A Glove and Software Algorithm that Translate American Sign Language to Text Utilizing Hand Positioning and Facial Expressions

Nikhil Chandra, Leland High School, San Jose CA

Abstract

There are currently over 10 million individuals that are hard of hearing or functionally deaf in the US. Despite this, there is no translation system on the market of American Sign Language to English text. Thus the objective of the project was to design a glove and software algorithm that could translate ASL to text while taking into account the facial expressions of the user, both key components of ASL. The glove had 6 sensors that measured 11 variables including the bend of each finger, and the 6 degrees of freedom. During pre-data collection 10 signs were initially signed 50 times each, and the averages for each data marker for that sign was recorded. During a live translation, an average for each data marker would also be found, and compared through their euclidian distances to the averages for each sign found during pre-data collection, and the least euclidian distance would point to the translated sign. 1000 trials were conducted to measure the glove's accuracy, and a 99.7% accuracy for the 10 signs was achieved with and without facial expressions. However, when facial expressions were taken into account the confidence of the model in its translations increased by about 20%. The glove then scaled to 40 signs, and maintained an accuracy of 99.3% with facial expressions. The application of this glove is far reaching by allowing hard of hearing individuals to more easily communicate on a daily basis to family members or others who do not understand the language.

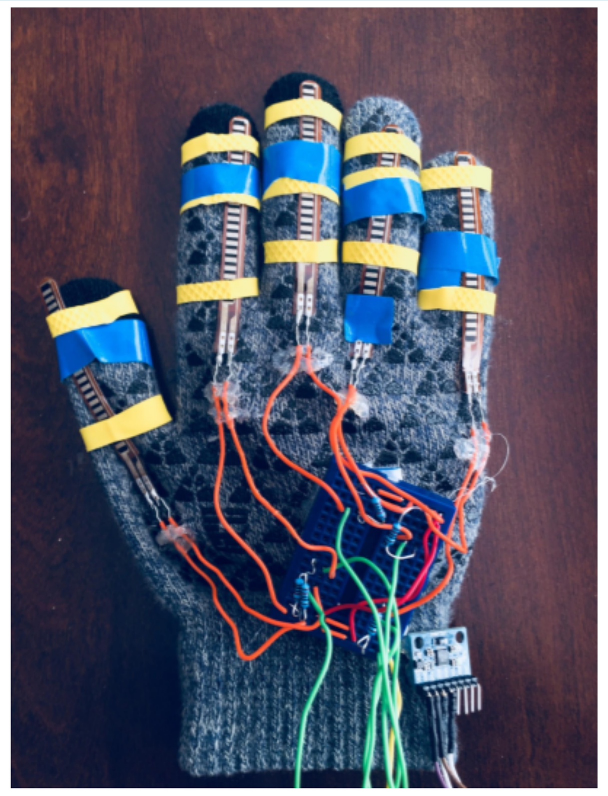
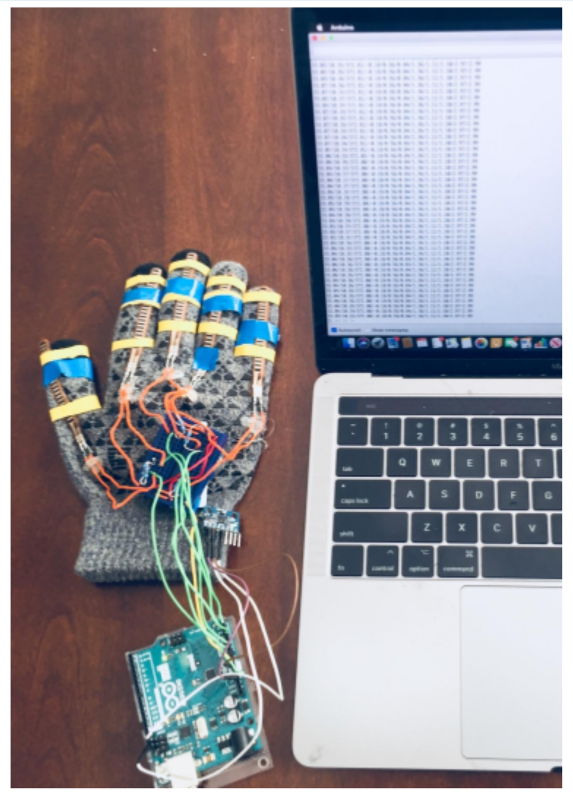
Problem

There are currently 10 million individuals in the US that are hard of hearing, 1,000 individuals that are functionally deaf, and over 500,000 that use American Sign Language(ASL) as their sole language. However, there is no ASL translation system on the market. Unlike other languages, deaf individuals interact with others who do not understand ASL on a daily basis, making a translator necessary, and human interpreters are not accessible to all.

Previous Research and Objective

- The majority of sign language translation systems only translate the alphabet such as research done by
  - Syed Atif Mehdi from University of Kaiserslautern, Germany, where he designed a glove to translate the ASL alphabet to text.
  - Or JanFizza Bukhari from National University of Sciences & Technology who designed a similar ASL alphabet translation glove.
- There is only a small amount of research done for translation gloves or software to translate full words such as by
  - J. Hernandez-Rebollar from George Washington University where a glove was designed to translate ASL to text on a small scale; however, was bulky and had to be fitted upon the entire arm.
- It is necessary to note that there has been no research done on designing a glove to translate words to text that takes into account facial expressions into the algorithm, an absolutely vital component of ASL.
- Thus, throughout the project, the purpose will be to design a glove and software algorithm that translate American Sign Language to text while allowing for the input of the user's facial expressions.

Materials and Methods: Hardware Design

Flex Sensors - There are 5 flex sensors connected to the back of each finger to collect voltage data on the bend of each finger. As the sensor straightens, there is less resistance and a higher voltage. Naturally, each sensor has a different resistance when made, and the data will be normalized to account for the varying voltage scales from fully bent to straight.
MPU 6050 - This serves as an IMU and is positioned at the bottom of the glove collecting data on the 6 degrees of freedom, specifically the roll, pitch, and yaw, and the acceleration of the glove in three axes. The data will be used to measure the hand positioning and path of motion of the glove.
Arduino Uno - connected to a laptop, the Arduino delivers a 5 volt supply to each of the flex sensors connected to the breadboard, and to the MPU 6050. It collects data from each of the sensors which are then appended to a serial monitor and text files at a rate of 19200 Hz utilizing Putty.
External Materials - The yellow bands are put in place in order to allow the flex sensors to bend and stay in place with the finger rather than snapping. The clear substance is an adhesive that holds the flex sensors and the MPU 6050 in place.

Materials and Methods: Software

Pre-Data Collection

- Ten signs were used: Hello, Deaf, Silly, Don't, Sorry, Funny, Very Funny, Fine, Hearing, Same.
- The signs were chosen such that if the model proved successful, then it would be able to scale. The signs include a diverse range of motions, with some signs such as Same and Silly having very similar finger bends and slightly different paths of motion. Funny and Very Funny are examples of signs that have the exact same motion but different meanings based on the facial expressions of the user.
- 50 trials were conducted for each sign.
  - For each trial, data during the period of the sign was continuously collected for each flex sensor and the six degrees of freedom.
  - For each of these data markers, an average was taken from the data points collected and recorded in a text file for that data marker.
  - At the end of the 50 trials, another average is taken from those 50 trials to obtain a final average.
- Common ASL facial expressions were quantified on a scale of 0.0 to 1.0

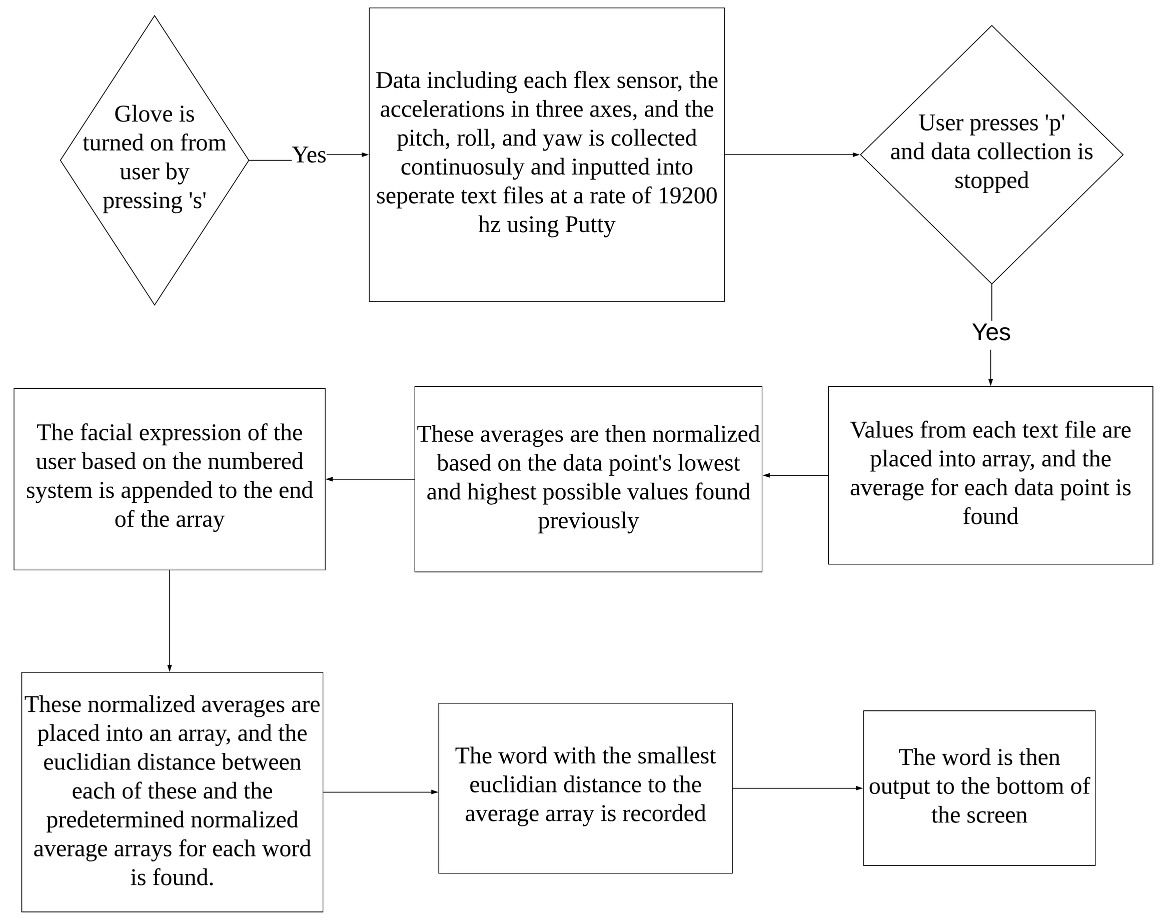
Anger	0.0
Fear	0.1
Disgust	0.2
Depressed	0.35
Sad	0.4
Neutral	0.6
Happy	0.8
Joy	0.85
Excitement	0.9
Surprise	1.0

$$Distance(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

x - the array of each preset averaged data point found for a sign
y - the array of each averaged data point found during a live trial

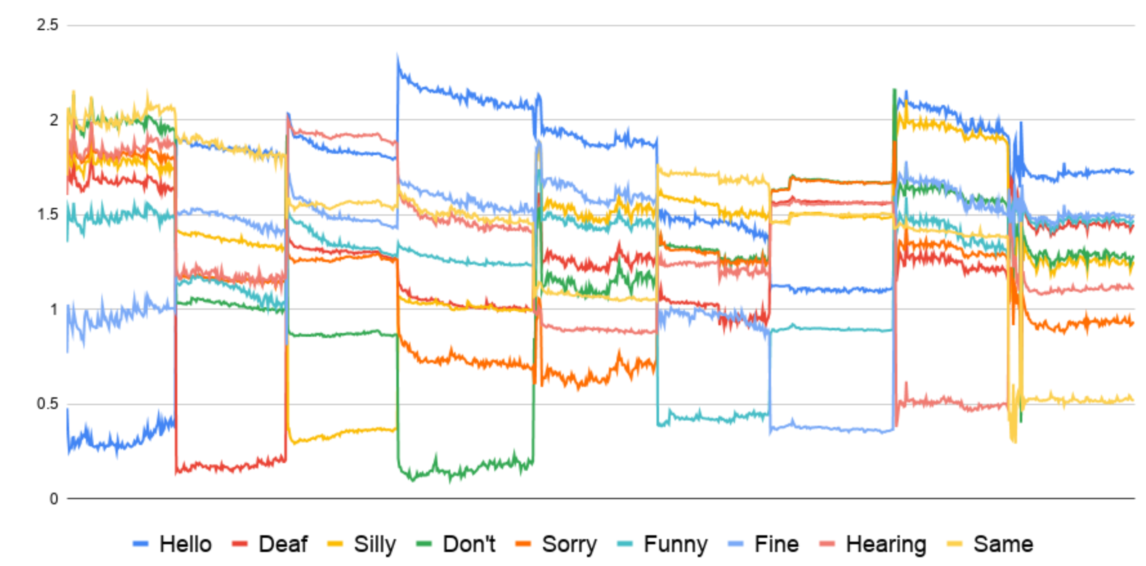
$$Confidence(x) = \frac{x_1 - \sum_{i=2}^{q}\frac{x_i}{q}}{max - min}$$

q - number of signs being evaluated
max- highest recorded euclidian distance. min is always 0.
x - array of euclidian distances, with x1 representing the translated sign

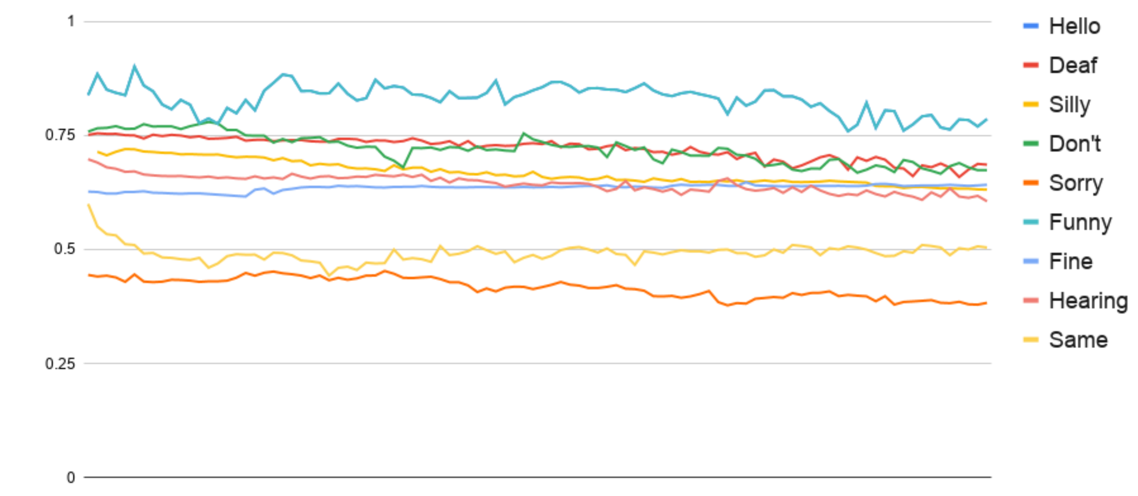


Results and Interpretation

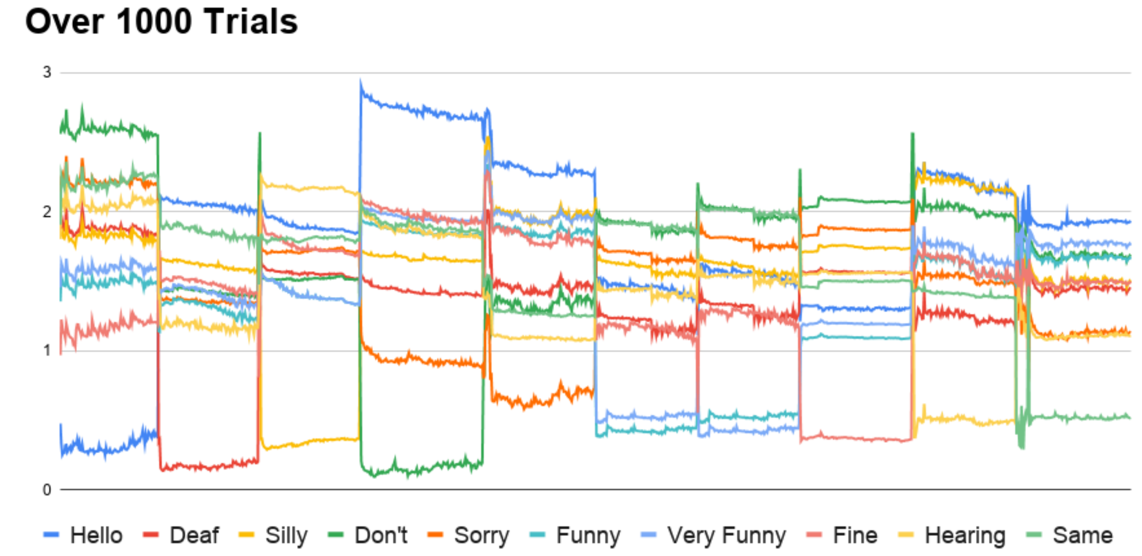
Euclidian Distances of Ten Signs Over 1000 Trials



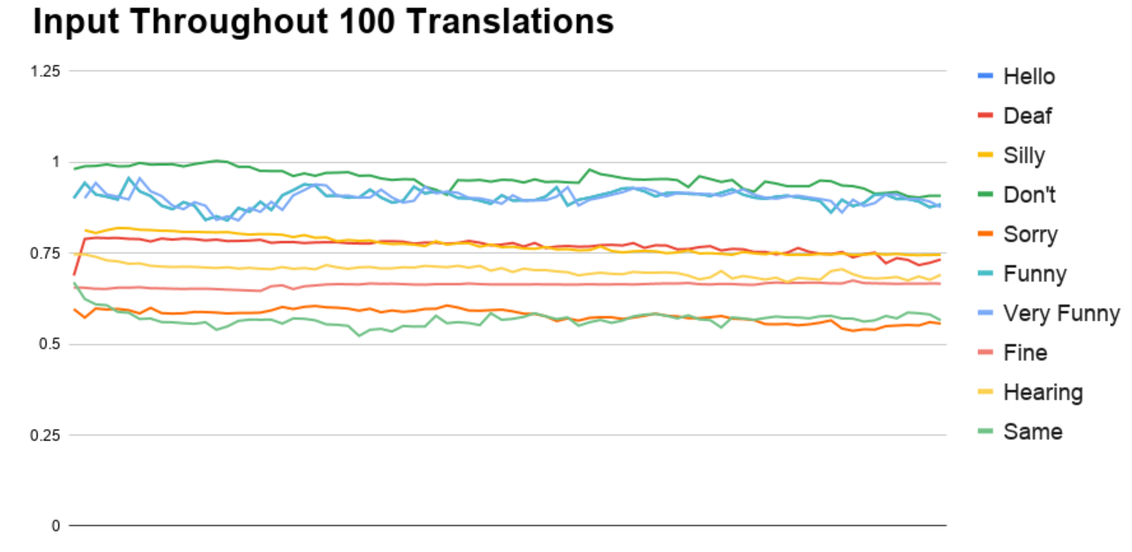
Algorithm Confidence for Each Sign Throughout 100 Translations



Euclidian Distances of Ten Signs With Facial Expressions Input Over 1000 Trials



Algorithm Confidence for Each Sign With Facial Expressions Input Throughout 100 Translations



Results and Interpretation Continued

- 1000 trials were conducted with facial expressions being used and without. The confidence for each translated sign was also graphed for each. The euclidian distances of each sign's preset averages to the averages during the trial were graphed as the y-axis. And the signs: hello, deaf, silly, don't, sorry, funny, very funny, fine, hearing, same, had 100 trials conducted for each in the x-axis for that order.
- "Funny" and "Very Funny" could only be differentiated between utilizing their facial expressions, which is why "Very Funny" was not in the graph with facial expressions.
- 99.7% accuracy was achieved for the model with and without facial expressions. Three signs were mistranslated during the 100 "hearing" trials.
- The gap in the distance graphs between the translated sign's euclidian distance to each of the other sign's distances also shows the confidence of the model for each translation.
- Normalization also proved essential as various data markers such as the pitch, roll, and yaw naturally had unequal weight on a 0°-360° scale, while the flex sensor were on about a 0 V - 5 V scale. The algorithm also proved efficient not requiring large amounts of data for each sign to be trained, and only an average of 5 minutes to register a new sign.
- Although the model using facial expressions had the same accuracy, referring to the confidence graphs, the confidence for each translation increased significantly.
- Words such as "Don't" and "Funny" had high confidence levels because of their strong difference when compared to the hand orientation and movement of other signs.
- Similar words such as "Sorry" and "Silly" with very similar finger bend averages, still were translated with high confidence from about 60% to 80% using their hand motions.
- As a whole, the model with the use of facial expressions proved successful at a 99.7% accuracy with confident predictions about 60% to near 100%, and the use of facial expressions proved to increase the confidence of predictions along with differentiating between words with equal hand positioning and movements.

Conclusions

- Throughout, the purpose was to design a glove and software algorithm to translate American Sign Language to English text while taking into account the facial expression of the user.
- The algorithm found the averages of various data points: the finger bends, the acceleration in three axes, the angle of rotation around three axes, and the quantified facial expression. It then compared the euclidian distance of these averages to the preset averages found during experimentation for each sign. The smallest distance would indicate the translated word.
- The model using facial expressions proved successful by achieving a 99.7% accuracy over 10 signs, outperforming the current research done on ASL gloves which only translate the alphabet and the research done on attempting to translate words.
- The algorithm proved to be scalable as the 10 signs included diverse and entirely different signs along with signs that had very similar hand positions and motions with slight differences, along with signs that were exactly the same but with different facial expressions. With this, the model still showed high confidence from about 60 to about 100% in translating each sign.
- The algorithm's success also shows the viability of using facial expressions as another data point to translate ASL to text, allowing a key component of the language to be accounted for in translations.

Relevant Applications to Biotechnology

- The algorithm was recently scaled to 40 words, and maintained an accuracy at 99.3% further showing its scalability, the highest out of current literature.
- The glove and algorithm enable individuals that do not understand ASL to better understand hard of hearing family members or friends, where there might have been a greater disconnect before.
- The glove also takes into account facial expressions, an absolutely vital component of ASL which is the major reason for why no previous translation systems have gone to market.
- Additionally, the algorithm in the future could be scaled and altered such that it delivers live translations. This can be done by using the same algorithm to constantly analyze incoming data, but only deliver a translation when the confidence level is greater than for example, 90%.
- The algorithm could be utilized within a software application that could detect facial expressions through machine learning and connect to the glove via bluetooth.
- Additionally, although based on the algorithm it should not have an effect, the speed at which an individual signs at and its effect on the accuracy of translations can also be analyzed.

Acknowledgments

I would like to thank my parents who supported me throughout the project by purchasing necessary materials and equipment that was essential in conducting the research successfully. Also to Mr.Wallace for providing insightful tips and suggestions to improve the project along with greater knowledge on various essential physics topics to better the results of the algorithm.