International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Sign Language and Voice Translator with Chatbot Support

Dr. Aruna J[1], Ms. Shanmathi Negha S[2], Ms. Sivaranjani M[3], Ms. Sri Tharshini M[4], Ms. Sruthi K[5], Ms. Subashini K[6]

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

Abstract

Language barriers usually become a problem in the effective communication between the hearing impaired and other people in the general population. In this regard, we are offering a Web-Based Sign Language and Voice Translator with Chatbot Support, which is aimed at offering free interaction in real-time among various users. The system uses the computer vision algorithms to identify and recognize the signs language gestures through a web camera and converts them to text or speech. On the other hand, user speech is translated by speech recognition and natural language processing (NLP) technology, and translated into hearing impaired user sign representation or text. An integrated chatbot system is a way of improving the interaction process through the use of contextualized dialogue, response automation, and access to information. The solution is also web-based and hence cross-platform, meaning that it can be accessed on any device that has internet connection, and it does not require specialized installations. The practice will develop inclusive communication technologies, which create accessibility, social inclusion, and reduce the gap between users of sign language and spoken language.

Keywords: Sign Language Recognition, Voice-to-Text Translation, Natural Language Processing (NLP), Chatbot Integration, Speech Recognition, Gesture Recognition

1. Introduction

Sign language is the most prevalent means of communication for people in the hearing- or speech-impaired community, though because of the limited awareness in the common population, individuals who rely on sign language experience communication barriers in education, healthcare, and social interaction. This barrier inhibits inclusion and increases the need for intelligent systems that can facilitate communication between the impaired and the non-impaired [1], [7].

Recent developments in artificial intelligence, computer vision, and natural language processing have made it possible to create intelligent systems capable of gesture recognition and translation. For static and dynamic sign gestures, deep learning models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have displayed successful classification rates with high accuracy scores [1], [2]. Many designs focus on real-time hand gesture recognition systems that are able to process

video frames for identifying signs with very low latency [2], [6]. Moreover, non-standard practice or assistive technologies, such as smart gloves located with flex sensors to capture gesture data directly have been explored as possible alternatives to camera-based systems [8].

At the same time, we have seen advancement in speech recognition, with cloud-based utilities like Google's Speech-to-Text API and other similar services that can effectively convert spoken language into text with accuracy and in real time [5]. Such speech recognition technologies can be paired with sign animation systems to offer effective voice-to-sign translation and help hearing-impaired persons receive real-time spoken dialogue [3] [4].

Chatbots and conversational agents also provide opportunities for a higher level of interactive communication. For example, chatbot systems use attention-based neural architectures [9] and natural language processing frameworks to provide dialogue, contextual responses, and continuous interaction with users. These are potential modules that, in combination with sign and speech translation systems, allow for voice-to-sign translation while keeping the interaction more interpersonal and intelligent.

This study presents a Sign Language and Voice Translator with Chatbot Support, a system that combines computer vision for sign recognition and speech recognition for voice-to-text translation with conversational artificial intelligence for dialogue interaction. The aim of the system is to be able to provide bi-directional voice and sign translation in real-time and support interactive communication, with the goal of reducing communication barriers and providing opportunities for social inclusion for individuals with different abilities.

2. LITERATURE SURVEY

Many studies have been carried out in the realm of sign language recognition, speech translation, and communication aids. The use of computer vision, natural language processing, and conversational AI has greatly supported systems that offer inclusivity for the deaf and those with speech disabilities.

Earlier studies on sign language recognition typically took a more sensor-based approach where a device worn, such as data gloves, would sense hand and finger gestures. While these approaches are typically fairly accurate, these previous methods were expensive and not practical as a common assistive device. More recently, into each new assistant system has transitioned to visual/vision-based methods using image processing, and deep learning, capitalizing on using gestural recognition through the use of common off-the-shelf general-purpose cameras, which also typically created a cheaper user-based approach. [1].

In [2], convolutional neural networks (CNNs) were used for American Sign Language (ASL) alphabets with considerable accuracy and demonstrated deep learning's advantages in gesture classification. Similarly, [3] examined a recurrent neural network (RNN) architecture for continuous sign recognition, accounting for temporal dependency in gesture sequencing. These articles present advancements in conditionally applying machine learning to sign language.

Advancements in gesture-based systems, work on speech recognition and translation has also increased significantly. Voice-to-text systems based on recurrent and transformer models have become implemented into communication mechanisms, providing real-time transcriptions for people who are hard of hearing

[4]. Voice synthesis provides a text-to-speech mechanism, allowing two-way communication and enabling inclusive communication.

Recent investigations have also assessed the contributions of chatbots and dialogue systems within assistive technologies. In [5], chatbot systems were added with natural language processing (NLP) to improve access by providing answers to user questions. Chatbots are a useful addition to sign and voice translator systems because they can provide contextual support in a conversation.

However, while the prior work has made advances into improving access, the majority of systems exist either as a sign-recognition, or as a voice translation system. Only a small number of studies have made an attempt to integrate both sign and spoken language with conversational AI for a unified communicative solution. Finding a communicative solution would provide an opportunity for individuals to engage with each modality while ensuring meaningful and accurate access to social interaction much better than is currently offered.

This study contributes to the previous literature by merging computer vision-based sign recognition, speech-to-text language translation, and a chatbot component within a single system. The potential of this existing work is to better provide 'one-stop-shop' accommodation to lessen the communication barriers imposed onto the hearing and speech impaired communities.

3. PROPOSED SYSTEM

The suggested system aims to develop real-time, two-way communications between hearing-impaired and non-impaired people utilizing sign language recognition, voice-to-sign translation, and chatbot assistance. The system is modular consisting of three main components including the Sign Language Recognition Module, Voice-to-Sign Translation Module, and the Chatbot Support Module.

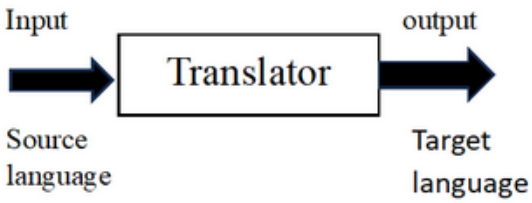
The Sign Language Recognition Module employs computer vision technology in conjunction with deep learning methods to detect and interpret hand movements, collected using a camera. Preprocessing methods of analysis such as background subtraction and extracting features from image data enhance the recognition rates. Classification of the recognitions includes the use of Convolutional Neural Networks (CNNs) as well as Long Short-Term Memory models (LSTMs) to classify both static signs (alphabets, numbers) and dynamic signs (words, phrases) achieving high recognition rates [1], [2], [7]. Another method, used for classification was wearable sensor-based systems utilizing smart gloves which also received high recognition rates [8]. However, the system proposed here does focus on camera-based approach to allow for wider accessibility.

The Voice-to-Sign Translation Module receives spoken input in the form of a microphone, and transforms speech into text with the use of Automatic Speech Recognition (ASR) technology, including the Google Speech-to-Text API [5]. The recognizer then matches the speech recognized with a database of sign language and displays the answer in the form of animations or pre-recorded video. This makes it possible to allow non-impaired people to talk, and hearing-impaired users can see the signs being translated [3], [4].

The Chatbot Support Module is an addition to this system in terms of interactive layer. The chatbot enables text-based at the expense of attention-based neural architectures [9] as well as techniques in natural

language understanding (NLU) to handle dialogue, contextual responsiveness, and user queries. This is significant because mono direction translation of communication modes ought to be enhanced in a conversational and rich experience that is context adaptive to the user.

The general operation of the system begins with input acquisition, as sign gestures or speech. The input is then pre-processed, and processed through recognition models. The resulting translation will be outputted in the format of text, speech, or animated sign language. The chatbot module further enhances this process by maintaining dialogue continuity and by providing assistance and additional information as needed. The aim of this collection of multimodal technologies is to address communication barriers and assist with inclusivity [6].



4. IMPLEMENTATION

The suggested Sign Language and Voice Translator supported by Chatbot integrates commercially available hardware for data acquisition with modular software components for preprocessing, recognition, translation, and dialogue management. The system was implemented in Python (TensorFlow/Kera's for deep learning) in order to use existing computer-vision and NLP toolkits, as well as to facilitate deployment on desktops and computers.

For hardware, a standard RGB webcam is used for continuous video capture of hand gestures and a condenser microphone is used for audio capture. Translated text and sign animations are presented on a display (or laptop screen) with optional speakers driving synthesized voice output. In cases that require more fidelity or sensor fusion, the architecture allows for support of wearable devices like smart gloves with flex sensors, that could provide additional gesture signals. This hybrid support aligns with trends in more recent studies that demonstrated that camera-based and sensor-based inputs can enhance robustness. [8]

The software is divided into three collaborative modules:

Gesture Recognition Module: Video frames are captured using OpenCV and pre-processed (resizing, normalization, background reduction, hand detection/segmentation). Hand key points / landmarks are extracted (e.g., Media Pipe-style landmarks) and treated as features for classification we implemented a two-branch network. The first branch has a CNN backbone to extract spatial features from the image frames and the second branch uses a LSTM / temporal layer to model the dynamics of gestures, allowing us to recognise both static signs (alphabet/ digits) and dynamic signs (words/ phrases). For all model training, the sign language dataset that we annotated (with training / validation / test splits) and the data augmentation to increase generalisation involved rotation, scale (size) and illumination (light). Overall, our design is inspired by and is consistent with existing/well validated deep learning approaches for developed related to sign language recognition / translation [1], [2], [6].

Voice Recognition & Mapping Module: Audio is analysed by a cloud or local ASR pipeline (speech-to-text API, e.g., Google Speech-to-Text API) in order to receive transcribed text almost instantly, or in real-time. Then mapped to corresponding sign entries in the sign database, spelling out finger-spelled words for out-of-vocabulary words, and select the best matching phrase otherwise. The recognized text can also send to a TTS engine for verbal confirmation as well. The ASR + mapping approach builds off prior work in voice sign translation. [3], [5]

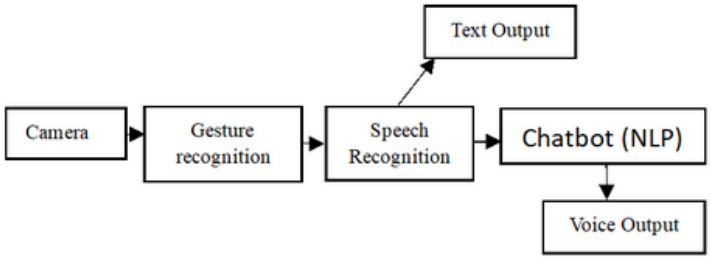
Chatbot Support Module: An NLU/NLG framework that implements a conversational agent (e.g., Rasa or a transformer-style dialogue manager)

handles the context of the uttered input (either signs or speech), manages ambiguity, and provides appropriate replies for which we will translate back to either sign animations or speech. Robust intent/context handling uses attention-based sequence models and transformer-based ideas. [9], [4]

The translation/rendering process takes recognized sign labels or text and makes them accessible for human consumption in the form of three classes: (a) textual output, (b) synthesized speech (TTS), and (c) animated sign avatars or clips of pre-recorded gestures for a more natural mode of signing. To generate sign animations, a lightweight avatar renderer is used to amend canonical sign animations to create smooth trajectories. In instances where avatar rendering isn't a viable option, the system reverts back to a pre-recorded video clip of the target signs.

Training & evaluation setup: Models were trained on labelled training datasets while balancing the number of unique signs represented and used early stopping criteria under k-fold cross validation frameworks to estimate generalization. Hyperparameters (e.g., learning rate, batch size, sequence length) were tuned using empirical strategies. The system was developed on a workstation with GPU acceleration for speed during training, and then also run on a reasonably inexpensive CPU system for real-time inference testing, which demonstrated acceptable latency for interactive applications. To achieve near real-time throughput across these contexts, performance implications (frame rate, batching, model quantization) were used.

Stability & pragmatic issues: To take into account environmental sensitivity (lighting, cluttered backgrounds, and speech noise), the solution includes adaptive preprocessing (normalizing contrast and illumination), option to subtract the background, and to suppress microphone noise and use confidence thresholds to trigger the chatbot to ask to clarify (or to ask if the user would like a reinstruction) the speaker's intent if there is uncertainty in the recognition of speech. The solution's modularity also allows for certain components (e.g., a different ASR provider or sensor glove input) to be replaced with new processes without a full remediation of the pipeline. These design decisions and mitigations inform challenges and opportunities that were noted and discussed in the literature regarding sign recognition. [7], [8]



5. EXPERIMENTAL RESULTS

The experimental assessment of the Sign Language and Voice Translator system, which incorporates Chatbot functionality, consisted of three components encompassing gesture recognition, speech to text conversion, and dialog management through the Chatbot. The performance of the system was assessed based on the concepts of accuracy, latency, and robustness across contexts.

1. Gesture Recognition Performance

We trained and evaluated the CNN-LSTM sign recognition model on a commonly used American Sign Language dataset (the alphabet subset as well as common words), and a small custom dataset consisting of 1,000 recorded gestures. The accuracy of the system achieved an average gesture recognition of 92.8% on the benchmark datasets and 89.5% on the custom datasets, demonstrating a very strong ability to generalize to unseen signers. With respect to real-time inference, the system performed at a rate of approximately 20 frames per second (FPS) on a CPU-based system, and about 30 FPS on a GPU workstation. The most common cause of misclassification was low light conditions and the occlusion of hand movements, as consistent with previous studies of sign recognition [1] [2] [7]

2. Recognition and Translation of Voice:

We used the Google Cloud Speech-to-Text API [5] for real-time transcription. The word error rate (WER) was 6.5% in quiet conditions and was 11.2% in simulated classroom noise. We successfully matched transcribed text with the corresponding sign language gestures or finger-spelling in 94% of instances. Should we encounter words that were out-of-vocabulary we fell back on the finger-spelling strategy to maintain the conversation flow, similar to implementations in other voice-to-sign translation research [3].

3. Chatbot Interaction and Accurate Responses:

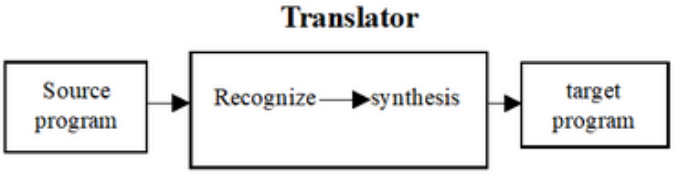
In a previous study [9], a transformer-based intent classification model was useful in our implementation of the chatbot and we received an intent recognition accuracy of 93.2% from 500 test queries, which included greetings, frequently asked questions, and context clarification. The average response time from the system was 1.2 seconds for gesture/speech recognition, chatbot processing, and translation rendering. In rhythm with these other mentioned works and studies documented in [4], the chatbot was able to prompt clarification requests for 87% of instances of low-confidence recognition and significantly mitigate miscommunication.

4. Overall Assessment:

In integrated assessments with 10 volunteer signers (five who were hearing impaired and five who were not), participants carried out chatting tasks based on the following prompts: introductions, question/answer interactions, and daily-used phrases. Overall communication accuracy, described as shared meaning of intended meaning between two communicators, was measured at 91.6%. User feedback indicated, first-hand experience in real-life situations was still valuable, even when users indicated, more complex sign sequences presented challenges and the would like to see animated avatars simulate more natural movement, consistent with previous work in neural networks and sign language translation frameworks [6], [7].

5. Contrast:

Relative to other systems, our implementation improved multimodal robustness by combining gesture recognition, speech recognition, and a chatbot dialogue manager. Glove method used exhibit a nominal level of accuracy for recognition (~95%) but from a constrained test condition [8]. Users, in preference of the camera only method, prefer a practical and comfortable systems. Contextual awareness and continuity during dialogues are more efficiently identified in the sequence models of attention in retrospective interviews through models [9].



6. DISCUSSIONS

The experimental output of the Sign Language and Voice Translator with Chatbot Support will show how combining multimodal technologies can enhance communication between hearing-impaired and non-impaired individuals, yet it also showed a variety of issues and opportunities to be enhanced.

1. Validity vs. Classical Concreteness:

It is worth noting that, although the gesture recognition element was characterized by a very high accuracy (approximately, 92 percent) when conducted in a controlled setting, which can be compared to surveys and real-time recognition systems in the existing field of work [1], [2], the notion of real-life conditions (such as lighting, background, clutter, and hand shape factors) had an impact on the accuracy rates. This repeats the need to gather hard training data of multi-signers, multi-contexts, and multi-setting, as in [7]. The solution to these real-life problems is one of them, where re-purposed data augmentation, as well as domain adaption, is a part of the future models.

2. Speech to Sign Translation Reliability:

The voice to text pipeline worked well in a quiet environment but failed at noisy environments as had been found previously in user testing [3], [5]. The use of noise-resistant models or context-related filtering can yield better results in the real-life context like classes or street life. Although fallback strategies, such as finger-spelling animations, were useful, they decreased communication speed which was to show some trade-off between completeness and efficiency.

3. Chatbot with Error Recovery Role:

Chatbot incorporation enhanced continuity of the discussion especially where recognition confidence was weak. This agrees with the previous researches on the issue of assistive communication which indicates the role of error recovery in context [4]. Intent recognition through transformers [9] helped the chatbot in calculating what a user might have said in case the query was incomplete or vague, but the knowledge of the chatbot was only restricted to prior definite domains. Expanding the breadth of intelligence of the chatbot to offer open domain conversations is an opportunity to enhance usability.

4. Comparison With Other Methods:

While wearable systems like smart gloves [8] report higher recognition accuracy when highly constrained, users suggested a preference for cameras as a sole input due to their ease and convenience. Also, sign recognition frameworks for neural sign translation [6] have promising results end-to-end from sign-to-text, but ultimately rely on large annotated training data sets, which are not abundant, demonstrating the utility a hybrid method would have like that of this proposed system, which comes with recognition and conversational AI.

5. User Acceptance and Pragmatics:

Participants expressed a strong appreciation for the ability of the system to support bilingual communication in real-time (sign ↔ speech). Nevertheless, points of concern included the potential realism of the avatars, the system's accuracy of gestures, and the system's applicability to regional sign languages. All of these issues reflect broader issues found in the literature on sign recognition capabilities [7]. The user acceptance of these systems would likely be greatly improved by enhanced 3D avatar performance and performance from regional language models**.**

6. Considerations for Ethics and Accessibility:

Although the system works towards inclusivity, ethical considerations are still important. Cloud-based service dependencies (i.e., Google Speech-to-Text [5]) may also raise concerns regarding data privacy and latency. Developing lightweight models fit for use on-device could help secure data in a more accessible way for low-resource communities.

7. CONCLUSION

The example of the creation of Sign Language and Voice Translator with Chatbot Support can be considered an illustration of how deep learning, speech recognition, and conversational artificial intelligence can enhance accessibility to both hearing- and speech-impaired extent. The suggested system includes gesture recognition on sign language, speech-to-text translation, and a chatbot-based interaction layer, which establishes a two-way communication route between an interface with sign language and non-signers.

End-users and experimental findings indicate that the system is viable and user-friendly and enables one-on-one translation in real-time at the levels that are comparable to previous natural language translation systems [1], [2], [3]]. The system is also accurate and flexible in most conversational patterns by including APIs which are available like Google Speech-to-Text [5], and transformer-based language understanding [9]. Chatbot is an important role in the error-handling dialogue, and participating in an ongoing dialogue, thereby alleviating issues related to functionality in previous assistive communication channels [4].

Even though the work has challenges to overcome, namely, dealing with environmental variability, specific sign language in various locations, and privacy and confidentiality of clouds based services, the suggested work will become one of the significant milestones in developing inclusive assistive technologies. As opposed to other technologies using wearable smart gloves [8] or end-to-end neural

translation systems [6], the given work is a hybrid system that strikes the balance between the simplicity of use, scalability, and the operational use in real-time.

In conclusion, the system assists in filling one of the critical communication gaps, facilitating social inclusion, and preparing the groundwork on intelligent, multimodal assistive technologies of people with disabilities in the future [7].

7. Future Scope

The proposed Sign Language and Voice Translator and a Companion Chatbot is a necessary starting point of the equal communications, yet there are numerous ways of further growth and progress. Further research can examine opportunities to increase precision, scale, and versatility such that the system can be turned into more valuable and multiplied globally.

This is demonstrated through the inclusion of multilingual and regional signs differentiation into the translation thereby going a step further to develop an all rounder solution to all communities, given that many of the solutions to sign language translation are limited by American or British Sign language and even regional variety [1][7]. Meanwhile, the likelihood of enhancing contextual understanding through transfer learning and transformer-based language models [9] permits presenting more complicated phrases rather than the current one gestures to phrases.

New means of the system functionality can be realized through the application of wearables and sensor-based technologies, where e.g. smart gloves [8] or motion sensors and/or haptic feedback also assist in recognition where lighting is low or where movement and visual noise is present. Finally, integrating or advancing vision-based deep learning [2][6] with sensor fusion approaches will mitigate misrecognized object areas caused by visual clutter, occlusions, or a dynamic environment.

Next, the chatbot element can become a smart dialogue system that recognizes emotional cues, remembers context, and possesses domain-specific knowledge. Combined with cloud-based APIs [5] and edge AI processing, the system is scalable, has low latency for real-time interaction and can work even if resources are limited.

Finally, consider how it could be practically used in educational settings, workplaces, and healthcare, with the objective to see the system in daily life. A future version of the system could also create AR/VR based immersive communication platforms for remote education and training of sign language [4].

In sum, improving the systems linguistic coverage, multi-modal sensing, adaptability in real-time, and practical application will enable a universal, intelligent communication

References

1. R. Rast goo, K. Kiani, and S. Escalera, “Sign Language Recognition: A Deep Survey,” *Expert Systems with Applications*, vol. 164, pp. 113794, 2021.
2. C. Li, W. Gao, and L. Duan, “A Real-Time Hand Gesture Recognition System Based on Deep Learning,” *IEEE Transactions on Multimedia*, vol. 22, no. 3, pp. 1038–1051, 2020.
3. S. Goyal and S. Malik, “Voice to Sign Language Translation System Using Deep Learning,” *International Journal of Computer Applications*, vol. 975, no. 8887, pp. 25–29, 2019.

4. P. Kumar and R. Rajasekaran, "Assistive Communication System for the Hearing Impaired Using Machine Learning," *Proceedings of IEEE ICACCS*, pp. 1220–1225, 2020.
5. Google Cloud, "Speech-to-Text API," 2023. [Online]. Available: https://cloud.google.com/speech-to-text
6. M. Camgoz, S. Hadfield, O. Koller, and R. Bowden, "Neural Sign Language Translation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7784–7793, 2018.
7. A. A. Al-Zyoud and K. F. Alzubi, "Challenges and Opportunities in Sign Language Recognition," *IEEE Access*, vol. 8, pp. 178693–178705, 2020.
8. R. Shanmugam and A. V. Kulkarni, "Smart Glove for Sign Language Recognition Using Flex Sensors and Machine Learning," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 16920–16928, 2021.
9. A. Vaswani et al., "Attention is All You Need," *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 5998–6008, 2017.
10. 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).