An Instant Sign Language Translation App - HandsTalk

LEE Cheuk Sum, SO Ho Mang Marcus and WONG Ho Leong

Hong Kong University of Science and Technology; {csleeao, hmsoab, hlwongbg}@connect.ust.hk

ABSTRACT

The World Health Organization (WHO) [6] reports that globally, over 430 million people require rehabilitation for their disabling hearing loss, and with over 300 different sign languages used around the world, communication barriers can make interactions challenging for those with hearing impairments. In Hong Kong, the shortage of sign language translators remains a significant issue, with only one translator available for every 3,000 deaf individuals [3]. This is especially concerning given that there are approximately 50,000 people in the city who are either totally deaf or hearing-impaired. As a result, there is a great demand for real-time sign language translation services.

Our mobile app, HandsTalk, which utilizes generative AI and advanced computer vision technology, is the first of its kind in Hong Kong to provide real-time translation from sign language to English using only built-in cameras. By incorporating built-in cameras, it eliminates the necessity for extra hardware or devices (e.g., Leap Motion or haptic gloves) to execute real-time sign language translation, thus enhancing its accessibility to a wider audience. Furthermore, Generative AI is employed to refine the identified words into a coherent sentence to ensure that the translation accurately conveys the intended meaning of the sign language.

Our evaluation showed that HandsTalk is more effective than existing approaches, and the translation process is seamless and does not require any specialized devices such as Leap Motion or haptic gloves. It overcomes the limitations of existing sign language translation systems and fills a crucial gap in real-time communication for individuals with hearing impairments. By providing accurate and immediate translation from sign language to English, it empowers users to communicate more effectively and confidently in a variety of settings. This represents a significant step forward in creating a more inclusive and accessible world for all.

I. INTRODUCTION

HandsTalk aims to improve real-time sign language translation services, empowering deaf and mute individuals to communicate effectively. By promoting inclusivity and connectivity, the initiative aims to enhance the quality of life for millions of people who face communication challenges in our interconnected world.

Here are five clear and concise selling points that showcase the benefits of HandsTalk:

1. **Real-time adaptive translation with advanced computer vision, deep learning, and generative AI.**
2. **User-friendly and easy to use for both sign and non-sign users.**
3. **Leap Motion and haptic gloves are not needed, making it accessible to a wider audience.**
4. **Facilitates seamless communication for sign language users.**
5. **A platform for learning and improving sign language skills.**

II. Methodology

A. Preliminary research and design approach

Our project is unique as it is the first-of-its-kind mobile application in Hong Kong that offers real-time translation of sign language into English. Hence, the development of HandsTalk required extensive communication with various stakeholders, such as academic advisors, sign language users, and sign language agencies. Furthermore, a significant amount of sign language data was gathered from various institutions to train the model. HandsTalk went through a thorough design process that involved conducting preliminary research, identifying problems, setting goals and objectives, analyzing existing solutions, processing collected data, selecting models for training, and creating the overall structural design.

Many existing sign language detection applications rely on specialized devices like leap motion or haptic gloves, which can be difficult for people with disabilities to use. Additionally, traditional sign language translation applications often only translate individual words or terms to sign language, making them more suitable for sign language learners who need to become familiar with sign language syntax. In contrast, we designed our application to utilize only built-in cameras, which are available on most mobile phones, to further enhance its accessibility to a wider audience. It offers real-time translation from sign language to English, enabling individuals with hearing impairments to directly express their intended meaning using sign language. This feature sets Handstalk apart and makes it a valuable tool for those who rely on sign language to communicate.

B. Design Journey

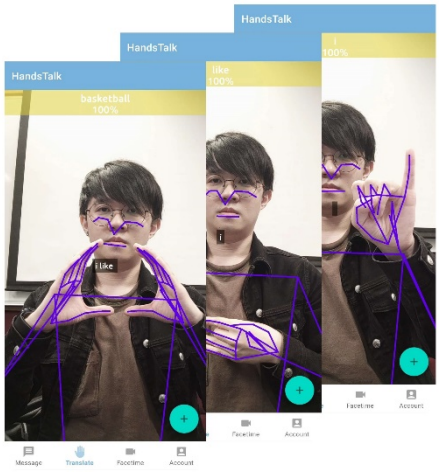
We were motivated by the challenges and findings faced by sign language users and embarked on a careful design process to develop an application that bridges the communication gap between sign language and non-sign language users. This design journey involved brainstorming, ideation, and testing to ensure HandsTalk met the needs of its users. To ensure effective, accurate, and seamless translation, we integrated advanced computer vision technology for sign language detection and generative AI for sentence completion as shown in Figure. 1.

To gather insights, requirements, and understand the challenges faced by sign language users, we conducted literature surveys and interviews with deaf sign language users. These efforts revealed variations in sign languages across regions and countries, as well as differences in sentence logic between sign language and English grammar. To ensure accurate and seamless translation, we proposed and evaluated various deep-learning algorithms in our project to identify the most suitable one. We also recognize the importance of creating an all-in-one channel for non-sign language users to communicate with sign language users, which promotes inclusivity and encourages the learning of sign language during communication interactions with different stakeholders in the project. Our ultimate goal is to enhance the humanity and inclusivity of sign language users in society.

C. Design Principles

The development of HandsTalk is guided by the design principles of user-friendliness, simplicity, and consistency, with the goal of making it easy to use for both sign and non-sign language users. To achieve this, we drew inspiration from other popular communication applications, as well as recent computer vision technologies and deep-learning algorithms to drive our sign language detection framework.

1. Recognize *"Start"* Gesture
Result List: ["*Start*"]



2. Recognize Three "Words"
Result List: ["*Start*", I, Like, Basketball]

3. Recognize *"End"* Gesture
Result List: ["*Start*", I, Like, Basketball, "*End*"]

AI Chatbot Output:
"I enjoy playing basketball (0.85)"

Figure 1. Key Steps Involved in Sign Language Translation

III. RESULT

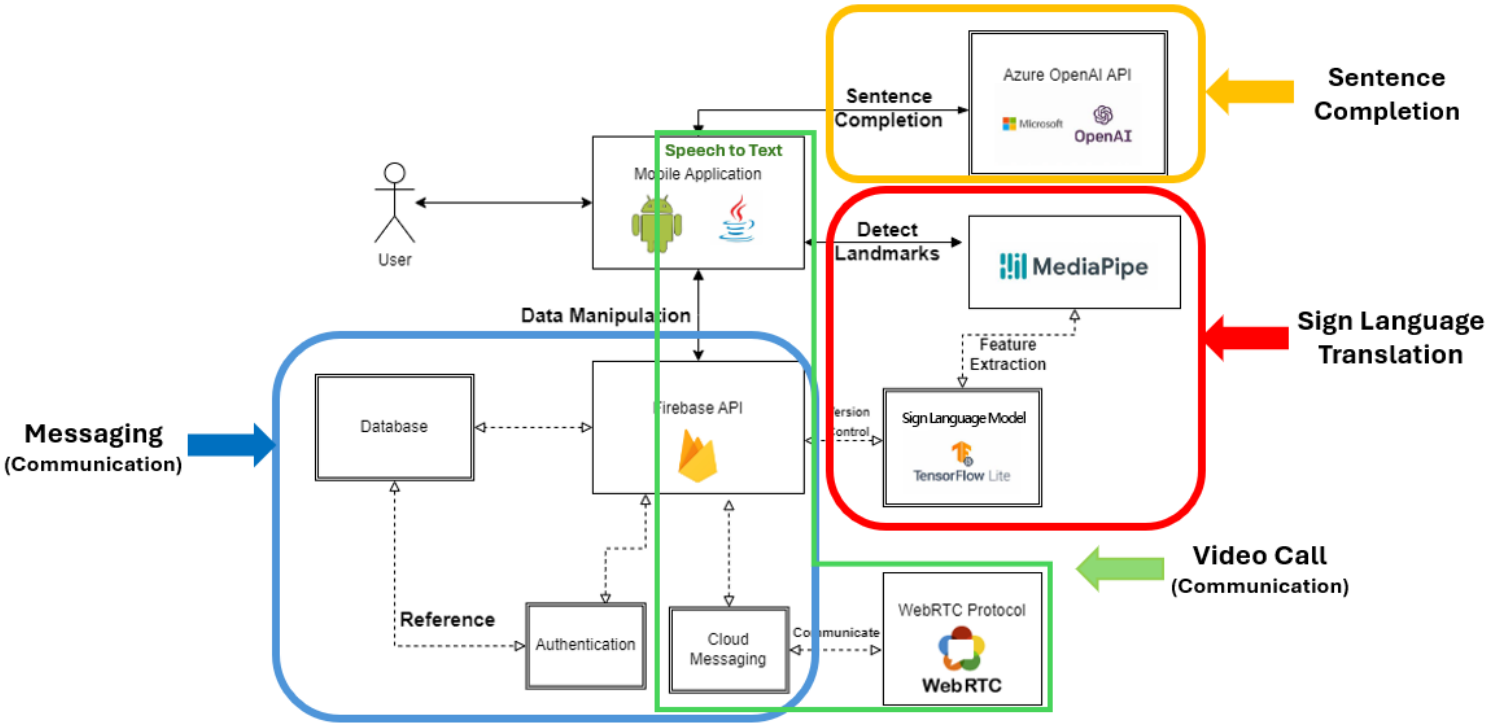


Figure 2: System Architecture of HandsTalk

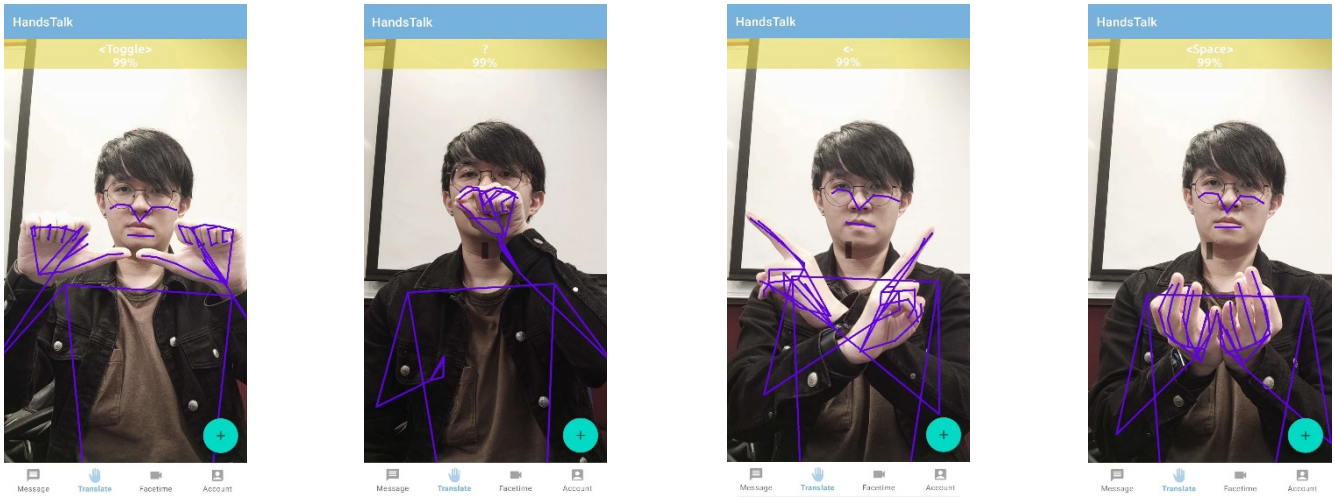
HandsTalk is composed of four main components that work together to facilitate effective communication, including **1) All-in-One User Interface**, **2) Real-time Body Posture Detection**, **3) Sign Language Detection (Deep Learning)**, and **4) AI-Driven Sentence Completion**.

1. All-in-One User Interface The user interface is designed to accommodate both sign and non-sign users, with basic messaging and video call features for communication purposes. Firebase API service [4] is used for account and database management, while WebRTC [7] is utilized for implementing the video call feature. HandsTalk allows standard speech-to-text during calls, which enables non-sign users who may not understand sign language to communicate effectively.

2. Real-time Body Posture Detection Computer vision algorithms are utilized to recognize and track human body postures and movements in real-time. Built-in cameras are used to capture images or video of the subject, which are then processed using computer vision algorithms to identify and track specific body parts and movements.

3. Sign Language Detection (Deep Learning) The problem is defined as a time series multilabel classification task, and the model is trained based on the WLASL [8] dataset. This dataset contains 2000 common words in American Sign Language, and a subset of 100 most common words are selected for the evaluation. The MediaPipe [5] Library is used to extract coordinates of hands and poses from the video data. The data is then fitted into a gated recurrent neural network with an encoder-decoder pattern for training the model.

4. AI-Driven Sentence Completion After detecting words and phrases using our sign language detection model, generative AI through an OpenAI chatbot is utilized to create complete sentences that accurately capture the original meaning conveyed by the sign user.



"Start/End" ***"Question Mark"*** ***"Delete"*** ***"Space"***

Figure 3. Special Gestures Sign Language Translation

As shown in Figure. 3, our approach incorporates four special gestures to allow sign users to select and modify words to be used in a sentence. The four gestures are **1) "Select", 2) "Space", 3) Delete", and 4) "Question Mark"**:

1. **"Select"**: This gesture is triggered when the user's hands are out of the screen. It is used to select instructions, words, phrases, and letters.
2. **"Space"**: This gesture indicates the end of letter spelling and helps distinguish between different words based on their spelling.
3. **"Delete"**: Users can use the "Delete" gesture followed by "Select" to remove the previous result from the list.
4. **"Question Mark"**: Placing this instruction at the sentence's end will result in an interrogative sentence.

Once the words, phrases, and letters are selected using the gestures, we use Azure OpenAI Chatbot to complete the sentence as shown in Figure. 4. To ensure that the output is desirable, we use techniques [1] such as Few-shot examples, Persona Patterns, and Contextual Augmentation to tailor make the prompt. These techniques help improve the accuracy and relevance of the output.

IV. DISCUSSION

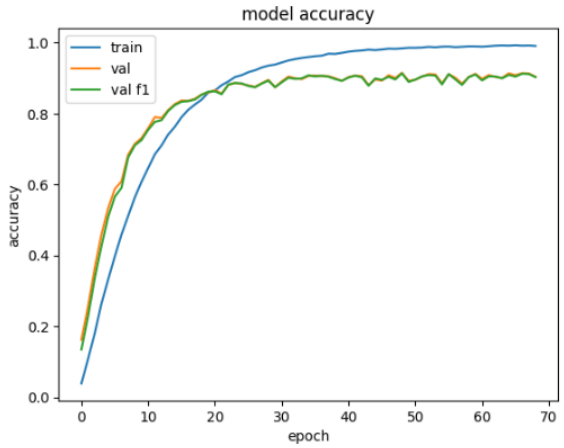


Figure 4: Performance Evaluation

*Val = Validation, Acc = Accuracy

	Val:Acc	Val:Loss	Test:Acc	Test:loss	Remarks:
Linear Regression (LR)	0.323	-	0.311	-	Val F1 = 0.373 Test F1 = 0.335
Random Forests (RF)	0.353	-	0.342	-	
LSTM	0.442	5.726	0.431	5.503	
GRU (1 layer)	0.462	5.435	0.458	5.467	
GRU (Encoder-Decoder Pattern)	0.552	4.723	0.566	4.789	
GRU (Data Augmentation)	0.655	3.231	0.672	3.112	
GRU (Standardize by Pose)	0.862	0.875	0.879	0.813	
GRU (Standardize by Hand)	0.864	0.861	0.849	0.898	
GRU (Both Standardization)	0.915	0.3287	0.918	0.3236	Val F1: 0.9137 Test F1: 0.9142

Table 1: Performance Evaluation

Model Performance Evaluation

We chose Recurrent Neural Networks (RNNs) [2] as the deep learning algorithm for HandsTalk because they are particularly effective at handling sequential data, including time series, natural language, and speech. To optimize the performance of our deep learning model, we also incorporated various data augmentation and different standardization approaches. As shown in Table 1 and Figure 4, our proposed model achieved an accuracy and F1 score of 91%, which is very promising and efficient.

Evaluation and Feedback with Sign Language Users:

We also collaborated with two sign language users to gather feedback on our application. The feedback we received was overwhelmingly positive, with users praising the application for its user-friendly experience and potential to facilitate communication for sign language users. Here are a few examples of the positive feedback we received:

Friendly All-in-One User Interface: *"The application is very user-friendly and easy to use."*
Benefiting a Wider Audience with Built-In Cameras: *"It's amazing that your application doesn't require specialized devices like leap motion or haptic gloves. This makes it more accessible to more people and allows for greater ease of use."*
Sign Language Communication App for the Deaf: *"This app has the potential to make communication easier for sign language users."*
Learning Platform for Sign Language Learners: *"It is a great tool for those who want to learn sign language or improve their skills.*
Immediate and Accurate Translation: *"I was impressed with how quickly and accurately the application detected changes in my pose or distance from the camera. It was able to recognize and translate the correct sign language most of the time, which made communication much easier for me."*
Real-Time Adaptive Translation: *"Great job! The application is highly accurate and able to quickly detect changes in pose or distance from the camera. It can successfully recognize and translate the correct sign language most of the time."*

V. Conclusion

In conclusion, the shortage of sign language translators in Hong Kong has made it challenging for deaf individuals to communicate effectively in various settings. However, our innovative mobile app, HandsTalk, which employs generative AI and advanced computer vision technology, has revolutionized the way in which sign language translation is performed. By using only built-in cameras, HandsTalk eliminates the need for additional hardware or devices, making it more accessible to a wider audience. We believe HandsTalk HandsTalk has the potential to positively impact the lives of many deaf individuals in Hong Kong, improving their ability to communicate and participate in various aspects of daily life.

VI. Acknowledgement

We would like to express our deepest gratitude to our esteemed project advisor, **Prof. Kenneth Wai-Ting LEUNG**, for his invaluable guidance and advice throughout the project.

In addition, we would like to extend our heartfelt thanks to sign language users **Wan Yongrui and Liao Yong** for their invaluable feedback that significantly contributed to enhancing our app.

VII. References

[1] Belagatti, P., “A Complete Guide to Prompt Engineering | Build AI Applications with SingleStore”, 2024. https://www.singlestore.com/blog/a-complete-guide-to-prompt-engineering/

[2] Smith, L. N., “Recurrent Neural Networks (RNNs): A gentle Introduction and Overview”, 2019.

[3] South China Morning Post, “Chinese University hopes new course will make life better for deaf”, 2018. https://www.scmp.com/news/hong-kong/health-environment/article/2168971/chinese-university-hopes-new-sign-language-course

[4] Google Firebase, “Firebase”, 2023. https://firebase.google.com/

[5] Google Developers, “MediaPipe”. https://developers.google.com/mediapipe/

[6] World Health Organization, “Deafness and Hearing Loss”, 2024. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss/

[7] WebRTC, “WebRTC Homepage” Webrtc.org, 2017. https://webrtc.org/

[8] WLASL, “WLASL Homepage”. https://dxli94.github.io/WLASL/