Gesture Vocaliser Using Leap Motion Controller

Nikul P[1, a)], Nithin X Johnson[1, b)], Sooraj B P[1, c)], Vaishak Sajeev[1, d)], Asha S[1, 2, e)]

[1] *Department of Electronics & Communication Engineering, SCT College of Engineering,* Thiruvananthapuram, India

[2] Assistant Professor

[a)] Nikul P: nikulpratheep@gmail.com

[b)] Nithin X Johnson: nithinxjohnson997@gmail.com

[c)] Sooraj B P: bpsooraj@gmail.com

[d)] Vaishak Sajeev: vaishaksajeev97@gmail.com

[e)] Asha S: asha@sctce.ac.in

Abstract

In present-day societies, talking is the main form of communication between humans. Different thoughts that are shaped in the human psyche are conveyed by discourse through the form of words, expressions, and sentences by applying some appropriate syntactic rules. One of the major challenges for people with hearing impairment is their fettered communication with the outside world. Deaf people make use of sign language or gestures to convey their thoughts but it is impossible to understand or comprehend for a person who does not understand sign language.

American sign language is the most standardized sign language used by ill abled people throughout the world for communication. We propose a gesture recognition system called 'Gesture Vocaliser' which detects American sign language alphabets (ASLA) as gestures . Upon showing these signs the gesture detection system detects the sign shown with the help of a classifier. The gesture detection system consists of a vision sensor which uses tracking technology to track the hand gestures. Whenever the proposed system senses any gesture, it converts these gestures into text, which is then converted into an audio message. There is a reverse phase in which human speech is recognised as a text and is then converted into sign language gestures for each corresponding character.

Introduction

- **Problem Statement:**

Deaf and mute people find it difficult to communicate with the real world. Establishing a means of fluid communication or interaction with these people cannot be neglected. These people interact or communicate through the means of hand gestures or signs. It is inconvenient to find a well versed and experienced translator for communication always but a human-computer interaction system for this can be used when convenient.

- **Objective**:

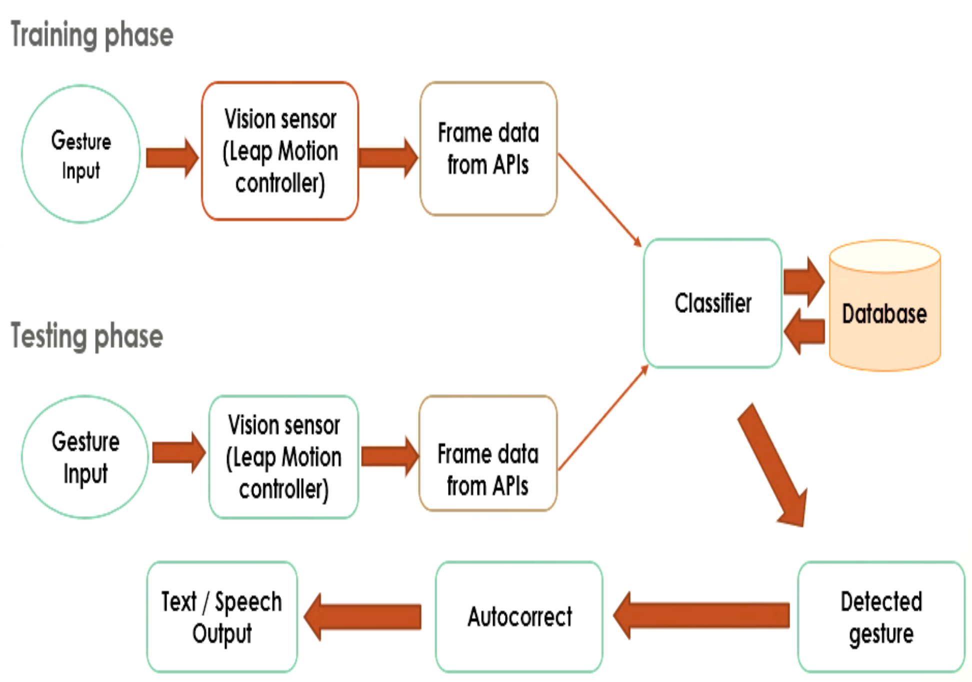
To assess the feasibility of developing a communication aid system using Leap Motion Controller which captures the sign language (ASLA) and converts it into an artificial human voice.

- **Overview:**

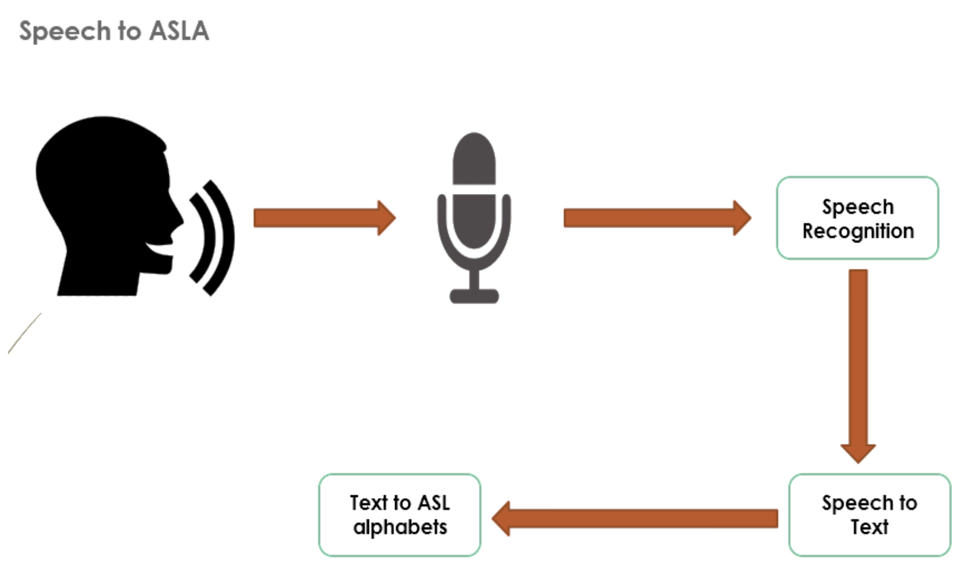
The system includes an autocorrect mechanism in which the incorrect or misspelled words after prediction will be automatically corrected to a meaningful word which in turn is converted into speech using a Pyttsx module. The deaf & mute people can communicate more effectively by using their own sign language. For that purpose, we have designed a reverse phase system in which our human speech is converted into American Sign Language alphabets.

Leap Motion Controller
(*Source: https://www.hardwarezone.com.sg/review-taking-leap-faith-leap-motion-controller*)

Methodology



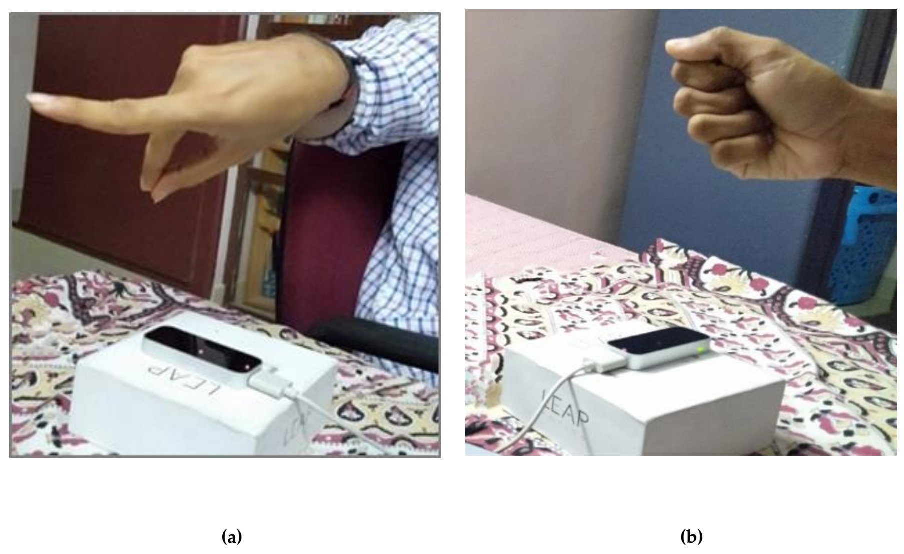
- A dataset created using the data obtained from the LMC is used to train the SVM classifier.
- The data collected is divided into training and testing phases.
- The dataset is divided at an 8:2 ratio for training and testing.
- The training phase provides the framework for recognizing the gestures.



- The system is also capable of converting speech into sign language alphabets.
- Using a speech to text engine provided by Google, the content of speech is converted to text and then displayed as images of ASL gestures.

Results & Discussions

- Considering the ASL alphabets, twenty-two out of twenty-six were trained to be recognized with 50 samples being taken for each letter. The remaining four include 'J', 'M', 'N' and 'Z'.
- The gestures for 'J' and 'Z' are dynamic gestures i.e. they include some motion and under the current scope, it is difficult to track a dynamic gesture due to the limitations of the predictive model employed in the Leap SDK.
- A possible solution to this would be to consider the gestures as static and train the classifier accordingly.
- The gestures for 'M' and 'N' are having somewhat similar features when they are viewed through the sensor's visualizer.
- False gesture detection may occur and because of this reason distinguishing 'M' and 'N' gestures are a difficult task.



Training the classifier for ASLA (a)'P' with changed orientation.(b) T' with changed orientation.

- Proper orientation of the hand is required to detect some gestures.
- For instance, 'P' and 'T' were trained in such a way that its ASLA gesture remains unchanged but the orientation of hand had been changed slightly and because of this, the letters can be properly detected without changing their gestures as shown in fig. (a) and fig. (b) respectively.
- There is a possibility that errors can happen during spelling and to tackle this issue a failsafe autocorrect has been implemented which would correct and replace the misspelled word with the most probable word from a word frequency list.

- The speech recognition works efficiently by setting a proper threshold level.
- The threshold setting is done to distinguish speech from ambient noise levels. Without appropriate thresholding, proper recognition of speech becomes difficult.
- Setting up a threshold level according to ambient noise levels is much better than setting a fixed threshold. For that purpose, an ambient noise level scanning for a duration of 3 seconds is employed. Better recognition can be achieved by increasing the scanning time with the trade-off of delay.

Predicted label (columns) / **Actual Label** (rows)

	A	B	C	D	E	F	G	H	I	K	L	O	P	Q	R	S	T	U	V	W	X	Y
A	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	1	0	0	0
I	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	1	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
U	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	6	0	0	0	0
V	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	10	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0
X	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11

Confusion matrix for 22 alphabets (excluded 'J', 'M', 'N' and 'Z')

- The total dataset is split into training and testing data. Out of 1100 samples in the dataset 220 samples are taken for testing. From the confusion matrix as shown in above figure.
- The overall accuracy is computed as the sum of all the diagonal elements in the confusion matrix divided by the total number of samples taken for testing, i.e., 220 samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- ***TP - True Positive***
- ***TN - True Negative***
- ***FP - False Positive***
- ***FN - False Negative***

Accuracy = 97.27%

Conclusion

- The proposed project can identify & convert ASLA gesture to text with substantial accuracy with the exception of gesture for 'M','N','J' & 'Z' which can be overcome by using the tailor-made replacement of these gestures which can be aptly detected.
- The Leap Motion Controller & its SDK has two limitations, both of which can be overcome:

(i) The predictive model it employs sometimes results in false identification of gestures. This occurs during inputs from gestures which involve closed palms or from crisscrossing of fingers. This limitation has been overcome by using an auto-correction feature implemented. The accuracy would hence also depend on the quality of spellchecker used and its database.

(ii) The power consumption coupled with the heating of the sensor causes poor recognition but this can be fixed by implementing a cooling solution.

- The reverse-phase which converts the speech to text as well as images of apt gestures work to precision. Furthermore, the current system can be expanded to detect dynamic gestures and phrases in the future.

Future Scope

- The system can recognize ASL alphabets and words can be formulated. This can be further developed to formulate sentences and entire paragraphs by introducing gestures for spaces and punctuation marks.
- Currently, this system utilizes auto-correct to rectify possible errors in gesture prediction. Further developing this, auto-completion and word prediction can aid the ASL users to communicate in a much faster and efficient way.
- Based on this system, a Personal Assistant can be developed to facilitate smoother conversations between an ASL user and an individual who cannot comprehend ASL. With a good word prediction algorithm, the detection time delay will be balanced out and the data input rate will be increased.

Acknowledgment

We have to express our appreciation to Sanju Mathew for sharing his pearls of wisdom with us during the course of this project. We are also immensely grateful to our colleagues for their comments on an earlier version of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed professionals.

References

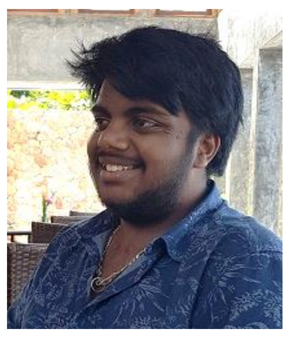
1. David Huggins-Daines, Mohit Kumar, Arthur Chan, Alan W Black, Mosur Ravishankar, and Alex I. Rudnicky," Pocketsphinx: a free, real-time continuous speech recognition system for hand-held devices", ICASSP 2006, pp 185 to 188.
2. G. Marin, F. Dominico, P. Zanuttigh, "Hand gesture recognition with Leap Motion and Kinect devices", 2014 IEEE international conference on image processing (ICIP), 27-30, October 2014, pp. 1565-1568.
3. Jayashree R. Pansare, Shravan H. Gawande, Maya Ingle, "Real-time static hand gesture recognition for American Sign Language (ASL) in complex background", Journal of Signal and Information Processing, 2012, 3, August 2012, pp. 364-367.
4. Lionel Pigou, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen "Sign language recognition using Convolutional Neural Networks", Springer International Publishing Switzerland 2015, pp. 572–578.
5. Ling Shao, Dong Xu, Jamie Shotton, and Jungong Han, "Enhanced computer vision with Microsoft Kinect sensor: A Review", IEEE Transactions on Cybernetics, vol. 43, no. 5, October 2013, pp 1427 to 1329.
6. Ming-Chang Liu, Liangyin Yu, Quan Yuan Pub, "Hand gesture detection patent from Sony Corp.", Patent number: US8792722B2, September 18, 2014.
7. http://blog.leapmotion.com/hardware-to-software-how-does-the-leap-motion-controller-work/
8. https://scikit-learn.org/stable/tutorial/basic/tutorial.html

Authors

Nikul Pratheep:
https://www.linkedin.com/in/nikul-p-a88582106

Nithin X Johnson:
https://www.linkedin.com/in/nithin-x-johnson-2495b6168

Sooraj B P:
https://www.linkedin.com/in/sooraj-b-p



Vaishak Sajeev:
https://www.linkedin.com/in/vaishaksajeev

Asha S:
https://www.linkedin.com/in/asha-s-00082440

- **View the conference paper: :**
- **View the Git Repository :**