A Proposed Method for Translating Sign Language to Text and Speech Using Deep Learning and Natural Language Processing Techniques

Tahsin Maisha Rahman
Department of Computer Science & Engineering
BRAC University
Email: tahsinmrahman@gmail.com

Mohd. Hamim
Department of Computer Science & Engineering
BRAC University
Email: mohd.hamim@g.bracu.com.bd

Abstract

Hearing-impaired individuals find it difficult to convey their message to hearing individuals due to a lack of understanding of the English language itself. While a hearing person learns a new word by listening and speaking the word inside while a non-hearing person learns a new word by associating a visual body. Hence, it is very difficult for a non-hearing person to speak English as fluently as a hearing one. We propose a wearable device based on finger movement sensors, support vector machine classifier and attention-based transformers to assist a non-hearing person to convey their messages into either voice or text format.

Keywords: support vector machine classifier, natural language processing, transformers, sensors

1. Introduction

According to World Health Organization (WHO), there are approximately 466 million persons with disabling or hearing loss in the world and the number has been predicted to rise to 700 million people by 2050 [1]. In addition to their hearing impairment, it is difficult for them to communicate effectively with non-hearing impairment individual and face challenges in understanding the English language itself. What could be the reason? According to psychologist, David Ludden, hearing individuals decode a written symbol into speech sounds to learn the written text [2], whereas hearing-impaired individuals try to associate each word with a sign language making it difficult for them to communicate verbally. The hearing-impaired population use sign language for communication which has its own set of grammars at its core and is different from the spoken language. Hence, it is only normal for the hearing population to understand the language of the non-hearing population leading to a major communication barrier.

2. Proposed Model

We propose a project that would translate sign language into spoken English language using Support Vector Machine (SVM) classifier and Attention-based NLP models - Transformer. A flex sensor has been used to detect finger movement and motion for each word in the American Sign Language (ASL). These analogue signals are processed and converted to digital signals using an ADC converter. Digital data obtained are fed into the SVM classifier which is trained to classify each digital signal into corresponding words. Words, numbers, or sentences obtained are fed to the NLP model which has been trained to understand the American Sign Language and can translate them into English sentences or words.

3. Implementation of the Proposed Model

3.1. Data Set

For this project proposal, we aimed to use the American Sign Language database because of its availability of availability and pre-defined grammatical structure which are easy to understand. For classification training, we plan to acquire the data set with the help of five different individuals who will use the American sign language for at least 100 words, and we will manually annotate each of them into their corresponding English words. To reduce inaccuracies, everyone will repeat sign language 5 times. Following that, the signal will be averaged to acquire the most important feature.

For training the NLP model, we plan to use the American Sign Language database which will be

analyzed by the model and trained to convert ASL into spoken English words.

3.2. Model Implementation

The following figure illustrates the entire process in a flow-diagram format

Flex Sensors
Detect finger movements and orientation

↓

Signal Amplifiers

↓

Low-pass filters

↓

Microcontroller with ADC and Bluetooth Transceiver

↓

Feature Extraction

↓

Multi-Class SVM Classifier

↓

Words, senteces, numbers

↓

NLP Model (Attention based models - Transformers analyses the ASL and has been pre-trained to conver ASL into Spoke English Language

↓

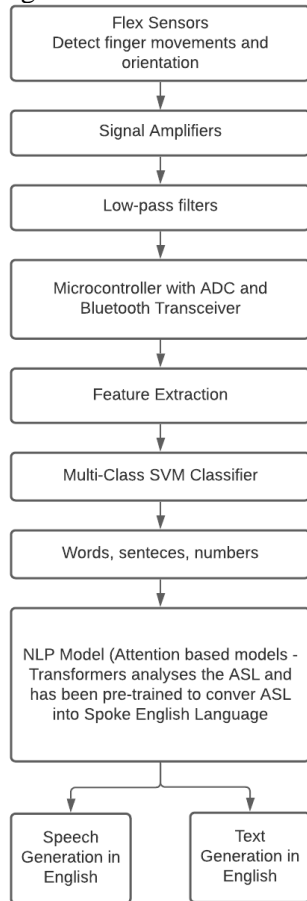Speech Generation in English   Text Generation in English

Figure: The proposed architecture of the solution.

Each component in the block performs a specific role which has been described as follows:

- **Flex sensors**: resistance-based sensors that can be deployed at the top of each finger and its movement is measured whenever a sign will be made
- **Signal amplifiers**: to amplify the signal acquired
- **Low pass filter** removes any redundant signals from the acquired signal
- **Microcontroller with ADC and Bluetooth transceiver:** the filtered signal will be converted to a digital signal and is wirelessly transmitted to a computer or any remote device for processing
- **Feature extraction**: only the principal features of the entire signal would be extracted to train the classification model

- **Support Vector Machine (SVM):** the model is trained to classify acquired signals or extracted features into their corresponding sign languages. The output could be either in words or numbers
- **Transformers (attention-based model):** the model is pre-trained to convert the American sign language into spoken English language. After it receives an output from the signal classifier, the NLP model organizes the acquired words into ASL format and translates them into the English language

4. Conclusion

This project proposal aims to assist hearing-impaired individuals with the ability to communicate with hearing individuals. Initially, we expect the model to easily classify words or numbers and provide the corresponding output. However, the classifier would provide an output one word at a time and this could be challenging for the NLP model to organize those words into a sentence the individual wanted to convey. Other challenges are, acquiring exact information from the sensors, understanding the endpoint and start point of sign language and many more. Despite these challenges, this proposal holds a great potential to reduce the communication barrier between hearing and non-hearing persons.

References

1. World Health Organization, "Prevention of deafness and hearing loss". Available Online: https://tinyurl.com/3ktyhmyj
2. David Ludden PhD (2015) "Can you read a language you can't hear?", Psychology Today. Available Online: https://tinyurl.com/2e8nn2fs
3. Zhou, Z., Chen, X. et al. Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays. Nat Electron 3, 571–578 (2020). https://doi.org/10.1038/s41928-020-0428-6
4. Sampada S. Wazalwar & Urmila Shrawankar (2017) "Interpretation of sign language into English using NLP techniques", Journal of Information and Optimization Sciences
5. Ashish Vaswani, Naom Shazeer, Nikki Parmar, X. et al. "Attention is all you need", Advances in Neural Information Processing Systems 30 (2017)
6. Sung-Woo Byun and Seok-Pil Lee, "Implementation of Hand Gesture Recognition Device Applicable to Smart Watch based on Flexible Epidermal Tactile Sensor Array", MDPI