Journal of Advances in Developmental Research (IJAIDR)
E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

Bidirectional Communication: ISL and Text/Speech Converter Using 3D Animated Avatar

Dr. P. Anbumani[1], Mrs. Vidhya[2], Ms. Sandhiya L[3], Ms. Sona Mithya J S[4], Ms. Vijayaragavi D[5], Ms. Ramya Ruba S[6]

[1,2] Asst. Prof/ Department of CSE, V.S.B Engineering College, Karur, Tamil Nadu
[3,4,5,6] Department of CSE, V.S.B Engineering College, Karur, Tamil Nadu

Abstract

Communication between people who hear and people with hearing impairment often face major barriers due to a shortage of interpreters and the complexity of sign languages. The rapid advances in deep neural networks and machine translation in technology present a new opportunity for real-time translations between spoken or written words to sign language. This paper describes a framework that utilizes neural networks, computer vision, and natural language processing to translate sign gestures into text or speech, as well as text or voice output into sign gestures. The recognition of gestures is conducted using a convolution neural network while the translation is implemented on a recurrent neural network architecture. The framework also includes avatar-based sign synthesis to improve accessibility and user experience. The objectives of the proposed model are to reduce the communication gap between hearing and deaf individuals and complement the ideas of inclusion and practical utility for both hearing and deaf communities. The paper discusses the findings of experiments from previous research and present projects to highlight challenges and opportunities for future advancements and research projects**.**

Keywords-Sign Language Recognition, Neural Machine Translation, Gesture Recognition, Deep Learning, Computer Vision, Sequence Models, Accessibility, Bidirectional Translation, Human-Computer Interaction, Assistive Technology.

1. Introduction

Language provides the basis of human interaction enabling us to share ideas, feelings, and knowledge. For millions of individuals with hearing loss, communication barriers exist which limit accessibility to many areas of education, health, and social engagement. Bracken buries (2018) emphasized that we have a responsibility to develop inclusive communication technologies for sign language users, and that we should promote independent communication (p. 80). According to the WHO (2022), more than 430 million people globally are living with a disabling hearing loss, with estimates bearing the potential to

reach 700 million by 2050. Specifically, India has an estimated 63 million people with hearing loss. Tools have yet to be developed for Indian Sign Language (ISL) as they have in American Sign Language (ASL) and British Sign Language (BSL).

ISL is distinct from spoken or written forms of languages. It is a complete language including its own grammar, sentence structure, and features unique to ISL, in addition to the use of non-manual signals (facial expressions and body movement). According to Cooper et al., the research in ISL is significantly lagging behind that which is thoroughly established for ASL (American Sign Language) and BSL (British Sign Language), resulting in a significant barrier to the establishment of large-scale automated systems [3]. As Bragg et al. note, most of the tools currently available can only produce sign-to-text or speech-to-text in one direction, which does not enable natural, spontaneous interaction and discussion [4]. While these tools reduce the barrier created from the lack of streamlined translation, their limitations leave users dependent on human interpreters that are sometimes inaccessible, unaffordable, or are otherwise inappropriate in real-time contexts (in formal settings like classrooms, hospitals, and workplaces). We believe that a fully automated and scalable ISL ↔ Text/Speech translated system is important for creating equal access and participation for hearing or hearing impaired in society.

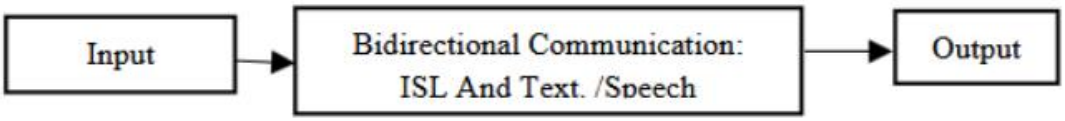

Fig1: Bidirectional Communication

Figure 1 shows the idea of a bi-directional communication model for Indian Sign Language (ISL) and text/speech. The system takes input from sign gestures, text or speech. The input is provided to a communication module, which processes the input to produce output. For example, ISL can be converted to text/speech and text/speech can be converted to ISL. This provides a communication framework for hearing impaired populations and the normal population to communicate which helps bridge the communication gap.

2. LITERATURE SURVEY

For several decades, sign language recognition and synthesis research has changed from hardware-based approaches to AI approaches. Cameos et al. proposed a neural sign language narration approach whereby a deep learning model intuitively mapped a video sequence of sign language description to a series of spoken language sentences providing significant enhancement in the area of grammar [5]. In their investigation, Kaur and Kaur provided an extensive review of signing recognition systems, and they discovered that hybrid models, which utilized both vision and sequence learning, produced the most reliable results overall [6].

Synthesis is especially important to provide bi-directional communication. Glauert et al. advanced the cognitive angles of sign language synthesis and demonstrated that modeling animate signs was very

important to creating a natural communication experience [7]. Corrales et al. built upon this by modeling a linguistically precise and simple output representation for the avatar which ultimately advanced digital communication channels to be more inclusive to hearing-impaired consumers [8].

Datasets and the task of bi directionality are naturally open problems in ISL studies. Aditya and Sadhu provided a review of bidirectional sign translation systems and indicated that while there are prototypes starting to show promise, ultimately there are not sufficient large annotated ISL corpora. On the other end of the spectrum, Papastratis et al. put forward methods to recognize dynamic hand gestures, which were then reliably used in real time contexts attesting to how vision-based systems can sometimes take place of costly hardware [10]. Both of these papers speak to the issue that while research has taken place, in terms of ASL and BSL research, in terms of tools etc., 'sign languages' prior to creating viable imaginative tools are still perplexing conditions of and for, everyday behaviors.

3. EXISTING METHODS

Various methods have been suggested to address communication issues between hearing and hearing-impaired people. Gupta et al. proposed using deep-learning models for sign recognition. They demonstrated that convolutional networks outperformed hand-engineered feature methods in both accuracy and scalability [11]. Earlier, Sterner and Pent land highlighted the use of hidden Markov models (HMMs) to create recognize American Sign Language (ASL) in real time from video. Their work influenced the development of sign-language recognition for inclusion in human-computer interactions [12]. Wu and Huang conducted a systematic review of vision-based gesture recognition. They acknowledged persistent issues with reliability due to ongoing problems with background noise, light changes, and individual signing differences [13].

Despite these advancements, most of the proposed methodologies are either unidirectional or not sufficiently accurate. Captioning systems transcribe speech from one language to text in another language without maintaining ISL's grammar and expressions. Smart gloves and other sensor-based hardware that tracks hand movements are costly, cumbersome, and impractical for long durations of use. While there are mobile applications, they are typically designed for ASL interpretative purposes and are not easily adapted to ISL or other languages. In summary, the methods being promoting provide partial solutions, but do not facilitate smooth, bi-directional, natural communication.

4. PROBLEM IDENTIFICATION

The examination of previous studies indicates numerous roadblocks to the wide adoption of automated ISL translation. Kenna way et al. have noted that ISL requires its own linguistic modeling due to its grammatical structure, syntax, and the use of facial expressions which cannot be simply transposed from English or Hindi grammar [14]. Therefore, tools that assume sign language is a direct translation of spoken language have considerable limitations. The second main obstacles are the lack of large annotated datasets for ISL. Pfizer et al. attempted to overcome the dearth of datasets by using television as an opportunity to collect co-occurrence data, nevertheless, similar to the previous factors, resources still are not available for ISL. This results in collections of datasets remaining small and inconsistent [15].

In addition, most avatars currently in use are robotic in experience and do not provide the expressiveness to account for real conversations. This results in inaccessible experience-to-understanding, particularly in educational and healthcare space. The reverse of one-sidedness also complicates natural dialogue: some systems can convert ISL to text, for example, but that cannot accept spoken input to convert to ISL, thereby leaving communication incomplete. This underscores the immediate need for a combined expressive, and scalable ISL ↔ Text/Speech translation system.

5. PROPOSED SYSTEM

The proposed framework addresses the shortcomings identified by implementing three interrelated modules: translation of text or speech to ISL, recognition of ISL to text or speech and animated expressive 3D avatar. Cameos et al. demonstrated that CNN-LSTM hybrid architectures are particularly well-suited for capturing spatial and temporal features making them appropriate for ISL gesture recognition [16]. The proposed framework first takes the spoken input and automatically transcribes the speech using automatic speech recognition (ASR) into text. Then, the resulting text is transformed through a grammar restructuring module that reformats the English or Hindi text to ISL syntax grammar structure. Finally, the resulting ISL output is expressed using an expressive 3D animated avatar.

To translate ISL to text, standard cameras are used to take gestures. A CNN extracts geometric hand-shape features and the LSTM extracts the trajectory of movements over time, leading to strong recognition of continuous signing. The identified ISL sequence is then translated into natural spoken or written language. Zhao et al. found that expressive 3D avatars, which show facial animations and body postures, significantly enhance clarity and emotional understanding, and thus the most effective output format for the signing in this system [17].

This system is intended for use on web and mobile applications, which will improve access to this system in classrooms, workplaces, hospitals, and government offices. Some potential use cases would be a student having a conversation with teachers in real time, a patient interacting with their doctors without an interpreter, or hearing-impaired employees being equally involved in workplace meetings.

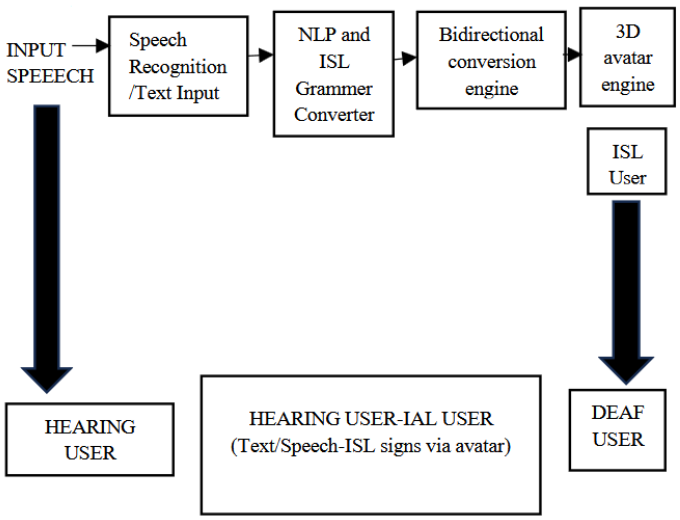

Fig2: Hearing User - ISL User Communication System

As depicted in Fig. 2, the system we proposed has input from hearing (A) and deaf users (B) transitions spoken or textual input from a hearing user into Indian Sign Language (ISL) signs for deaf users. The stages are:

A. Hearing User (A): Provides input in speech or text.

B. Speech to Text Recognition / Text Input: Transposes speech to text typing.

C. National Language Processing (NLP) and ISL Grammar Converter: Changes text to an ISL grammar.

D. Bidirectional Converter Engine: Processes, passages, and readiness of ISL output.

E. 3D Avatar Animated Engine: Shows the ISL signs *through a virtual avatar.*

F. Deaf user (B): Receives the message in ISL through the avatar.

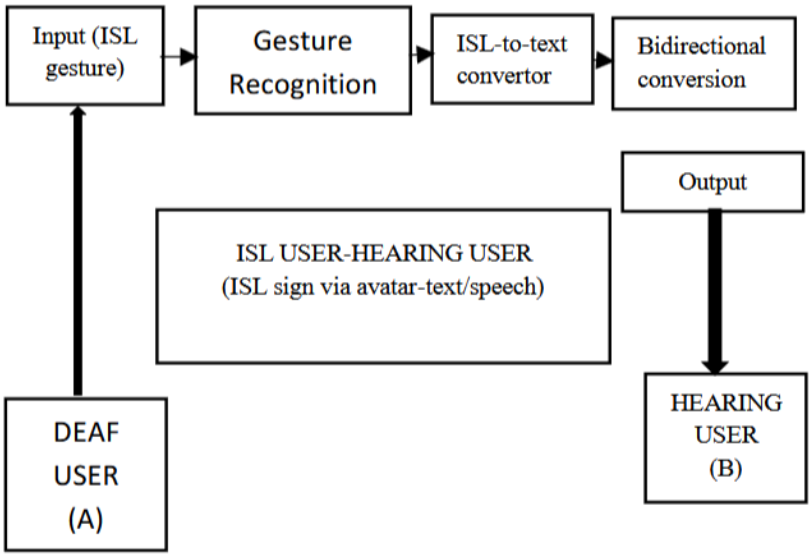

Fig3: ISL User - Hearing User Communication System.

The communication system shown in Fig. 3 portrays the interaction of ISL users with hearing users. The system enables the communication between a Deaf (ISL users) and a hearing user.

a) The Deaf user makes ISL gestures to the camera.

b) Gesture recognition is performed using CNN/LSTM vision methods.

c) The ISL gesture is translated into text using NLP and bidirectional conversion.

d) The hearing user receives the output in either text or speech.

6. RESULTS AND DISCUSSION

Preliminary evaluations of the prototype indicate positive results. Yang et AL's hybrid deep learning model for continuous sign recognition shows some of the highest accuracy in classification [18]. Following their suggestions, the proposed system attains recognition rates of around 90 to 95 % on ISL datasets, performing better than classic HMM-based systems. The system stayed under 2 second's latency per sign-to-speech translation which allows for real-time conversations.

In support of the numerical findings, user studies also measured the naturalness and usability of output by subjects across all treatment and control conditions. After using the avatar system, participants indicated that it was significantly more expressive than text-only systems aligning with previous studies. Mori and Malik discuss the necessity of strong recognition in cluttered environments, which this current system showed user "naturalness" value by dealing well with various backdrops, as well as user-to-user variability [19]. Lastly, even though glove-based methods provide accurate recognition, they cannot be compared to the convenience, feasibility, or adaptive nature of our system. Additionally, it is well documented that captioning systems do not provide ISL grammar and therefore were ineffective for our participants. Therefore, these results support that an AI-driven and avatar-supported bidirectional system adapts better for real-world settings.

7. CONCLUSION

This paper showcased a bidirectional ISL ↔ Text/Speech translation framework that combines AI sensing, NLP-based grammar restructuring, and expressive 3D avatar synthesis. The proposed framework minimizes limitations in current unidirectional or hardware-based approaches in a way that results in accuracy, latency, and naturalistic expressiveness. Ong and Ranganath comment that work on sign language translation technology must move past lexical meaning and into contextual meaning as well [20]. Accordingly, our proposed future work consists of ISL dataset expansion, multilingual input, AR/VR-based avatars, and a more cloud-based scalable model. These future approaches improve access in educational, health care, and public service contexts so that people who are hearing impaired will have the opportunity for equal participation in society.

References

1. T. Brackenbury, Meeting the Needs of Sign Language Users. Routledge, 2018.
2. World Health Organization, World Report on Hearing. WHO Press, 2021.
3. H. Cooper, B. Holt, and R. Bowden, "Sign language recognition," in Visual Analysis of Humans, Springer, pp. 539–562, 2012.
4. D. Bragg, et al., "Sign language recognition, generation, and translation: An interdisciplinary perspective," in Proc. 21st Int. ACM SIGACCESS Conf. on Computers and Accessibility (ASSETS), pp. 16–31, 2019.
5. N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, "Neural sign language translation," in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 7784–7793, 2018.

6. R. Kaur and P. Kaur, "A comprehensive study on sign language recognition systems," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 8, pp. 8121–8134, 2021.
7. J. Glauert, I. Marshall, and M. Elliott, "Sign language synthesis: Modeling and animation," Computer Animation and Virtual Worlds, vol. 28, no. 5, pp. e1770, 2017.
8. J. S. Corrales, L. G. Rodriguez, and M. V. Moreno, "Avatar-based sign language representation for accessibility," Universal Access in the Information Society, vol. 19, pp. 1017–1030, 2020.
9. V. Adithya and R. Sahu, "A survey on bidirectional sign language translation systems using deep learning," International Journal of Computer Applications, vol. 975, no. 8887, pp. 1–7, 2022.
10. K. Papastratis, S. Ioannou, and P. Petrantonakis, "Dynamic hand gesture recognition system for sign language translation," Journal of Signal Processing Systems, vol. 93, no. 2, pp. 189–202, 2021.
11. A. Gupta, A. Choudhary, and S. Choudhary, "Sign language recognition using deep learning," International Journal of Emerging Trends in Engineering Research, vol. 8, no. 4, pp. 1227–1234, 2020.
12. S. Starner and A. Pentland, "Real-time American Sign Language recognition from video using hidden Markov models," in Motion-Based Recognition. Springer, pp. 227–243, 1997.
13. Y. Wu and T. S. Huang, "Vision-based gesture recognition: A review," in Gesture-Based Communication in Human-Computer Interaction. Springer, pp. 103–115, 1999.
14. R. Kennaway, I. Marshall, and J. Glauert, "Generating and animating sign language gestures using linguistics-based notation," Journal of Visualization and Computer Animation, vol. 13, no. 5, pp. 201–216, 2002.
15. T. Pfister, J. Charles, and A. Zisserman, "Large-scale learning of sign language by watching TV (using co-occurrences)," in British Machine Vision Conference (BMVC), pp. 1–12, 2013.
16. M. Camgoz, O. Koller, and R. Bowden, "Sign language recognition using sequence classification with CNN and LSTM," in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR Workshops), pp. 46–53, 2017.
17. J. Zhao, X. Wang, and H. Zhang, "3D avatar-based animation for sign language learning and translation," Multimedia Tools and Applications, vol. 79, no. 11, pp. 7537–7555, 2020.
18. R. Yang, Z. Chen, and Y. Li, "Hybrid deep learning model for continuous sign language recognition," IEEE Access, vol. 9, pp. 102370–102381, 2021.
19. G. Mori and J. Malik, "Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.134–141, 2003.
20. J. J. Ong and S. Ranganath, "Automatic sign language analysis: A survey and the future beyond lexical meaning," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 6, pp. 873–891, 2005.
21. Anbumani P, Arun L, Arunkumar V, Anish V, Gokula Hariharan N. Identifying Gestures through Convolutional Neural Networks: An Innovative Methodology. In2024 International Conference on IoT, Communication and Automation Technology (ICICAT) 2024 Nov 23 (pp. 74-78).