Asphone – The Sign Language Interpreter Module

Pushpakumara N.P.S.S.
Department of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Nayanasara M.G.G.S.
Department of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Jayasekara L.H.M.
Department of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Ayodhya Shansala P.G.
Department of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Jagath Wickramarathne
Department of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Wishalya Tissera
Department of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

ABSTRACT
According to a survey conducted by the World Federation of the Deaf, the number of deaf-mute people in the world is constantly increasing, and there are currently 72 million deaf-mute people worldwide [4]. 63% of them have become impaired by birth itself whereas the rest have been impaired due to accidents. The primary mode of communication for these deaf-mute individuals is sign language which is accepted and used universally but communication between a typical person and this deaf-mute someone with a disability can be very challenging since the average person cannot comprehend the hand movements of a disabled person, and deaf-mute individuals frequently have hearing difficulties. This study focuses primarily on a real-time sign language interpretation technology that can be utilized to sustain effective communication between deaf-mute and hearing individuals. It includes a wearable glove and a "Portable Display Unit" (PDU). On the display of the PDU, the sign gestures done with the glove will appear as a readable plain text message. In addition, the "Asphone - Sign Language Interpret Module" which is another part of the system is capable of translating human voice messages into sign language. These translated sign language messages are also displayed on the screen of the PDU. Using this technique, deaf-mutes can communicate effectively with everyone they encounter in their daily lives. With the completion of this project, the world will be able to bridge the communication gap between deaf-mute and hearing individuals.

Keywords
PDU, Asphone, deaf-mute, Sign language.

1. INTRODUCTION
Studies show that sign language is a language commonly used in the world where more than 72 million people around the world who are visually and auditorily impaired use the language to communicate [1]. But the challenge is it is very rare that a normal person who is not impaired uses and gets to know sign language which makes it a challenge to communicate with each other [2]. To have a feasible solution to the said issue, a system is developed to cover the whole 360 of the problem and to make the lives easier for both the impaired people and normal people where they can communicate with each other without any trouble.

A real-time sign language interpretation device is the primary focus of this study's research solution to facilitate continuous two-way communication between a deaf-mute and a hearing person.

This system is inclusive of two main which are a “Gesture Capturing Glove” which is wearable and a “Portable Display Unit” (PDU). The Gesture Capturing glove captures the sign gestures and the Portable Display Unit displays the sign gesture received in plain text which can be read by anyone. There is another module which is the “Asphone - Sign Language Interpret Module” which can convert human voice messages into sign language animations. The PDU screen also shows the translated sign language messages. This method allows deaf-mute persons to communicate effectively with everyone they meet. If this study is fruitful, it will pave the way for deaf-mutes and the rest of the world to finally be able to communicate with one another. This research project mainly consists of four major components,

- The Gesture Capturing Glove
- Language Translation Process
- Continuous Sentences Interpretation System
- Portable Display Unit (PDU)

2. BACKGROUND REVIEW
There have been a lot of studies centered on automatic sign language interpreters. Computer vision and recognition techniques based on sensors are two primary areas of interest. All of these studies rely heavily on image processing techniques that begin with a camera-based image or video capture [3]. Images are used to analyze data, and then that data obtained is filtered using various pre-defined algorithms in these products. High-resolution cameras, which are used to capture the most detailed hand motions, have a greater need for storage space. To use these items, a deaf person must have access to a high-performance Capturing device, such as a camera which sometimes can be challenging for them [3].

There has been an additional study on the topic of a sign language recognition system that makes use of hand gloves [4]. In this scenario, the deaf-mute wears a glove with five flex sensor panels, and a motion tracker is attached to each finger [5]. Fingertips are the source of the data, which is then filtered using sensor information. The success of this study is predicated mostly on neural network principles that boost the effectiveness of the tool. This method has a quick turnaround time and can be put to use in time-sensitive situations. The wearable glove's sophisticated hardware modules limited its usefulness in real-world applications.

Research into this field of translation has uncovered attempted translations of spoken language into sign language. Yet this kind of testing are conducted in either Chinese or Russian [6]. They have not yet established the method of converting sign language into text, and instead, just translate human voice into sign language. This is known to be the root of many communication problems which is identified as a weakness. When attempting this translation procedure directly on video without gloves, there is a greater chance that sign language will become blurry [7]. Another issue is that you can't do any translation work without an active internet connection. This is because there is a lack of internet cafes and other such facilities in many areas. You can get more out of this device if you use it in offline mode.

3. METHODOLOGY

Asphone consists of two main physical devices which are the wearable Gesture Capturing Glove (GCG) which will be worn and used by deaf-mute people and the smartphone-like screen which is the Portable Display Unit (PDU) that will convert the gesture done by the deafmute person wearing theglove would make into plain text. As an answer to the given gesture the person receiving the message can respond to the deaf-mute person in words where the spoken words will be captured through the voice signals and the message will be interpreted as a message through sign language and will be displayed on the PDU screen. The application uses machine learning to the first cluster which will be fed to the system to input the gestures and to carry out the process where the application is developed on Machine learning and Python mainly. The used language in this system will be English which is the universal language that is most abundantly used in the world. This is a two-way system where the deaf-mute person can ask the question through the system and the person can answer the question with the relevant response which will be converted to animation so that the deaf-mute person can understand the vocal response.

Given below is a detailed elaboration of each component of the system.

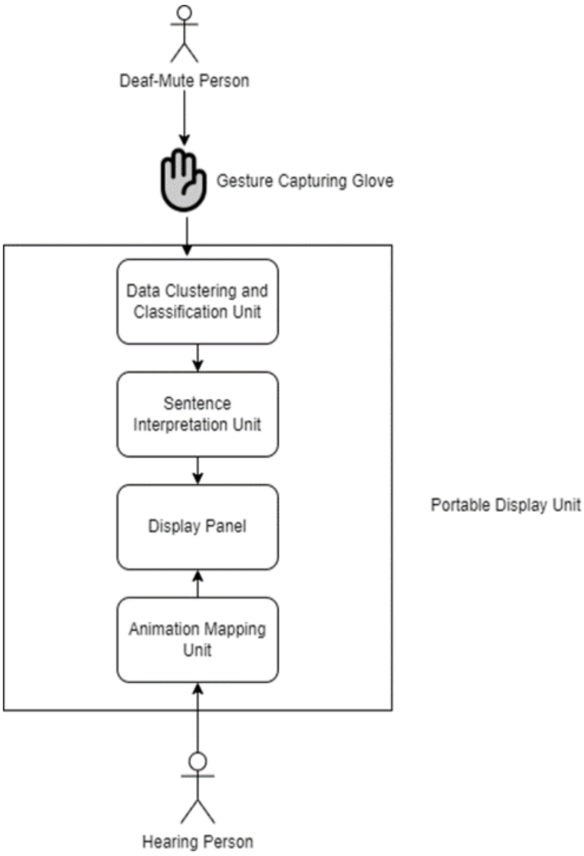


Figure 1: High-LevelSystem Diagram

3.1 Gesture Capturing Glove (GCG)

This is one of the main components of the Asphone where which is the key major input device that is used in the whole system [7]. A push button is used to input the boundaries of a single data frame corresponding to a certain sign in the data stream. A set of flex sensor panels are used to capture the gesture inputs of the finger movements and an accelerometer is used to capture the movements of the hand. With these two methods, the system can obtain two-dimension and three-dimension data of the sign inputs. After a set of data is recorded for a single sign language gesture it is stored in a CSV file. This way different sets of data are collected and stored (about 100) and those symbols are referred to as the mean dataset. Through the WIFI direct technology, the CSV file will be sent to the PDU where the system can be used offline or in network conditions. The data transmission process in the system is done by using a Node-MCU microcontroller which is located inside the Gesture Capturing Glove and the mean value calculation, and MSE error calculation process will be started after receiving the data from the PDU.

After getting the CSB file the PDU will calculate the meanvalue for each sign and then the mean square error (MSE)will be calculated. Through this, the system will monitor themost accurate signs recorded. If the MSE value is equal or less than 0.5, then the system will accept the sign as a valid sign.

$$\mathbf{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y)^2$$

- N - Total number of Data points
- yi - Actual output value
- $\hat{y}$i- Predicted output value ˆ
- yi-$\hat{y}$i - The absolute value of the residual

Also, a pre-trained model has been used to determine the optimum number of signs can be usedin this research project. Initiallyidentified 40 sign gestures that are mostly used in sign language, and labeled them from index 0 -to index 39. Then recorded 53 variations of data points from each sign. The following diagram indicates the finger variation output of the sign index 29, which is spread across 53 data points.

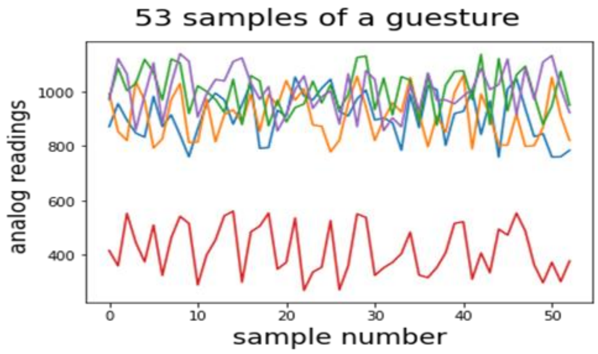


Figure 2 – Analog Behavior of Sample Gestures

Those recorded data points are fed to a K-Means clustering model, where can determine the optimum number of data clusters can be used. In the K-Means clustering model, it could determine 29 different signs can be used use in this research, and the other 11 signs were neglected as trash data.

The used technology for the system is

- Arduino
- Python
- K-means clustering
- Sklearn
- Seaborn

3.2 Language Translation Process

Through this part, the input message will be evaluated whether it's a sign gesture or a voice and if it is a sign gesture then this part of the system will be converting the numerical data that will be received into plain text. In the same way, once the voice is input to the system by detecting the silence of the part of the sentence that is spoken, the system will be splitting it into different parts and will convert it to text individually. If it is mostly silence recorded, then the data will be disregarded by the system. Acoustic, language and dictionary are used to convert the audio to text and after the process, the data is sent to the database. CMU sphinx is used for voice recognition.

Technologies used,

- Python
- Raspberry Pi
- Google Colab
- ML process

3.3 Continuous Sentences Interpretation System

In this art the words spoken or gestured using the glove will be arranged meaningfully to make sentences combining a bulk amount of words. This is a signal translator where which is a real-time device where it has a signal binocular paired with the glove as a signal converter device that can be seen outside.

Using the break-point algorithm and end prediction algorithm, one by one the meaning of the signal gestures is displayed on the screen as sentences. Python is used in the breakpoint analysis to detect the continuation of the sentences where it will add up word connectors to the sentence. First, the character list stored in the initial string will be observed and then the stored characters will be mapped with words using the word pool and will send the mapped words to the PDU where it will output meaningful sentences. The end prediction algorithm will mark the end of the sentence that is made by the breakpoint algorithm.

Technologies used for this development are,

- Python
- NLP

3.4 Portable Display Unit

This is the major output device of the system which is portable. This will showcase the animation related to the relevant words gestured pairing up with the gesture-capturing glove.

The technologies used for the development of the PDU are,

- Raspberry
- Mimix3D
- Blender
- Natural Language Processing
- Wondershare Filmora
- Python

For the environment setup, the Jasper offline platform should be installed. This way the device will work offline. This device will have two-way communication where the message gesture by the deaf-mute person will be displayed on the screen and the other way the response by the other person will be recorded and will be communicated to the deaf-mute person where the conversation can be continued.

4. TEST DATA

Taking a live scenario into consideration the technical process is elaborated below. Here the scenario of a deaf-mute person asking for direction from the bank is taken as the exemplary situation.

01. The clusters are identified as K= 29 (No. of clusters) where through machine learning this value is taken where even though the assumed number of gestures is 40 the sample that is taken to the system is 29 gestures.

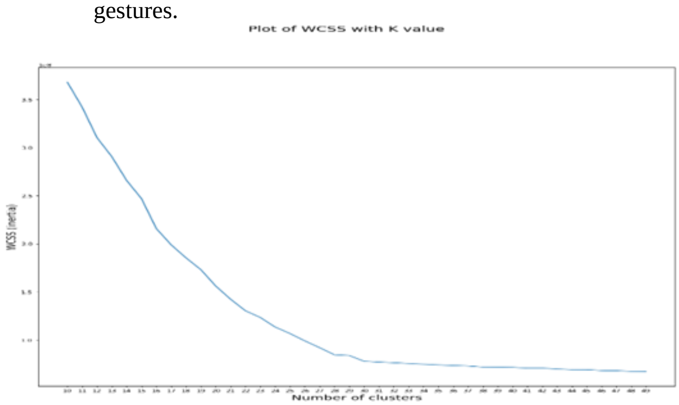
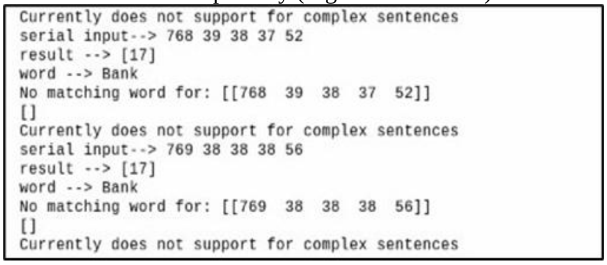


Figure 3 – Determining the K value

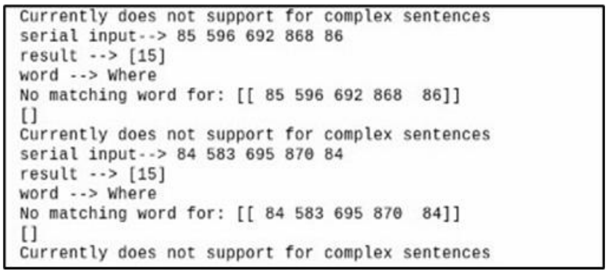
02. This data is pre-programmed to the glove to identify the sensor values of the given hand gestures for the 2 gestures made (Where/Bank) by the deaf-mute person.

03. The identified gestures are sent to the language processing unit, where the two words will be identified separately (E.g. - Where/Bank)



```
Currently does not support for complex sentences
serial input--> 768 39 38 37 52
result --> [17]
word --> Bank
No matching word for: [[768  39  38  37  52]]
[]
Currently does not support for complex sentences
serial input--> 769 38 38 38 56
result --> [17]
word --> Bank
No matching word for: [[769  38  38  38  56]]
[]
Currently does not support for complex sentences
```

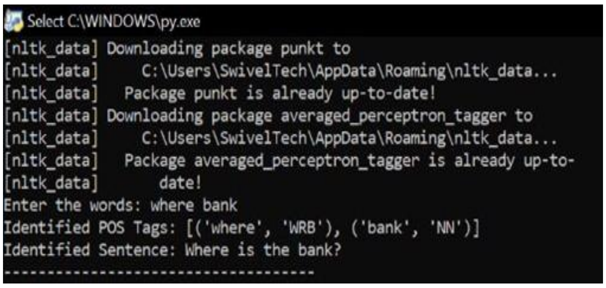
Figure 4 – Sample Word Interpretation 1



```
Currently does not support for complex sentences
serial input--> 85 596 692 868 86
result --> [15]
word --> Where
No matching word for: [[ 85 596 692 868  86]]
[]
Currently does not support for complex sentences
serial input--> 84 583 695 870 84
result --> [15]
word --> Where
No matching word for: [[ 84 583 695 870  84]]
[]
Currently does not support for complex sentences
```

Figure 5 – Sample Word Interpretation 2

04. The identified two words will be converted into a meaningful sentence by completing it using the connector words (is/the) and marking the end of the sentence by the Continuous Sentences Interpretation System. (E.g.- Where is the bank)



```
Select C:\WINDOWS\py.exe
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\SwivelTech\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     C:\Users\SwivelTech\AppData\Roaming\nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
Enter the words: where bank
Identified POS Tags: [('where', 'WRB'), ('bank', 'NN')]
Identified Sentence: Where is the bank?
-----------------------------------
```

Figure 6 – Sample Sentence Interpretation

05. The PDU will showcase the question created by the Continuous Sentences Interpretation System. After the question was displayed by the process elaborated above by the system the response monitoring will be carried out which the deaf-mute person will get the answer to the question asked.

06. The vocal response given in return will be recorded by the PDU.

07. The voice input will be converted to a natural language text.

08. The text recorded will be displayed to the deaf-mute person on the PDU screen as an animation.

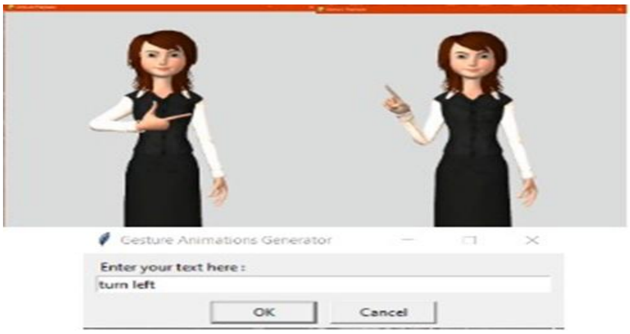


Figure 7 – Sample animation output

5. CONCLUSION

It is obvious around the world with the advancement of technology there are a lot of supportive systems invented for the betterment of certain types of people. Deaf-mute people are given major attention these days, especially with the social media influence and equality concept practiced around the world. There are several systems and devices invented to support deaf-mute people around the world to support their day-to-day life, to make their life easier, and to give them the experience of living a normal life. Because of the complexity of the language, they are using it has become quite a challenge to develop a comprehensive system for them to cover the 360 of their need for communication both two ways. Therefore, as a result of this considering the available studies, Asprone is developed to support the deaf-mute community making it easier for them to communicate and for others to respond to them.

6. REFERENCES

[1] D. O. J. F. -. Srilanka, "UN Universal Periodic Review - Sri Lanka," Third Cycle, Colombo, 2017.

[2] S. L. Government, "Department of Census and Statistics," Department of Census and Statistics, 2011. [Online]. Available: http://www.statistics.gov.lk/Other_CensusandSurveys. [Accessed 30 12 2021].

[3] S. L. D. O. J. Front, "UN Universal Periodic Review - Sri Lanka," Third Cycle, Colombo, 2017.

[4] H. N. Jan Bungeroth, "Statistical Sign Language Translation," Lehrstuhl fur Informatik VI, Germany, 2011.

[5] H. N. Jan Bungeroth, "Statistical Sign Language Translation," Lehrstuhl fur Informatik VI, New York, 2011.

[6] A. K. A. T. Shivakumar K.M, "Kannada Speech to Text Conversion," Amrita Vishwa Vidyapeetham University, Mysore campus, Mysuru.

[7] D. o. C. a. Statistics, "Department of Census and Statistics," Department of Census and Statistics, 2011. [Online]. Available: http://www.statistics.gov.lk/Other_CensusandSurveys. [Accessed 30 12 2021].

[8] A. Deaf, "AI Deaf," December 2021. [Online]. Available: https://aideaf.com/.

[9] E. Dreams, "Sign Language Interpreter Education Programme 2021 – Modules," Equal Dreams, 2021. [Online]. Available: https://equaldreams.sg/sign-language-interpreter-education-programme-2021-modules/.