue-THUMB = 255
sensorValue-INDEX = 255
sensorValue-MIDDLE = 244
sensorValue-RING = 214
sensorValue-PINKEY = 240

-------------------
sensorValue-THUMB = 40
sensorValue-INDEX = 21
sensorValue-MIDDLE = 255
sensorValue-RING = 26
sensorValue-PINKEY = 255
```

Fig.11 Gesture for 'D' and the corresponding text is displayed

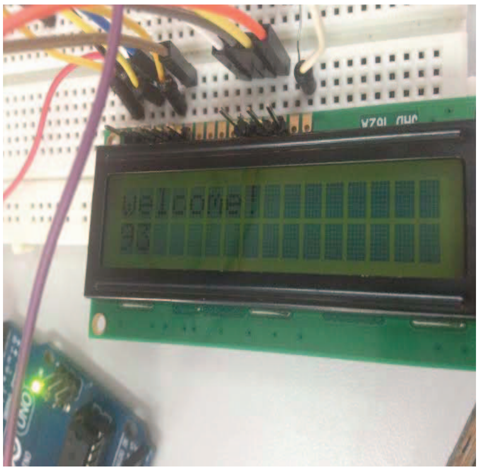


Fig.12 Recognition of the word "welcome"

Table III shows the performance analysis of sensor based system for gesture recognition. The performance of the system is calculated based on the probability of correctly recognizing the gesture by the system.

TABLE III Analysis of performance of sensor based system

	A	B	C	D	E	F	G	H	Performance (%)
1	✓	✓	✓	✓	✓	✓	✗	✓	87.5
2	✓	✓	✓	✓	✓	✓	✓	✓	100
3	✓	✓	✓	✗	✓	✓	✗	✓	75
4	✓	✓	✓	✓	✗	✓	✓	✓	87.5
5	✓	✓	✓	✓	✓	✓	✓	✓	100
6	✗	✓	✓	✓	✓	✓	✓	✓	87.5
7	✓	✓	✓	✓	✓	✗	✓	✗	75
8	✓	✓	✓	✓	✓	✓	✓	✓	100
9	✓	✓	✓	✓	✓	✗	✓	✗	75
10	✓	✓	✗	✓	✓	✓	✓	✓	87.5

In the current work the system is trained for a set of 8 gestures. Based on user interaction with the setup, the experiment is carried out for 10 times and the average recognition rate obtained is 87.5% across all the gestures and for each of these gestures the average recognition rate is found to be 80-90%.

VII. CONCLUSION

Deaf-mute people face difficulty in communicating with normal people who do not understand sign language. To overcome this barrier this work proposes a sensor-based system for deaf-mute people using glove technology. It requires fewer components such as flex sensor, Arduino and accelerometer and hence its cost is low compared to vision-based gesture recognition system. In this system the deaf-mute people wear the gloves (with the resistors and sensors attached to it) to perform hand gesture. First the system will convert the gesture to the corresponding text and then the speech is synthesized for the corresponding text by using the text-to-speech synthesizer. The system consumes very low power and it is portable. The sensor glove design along with the tactile sensor helps in reducing the ambiguity in gestures and shows improved accuracy. This paper can further be developed to convert words, phrases and simple sentences by concatenating the alphabets. More number of short flex sensors can be employed to recognize more gestures.

REFERENCES

[1] G Adithya V., Vinod P. R., Usha Gopalakrishnan, “Artificial Neural Network Based Method for Indian Sign Language Recognition” , IEEE Conference on Information and Communication Technologies (ICT) , 2013, pp. 1080-1085.

[2] Ahmed, S.F.; Ali, S.M.B.; Qureshi, S.S.M., "Electronic Speaking Glove for speechless patients, a tongue to a dumb," Sustainable Utilization and Development in Engineering and Technology, 2010 IEEE Conference on , vol., no., pp.56-60, 20-21 Nov. 2010.

[3] Chun Yuan; Shiqi Zhang; Zhao Wang, "A handwritten character recognition system based on acceleration," Multimedia Technology and its Applications (IDCTA), 7th International Conference on , vol., no., pp.192-198, 16-18 Aug. 2011.

[4] Dekate, A.; Kamal, A.; Surekha, K.S., "Magic Glove - wireless hand gesture hardware controller," Electronics and Communication Systems (ICECS), 2014 International Conference on, vol., no., pp.1-4, 13-14 Feb 2014.

[5] Hernandez-Rebollar, J.L.; Kyriakopoulos, N.; Lindeman, R.W., "A new instrumented approach for translating American Sign Language into sound and text," Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, vol., no., pp.547-552, 17-19 May 2004.

[6] Kadam K, Ganu R, Bhosekar A, Joshi S.D.,"American Sign Language Interpreter," Technology for Education (T4E), 2012 IEEE Fourth International Conference on ,vol., no., pp.157-159, 18-20 Jul 2012.

[7] Khambaty, Y.; Quintana, R.; Shadaram, M.; Nehal, S.; Virk, M.A.; Ahmed, W.; Ahmedani,G.,"Cost effective portable system for sign language gesture recognition," System of Systems Engineering, 2008. SoSE '08. IEEE International Conference on, vol., no., pp.1-6, 2-4 Jun 2008.

[8] Dhanalakshmi M,Anbarasi Rajamohan, Hemavathy R, "Deaf-Mute Communication Interpreter”, International Journal of Scientific Engineering and Technology (ISSN : 2277-1581) vol. 2 Issue 5, pp. 336-341,1 May 2013.

[9] Priyanka Lokhande, Riya Prajapati and Sandeep Pansare, “Data Gloves for Sign Language Recognition System” IJCA Proceedings on National Conference on Emerging Trends in Advanced Communication Technologies NCETACT, pp. 11-14, Jun 2015.

[10] Rajam, P. Subha and Dr G Balakrishnan, "Real Time Indian Sign Language Recognition System to aid Deaf and Dumb people", 13th International Conference on Communication Technology (ICCT), 2011, pp. 737-742.

[11] Ruize Xu, Shengli Zhou, Li, W.J, "MEMS Accelerometer Based Nonspecific-User Hand Gesture Recognition", IEEE Sensors Journal, vol.12, no.5, pp.1166-1173, May 2012.

[12] K. Tokuda, H. Zen, A.W. Black, An HMM-based speech synthesis system applied to English, Proc. of 2002 IEEE Speech Synthesis Workshop, Sep. 2002, pp 227-230.

[13] Tokuda, K.; Nankaku, Y.; Toda, T.; Zen, H.; Yamagishi, J.; Oura, K., "Speech Synthesis Based on Hidden Markov Models," in Proceedings of the IEEE , vol.101, no.5, pp.1234-1252, May 2013.

[14] Vinod J Thomas, Diljo Thomas, “A Motion based Sign Language Recognition Method”, International Journal of Engineering Technology, vol 3 Issue 4, ISSN 2349-4476, Apr 2015.

[15] Zhou Ren; Junsong Yuan; Jingjing Meng; Zhengyou Zhang, “Robust Part-Based Hand Gesture Recognition Using Kinect Sensor,” in IEEE Transactions on Multimedia, vol.15, no.5, pp.1110-1120, Aug 2013.

[16] https://en.wikipedia.org/wiki/Sign_language accessed on 20 Mar 2016.

[17] http://oomlout.co.uk/products/force-sensitive-resistor-interlink-fsr-402 accessed on 23 Mar 2016.

[18] http://www.engineersgarage.com/contribution accessed on 20 Mar 2016

[19] http://www.cstr.ed.ac.uk/projects/festival/manual/festival_12.html accessed on 20 Mar 2016.

[20] http://iosrjen.org/Papers/vol5_issue8%20 (part-4)/A05840104 accessed on 15 Mar 2016.

[21] http://www.ifuturetech.org/product/mma7361-mma7260-accelerometer accessed on 20 Mar 2016.

[22] https://www.arduino.cc/en/Main/ArduinoBoardUnoSMD accessed on 29 Mar 2016.

[23] http://www.spectrasymbol.com/flex-sensor accessed on 01 Apr 2016.

[24] http://www.ijser.org/paper/Real-Time-Gesture-Recognition-Using-Gaussian-Mixture-Model.html accessed on 02 Mar 2016.

[25] http://www.jtagelectronics.com/?p=75 accessed on 20 Mar 2016.