AI TTS Smartphone App for Communication of Speech Impaired People

Hanyoung Lee and Deawoo Park

Abstract Due to COVID-19 in August 2020, the announcement by the Central Disaster and Safety Countermeasure Headquarters is being made regular every day. Sign language interpretation for the hearing impaired is broadcast on TV next to the speaker, but language communication for each individual for specific quarantine is insufficient. In this paper, AI construct motion recognition of sign language interpretation as Big Data, extract features of motion recognition data, and apply it to AI machine learning. Design an app to communicate with the hearing impaired on a smartphone owned by the hearing impaired. When a hearing impaired person writes TEXT or signs language on a smartphone, the smartphone recognizes it. Sign language motion is converted to text through motion recognition. The converted text is delivered to the general public as text to Speech through AI voice recognition. When an ordinary person speaks on a smartphone by voice, it is converted into text and displayed to the hearing impaired. This study will be used as a means of transmitting information to the hearing impaired in the Untact era. Develop a method for speech-impaired persons with disabilities to communicate with ordinary people who have not mastered sign language, using 4th industrial technology.

Keywords Artificial intelligence · Mobile · Big data · Motion recognition · Machine learning · TTS (Text to Speech) · STT (Speech to Text)

1 Introduction

The announcement by the government’s Central Disaster and Safety Countermeasures Headquarters due to the Corona 19 (COVID-19) is being made on a daily basis through television and Internet media. With advanced fourth industrial revolution technology, wireless communication and internet are developed in areas of COVID-19 infected people to deliver information. A smartphone alone reveals the
movements of the COVID-19 confirmed patient and receives COVID-19 information through map search and AI voice search voice commands. At this time, disaster measures are being delivered on TV through sign language for the deaf next to the presenter of the Central Disaster and Safety Countermeasures Headquarters. The obligation to deliver information to deaf people is important enough to be specified by law.

However, communication is not easy for deaf people in real life. This is because sign language interpreters must accompany them in order to communicate with non-disabled people who have not learned sign language. As a result, communication between deaf and non-disabled people is subject to great restrictions. In a rapidly changing society, privacy is unprotected and only a limited number of people can talk to them, so they can be placed in blind spots of information. In addition, AAC (Augmentative and Alternative Communication), the existing supplementary alternative communication, was developed to meet the standards of people with developmental disabilities who use picture cards of simple word combinations. Like Fig. 1, there is an overwhelming majority over other types of disabilities, but the research required for deaf people is insufficient. After meeting sign language interpreters and hearing impaired people across the country, there is a need and demand for optimized AAC that allows hearing impaired people to easily communicate with non-disabled people [2]. This requires practical research to facilitate communication between the deaf and the non-disabled.

In this paper, AI machine learning is applied to hearing-impaired people’s smartphones to design APP for communication. When hearing-impaired people use TEXT on their smartphones or sign language, their smartphones recognize their movements and find the right words. In other words, sign language movements are converted to TEXT through motion recognition. Translated TEXT is expressed as sound to non-disabled people through TTS (Text to Speech). Conversely, when a non-disabled
person speaks to a smartphone through voice, it converts it to TEXT (STT: Speech to Text) and shows it to deaf people.

This paper research will embody the methods of communicating with ordinary people who have not mastered sign language in the Untact era through AI, and will be used as a means of information communication and smooth communication.

2 Relevant Research

Sign Language Law: The sign language law aims to improve the language rights and quality of life of hearing-impaired and Korean sign language users by revealing that Korean sign language is the native language of hearing-impaired people with equal qualifications to the Korean language and laying the foundation for development [3].

Motion Recognition: Four basic elements of sign language (hand motion, hand shape, hand position, hand orientation) that communicate with speech-impaired people are identified. The motion recognition data stored in advance is compared using the similarity method, and the sign language is recognized [4]. Although lip motion, which is recognized by converting to 3D values to increase the accuracy of sign language, is best applied, lip motion cannot be developed into a smartphone application that is easy to use because it can be accompanied by special equipment such as two infrared cameras and three infrared LED (Light Emitting Diode)s. TTS (Text To Speech) & STT (Speech To Text): TTS refers to the conversion of letters, sentences, numbers, symbols, etc. into auditory information generally spoken by people, while STT refers to the conversion of voice into letters, numbers, symbols, etc. TTS technology has the advantage of being able to do both visual-dependent tasks. In addition, it is the easiest and easiest way to deliver information that can change from time to time, and it is a very efficient means of providing information for people with speech and hearing impairments [5].

AAC (Augmentative and Alternative Communication): AAC is a system with four components: symbols, tools, strategies, and techniques. It is divided into low-tech AAC [6] and high-tech AAC [7] by using this to communicate their opinions to others. Even the current state-of-the-art AAC uses symbols (picture cards) for the intellectually disabled, so there is a limit to having conversations for communication that the deaf want.
3 Motion Recognition of Deaf People Analysis and AI Design

3.1 Motion Recognition Analysis on Smartphones

With the launch of new products every few years, smartphones are convenient to use features that reflect the Fourth Industrial Revolution technology. You can search the Internet, play games, e-commerce, and financial transactions that can only be done on existing PCs, and you can call or search information using voice commands because AI voice recognition is possible.

In particular, IoT (Internet of Things) sensors that are detected in smartphones are becoming more intuitive to develop UI than before by using motion, direction, and heart rate check functions. Smartphones have become capable of various functions beyond the role of talking on the phone. Initially, it only detected simple movements through the camera, but now it can make big data for AI. Enter the result value into the repetitive motion of the person entering the camera to make it a meaningful action Fig. 2. In other words, data can be created by entering the sign language used by the deaf. Different people’s hand gestures may make it difficult to extract accurate and meaningful words for each movement, but sign language is already used by people

![Image](image.png)

Fig. 2 Recognize ‘home’ through map learning using Google AI
with speech and hearing impairments, so it forms the same pattern. Entering the sign language data of many people is a way to improve the accuracy of sign language interpretation, as input may vary depending on the size of a person’s body type or movement Fig. 3.

You can use Google AI to learn sign language through instructional learning, and AAC allows you to enter your own sign language directly, so that you can save it in Big Data in the cloud and translate it accurately based on a number of learning samples. This is also available to those who wish to learn sign language, which could ultimately be an opportunity to break down the language barriers of the deaf and the non-disabled.

### 3.2 AI Machine Learning and TEXT Design

You can map and learn sign language movements on AI and build big data to understand the meaning of simple sign language movements and print them out in sentences. Through a combination of learned sign language words, a sentence can be completed by connecting the set words and TEXT the sentence. Figure 4 is
Fig. 4 Two different ways of exemplar extraction

A comparison of the two representative extraction methods. The left side is how to select one single representative for each motion, and the right side is how to select multiple representatives.

Even if one representative is selected in the best way from a number of instructional data, in the end, from a different perspective, only one training data is used for each motion. To reduce the error of sign language, hydration by various deaf people is essential, and as more data builds up, more accurate words and sentences can be generated.

Complete the information collected by Sequence to Sequence, an Encoder-Decoder model using the RNN (Recurrent Neural Network).

In the case of proper nouns, consonants and vowels are recognized individually and the sign language as symbols is patterned as words, so it is only a list of meaningless words when data is first entered. In addition, due to the characteristics of sign language, investigations such as ‘ya’, ‘ah’ are omitted, so the consonants and vowels are combined in order of words and the corresponding surveys are attached to complete the string Fig. 5.

The encoder receives the data and compresses the information into a single vector. The vector at this point is called a context vector, and decoder uses this context vector to complete the sentence shown above. The disappointing point is that the longer the data words entered are listed, the less complete the combination is.

3.3 Designed with AI Machine Learning TEXT and TTS

Sentences produced by motion recognition will be printed in Text to Speech (TTS). Once a character identified by the image is entered, the process of processing the string is carried out. Locate the voice to match the subsequent string in syllables. It then leads to the process of finding syllable rhymes and creating speech fragments. Finally, the pieces entered into the image make a word and complete the sentence with the words, and each of these speech pieces is synthesized in order and then printed out.

The work has been advanced when delivering voice to TTS, but so far there are differences in each TTS program. The core of TTS is how accurately it recognizes text and reads it naturally without any awkwardness.
3.4 **Designed with AI Machine Learning STT**

The ability of non-disabled people to communicate with hearing-impaired people is limited to visual representation. One way to get out of that limitation is STT (Speech to Text).

Anyone can use STT easily with the GCP technology provided by Google. You can run gcloud and Google APIs in a shell window, or you can connect to Google Compute Engine (GCE) and run APIs.

4 **AI App Design for AAC for the Deaf**

4.1 **Multi-AAC Development**

AAC is a necessary skill for many types of disabled people who have difficulty communicating with their language. However, the high-tech AAC, which was developed as a symbol card that allows simple information to be exchanged with simple clicks, is less suitable for people with hearing impairments. Therefore, research and development of AAC devices suitable for hearing-impaired people who can use AI together with symbol cards for simple communication is carried out.
4.2 **Sign Language Recognition Analysis APP on Smartphones**

The AAC, developed as a symbol card for deaf people to have a unique or deeper conversation, has a distinctive limitation. Google AI allows you to expand the scope of AAC, which is limited to recognizing and learning sign language Fig. 6.

4.3 **AI Machine Learning String Learning Analysis APP**

Converts to a string through a learned motion with an input image. Using RNN, complete the information collected by Seq 2seq with sentences and complete with TEXT. When the string is complete, it is delivered to non-disabled people by TTS Fig. 7.

4.4 **AI Machine Learning Voice Learning Analysis & TEXT Display APP**

In the process of recognizing voice and converting to TEXT, words are often omitted or distorted. This is due to the mechanical properties (micro sensitivity, ambient noise) that require continued development of hardware. STT has a difference in performance between a directional microphone and a microphone built into a smartphone. AI can improve accuracy by carrying out a number of studies through learning Fig. 8.
4.5 *Comparison of Information Transfer for the Deaf*

See Table 1.

5 Conclusion

To cope with social risk situations such as COVID-19 and disaster accidents, studies on communication with deaf people are needed. Although existing technologies such as TTS, STT, and motion recognition exist separately, they are not optimized for the role of AAC for the deaf.

The research in this paper aims to develop auxiliary devices that can communicate with non-disabled and hearing-impaired people who cannot sign at all. In this paper, the TTS STT technology is applied to the hearing impaired through the smartphone, combining IoT precedence and AI voice recognition learning technologies in
smartphones using the fourth industrial revolution technology. Through AI machine learning, a smartphone app was designed to learn the behavior recognition of sign language of the deaf and to accurately judge, and to design and suggest ways to use smartphones to communicate with non-disabled people. The study in this paper could protect the privacy of the deaf and have a positive effect on communication with the non-disabled due to the development of this device.

For future research, it is necessary to study ways to directly communicate normal people's language delivery from smartphones through hearing-impaired people’s smartphones and directly using IoT hearing aids or user sensors.
Table 1  Comparison of AAC development for the deaf

|   | Existing | AI | Effect |
|---|----------|----|--------|
| AAC | By symbolic drawing alone limited in use | Communicable with sign language and symbols | Extended communication between the deaf and the non-disabled |
| Motion | Sign language recognition is inaccurate and serviceability is limited | Machine learning to recognize behavior and make more accurate sentences | Basic DB allows disabled people to teach themselves and provides learning opportunities for those who want to learn sign language. |
| TTS | In the context of pronouncing the exact words resulting in heterogeneity | Natural pronunciation | Natural conversation possible |
| STT | High missing words and error rates | Machine learning reduces word missing and typographical errors | Conveniently communicate with non-disabled people |

References

1. datasom. http://www.datasom.co.kr/news/articleView.html?idxno=105467 (2020)
2. Hearing impaired; Gapyeong (Kim and 12 others) Sign language interpreter: (Lim and 12 others) (2020)
3. Korea Legal Information Center; http://www.law.go.kr/law/Korean Sign Language Method (2020)
4. Lee B-H, Lee K-S (2004) 2004 Recognition of finger language using image from PC camera. 102–104
5. Kim T, Kim B, Choi D, Lee Y (2012) Implementation of Korean TTS service on Android OS. 9–16 (2012)
6. Special Education Blog; https://m.blog.naver.com/bjs718/221797984500 (2020)
7. https://www.pacific.edu/about-pacific/newsroom/2018/october-2018/aac-camp-creates-space-for-communication.html (2018)
8. Heo D, Kim C, Kim S, Song B (2019) Comparative study on the speech recognition assistant device for participation in education of hearing impaired. 101–107
9. Kwon S, Choi Y, Lim S, Joung S (2016) A implementation of user exercise motion recognition system using smart-phone. 396–402
10. Yun T-J, Seo H-J, Kim D-H (2017) Text/voice recognition & translation application development using open-source. 426–427