International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 10 October 2025
Page No: 3978-3984
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0581

Study and Overview of Multilingual Voice & Text to Sign Language Translator

Khushbu Mahale[1], Vedant Sawant[2], Shardul Mane[3], Prof. Neha Nandurkar[4]
[1,2,3]UG Scholar, Dept. of AI&DS, Rajiv Gandhi Institute of Tech., Mumbai, Maharashtra, India.
[4]Assistant Professor, Dept. of AI&DS, Rajiv Gandhi Institute of Tech., Mumbai, Maharashtra, India.
Emails: *Khushbumahale12@gmail.com[1], vedantsawant0503@gmail.com[2], shardulmane20@gmail.com[3], nehanandurkar196@gmail.com[4]*

Abstract

This paper introduces a new approach to improving communication for deaf and hard-of-hearing individuals by developing a multilingual voice and text–to–sign language translator. Unlike existing tools that often lack support for multiple languages, struggle with real-time accuracy, or only handle text or voice separately, our solution combines both voice and text inputs in one system. The framework uses advanced AI technologies, including speech recognition (ASR), natural language processing (NLP) for context-aware understanding, and deep learning to generate realistic sign language. It is designed to support different sign languages (such as ASL and ISL) through specialized datasets, while addressing common challenges like limited vocabularies, robotic or unnatural avatars, and weak handling of continuous signing. The results show that this integrated AI model can create a more natural and inclusive communication experience, with the potential to promote social inclusion, expand educational opportunities, and bring economic benefits.

Keywords: *Multilingual Communication, Natural Language Processing, Sign Language Translation, Speech Recognition.*

1. Introduction

Communication is the foundation of human interaction, yet for the Deaf and Hard of Hearing (DHH) community, communication with the hearing population remains a persistent challenge. According to the World Health Organization, over 430 million people worldwide live with disabling hearing loss, and many of them rely primarily on sign language for daily communication. However, unlike spoken languages, sign languages are not universal. Each country, and sometimes even regions within a country, has its own sign language with unique vocabulary, grammar, and cultural context. For instance, American Sign Language (ASL), British Sign Language (BSL), and Indian Sign Language (ISL) differ significantly from one another, making multilingual communication an even greater challenge. Despite advances in Artificial Intelligence (AI) and Human-Computer Interaction (HCI), the availability of robust, real-time sign language translation systems is still limited. Most existing approaches suffer from three key limitations. First, many systems are restricted to static gestures, such as alphabets or digits, which fail to represent the dynamic nature of continuous singing in real-life conversations. Second, there is a lack of multilingual support. Current tools are often tailored to a single language pair, such as English-to-ASL, without considering linguistic diversity or the needs of users in non-English-speaking regions. Third, systems often lack naturalness in translation—omitting facial expressions, non-manual markers, and smooth gesture transitions, which are essential for accurate interpretation. The research further examines four datasets Sign Language MNIST, ASL Alphabet Dataset, WLASL, and ISL-Corpus to evaluate their suitability for different stages of system development. Static image datasets provide efficiency in prototyping, while dynamic video datasets enable scalability for real-world conversational systems. Beyond improving accessibility, the potential applications of such a translator are vast. In education and e-learning, it can empower Deaf students to access lectures, digital courses, and assessments in real time. In public services and healthcare, it can reduce reliance on human interpreters and ensure equal access to critical information. In social and professional settings, it can foster inclusivity and independence for members of the Deaf community. This paper aims to provide a comprehensive study and overview of multilingual voice and text to sign language translation, highlighting current research gaps, proposed methodologies, datasets, and future directions for building scalable, inclusive systems.

2. Literature Survey

[1] This paper presents a gesture-based voice and language translator designed specifically for individuals with speech impairments, addressing the lack of effective communication systems for this group. The authors implemented a system using MPU6050 sensors coupled with a Raspberry Pi to capture hand gestures corresponding to the English alphabet need for effective communication systems within abets. A trajectory recognition algorithm was developed, which performs raw data preprocessing, feature extraction, and classification through K-Nearest Neighbors (KNN). The recognized gestures are converted into voice using the Voice RSS API and further translated into multiple Microsoft-supported languages through Microsoft Translator. The study highlights three modes of operation—training, testing, and translation—and reports successful recognition of A–Z alphabets with multi-language voice output. The system overcomes limitations of earlier glove- and camera-based approaches by simplifying gesture representation, reducing noise through feature selection, and supporting multilingual output, thereby offering an artificial voice to aid communication for the speech impaired. However, it remains constrained to alphabet-level recognition and requires future expansion to words and symbols. [2] This paper introduces a more advanced real-time sign language translation system leveraging IoT-enabled wearable gloves embedded with flex and IMU sensors. Unlike alphabet-only recognition, this system focuses on dynamic sign language gestures and integrates deep learning through an optimized LSTM model, achieving a classification accuracy of 97.2% with a translation latency of just 180ms. The system processes sensor data via edge computing and cloud-based AI to deliver instant multilingual translation across more than ten languages, addressing the limitations of earlier vision- and glove-based systems that suffered from latency, high power consumption, or lack of multilingual capabilities. User evaluation reported high satisfaction (4.8/5), and the device operated for 12 hours per charge, making it both practical and portable. By integrating IoT, deep learning, and multilingual cloud translation, the system demonstrates a significant improvement over earlier models in terms of speed, accuracy, and usability. [3] This paper proposes a model that converts speech to Indian Sign Language (ISL) for six regional languages: Telugu, Hindi, Malayalam, Marathi, Kannada, and Tamil. The model's architecture involves three main phases: speech recognition, text translation, and sign language generation. For speech recognition, the system uses Wavelet-based Mel-Frequency Cepstral Coefficients (MFCC) with a Gaussian Mixture Model (GMM). Research shows that Gaussian models perform well in recognition tasks. The text translation phase uses an encoder-decoder-based Long Short-Term Memory (LSTM) model. An attention layer is added between the encoder and decoder to handle longer sequences and focus on relevant parts of the sequence. Finally, the translated text is converted to ISL using a direct translation method, which maps each character to a corresponding sign. The system achieved a validation accuracy of over 80% for each of the six languages. [4] Another paper explores sign language translation across multiple languages by leveraging deep learning algorithms, including

Convolutional Neural Networks (CNNs). The project creates a platform that translates both Indian Sign Language (ISL) and American Sign Language (ASL) into multiple Indian regional languages. This serves a dual purpose: enabling communication between people familiar with different sign languages and providing written text for those who don't know sign language but are proficient in regional languages. The methodology involves training ISL and ASL models on a dataset created from webcam images. The data is preprocessed, augmented, and used to train a CNN architecture. A key finding from the research is that the CNN model achieved impressive accuracy rates of 99.90% for grayscale images and 99.72% for color images for ISL recognition. The ASL recognition model also showed a significant 9% improvement in accuracy compared to previous models. [5] A separate paper addresses the complexities of sign language recognition and translation by focusing on the challenges of converting sign language to text/speech and vice versa. A core problem identified is the unique grammatical structures of sign languages, which complicates the development of robust computational models. The paper proposes using a "reversible CNN" for feature extraction from sign language gestures, a model that can not only extract features but also reconstruct the original data, which is useful for anomaly detection and data compression. This reversible CNN is a 12-layer model that processes voice and gesture inputs. The system's sign-to-text conversion achieved a promising overall accuracy of 70% and a precision of 46.6428%. The paper also discusses a text-to-sign conversion module that uses Natural Language Processing (NLP) techniques and a database of gestures to create animated signs. Looking ahead, the research aims to extend its capabilities to encompass a broader range of Indian regional languages and move beyond single-word or isolated sign translation to full sentence-level expressions. [6] This paper details the development of a real-time sign language translator that integrates speech recognition to bridge the communication gap between sign language users and non-signers. The system translates spoken or typed words into corresponding American Sign Language (ASL) and Indian Sign Language (ISL) gestures using a database of sign representations. The methodology relies on a lightweight and accessible technology stack, using standard web technologies like HTML, CSS, and JavaScript for the front end and the Web Speech API for real-time speech-to-text transcription. The core of the system is a sign language mapping database that links words or phrases to visual representations like GIFs or videos. While the system shows promise, challenges remain, including decreased accuracy in noisy environments, a limited vocabulary, and the inability to handle complex grammar or non-manual components like facial expressions and body posture. Future enhancements are planned to address these limitations by incorporating AI-based gesture recognition, crowdsourcing for a larger vocabulary, and using deep learning for speech enhancement. [7] This paper reviews existing sign language recognition and translation systems, highlighting their strengths and limitations. It notes that sensor-based systems, such as those using wearable gloves, offer high precision but are often expensive, bulky, and not user-friendly for daily use. In contrast, vision-based systems, which rely on computer vision and deep learning, are more convenient but are highly susceptible to environmental factors like lighting and background clutter. The authors also explore advancements in speech recognition, mentioning tools like the Web Speech API, which perform well but still face challenges with diverse accents and background noise. The literature review points out a significant gap in research regarding multimodal translation systems that seamlessly integrate both speech recognition and sign language translation in a comprehensive framework. This gap suggests a need for more holistic systems that can effectively bridge these modalities. Additionally, the review highlights the overlooked importance of non-manual components of sign language, such as facial expressions and body posture, which are crucial for conveying full semantic meaning but are often not accounted for by current technologies. The paper

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 10 October 2025
Page No: 3978-3984
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0581

concludes that the integration of AI and Natural Language Processing (NLP) could be a key factor in overcoming these limitations, offering more fluid and accurate translations by providing contextual understanding and language modeling. [8] This paper highlights a significant disparity in research, noting that a vast amount of work has been done on American Sign Language (ASL) compared to Indian Sign Language (ISL). This scarcity of research on ISL is primarily attributed to a lack of dictionaries and public exposure. Existing software in this field has predominantly focused on ASL or British Sign Language (BSL). The few systems that exist for ISL translation are described as scarce, ineffective, inconvenient, and not user-friendly. The document states that the fundamental architectures of these systems are typically based on direct translation, which may not meet expectations; machine transformation, which requires large, often unavailable parallel databases; or transfer-based architecture, which uses grammar rules to specify translations. In contrast to these approaches, the proposed work is presented as a more user-friendly, web-based, and optimized solution that utilizes video representation for translation, contrasting with other methods that use images or animations. The paper cites previous works, noting approaches such as using signal notation for language representation, synthetic animation for conversion, and storing videos for each word in a preset database.

2.1. Tables

Table 1, The comparative analysis of prior works highlights varying levels of performance across sign language recognition and translation systems.

Table 1 Performance Metrics

Paper	Evaluation Metrics	Reported Performance
[5] Sign Language Recognition and Translation Systems for Enhanced Communication	Accuracy, Precision, Recall, F-measure	70% overall accuracy
[7] Sign Language to Text and Speech Translation in Real Time Using CNN	Accuracy	95% accuracy
[3] Speech to Sign Language Translation for Indian Languages	Validation Accuracy	>80% for all languages; 81.4%–91% depending on test cases
[4] Sign Language Translation Across Multiple Languages	Accuracy (comparative)	Prior CNN: 99.90% (grayscale), 99.72% (colored); Proposed model: +9% ASL accuracy vs baseline
[6] Sign Language Translator with Speech Recognition Integration	Not specified	No specific values; notes accuracy impacted by background noise & varied speech

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 10 October 2025
Page No: 3978-3984
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0581

Accuracy remains the most common evaluation metric, with CNN-based approaches achieving high results, such as 95% for finger spelling and up to 99.9% in controlled ISL recognition tasks. Multilingual systems, like the Speech-to-Sign model for Indian languages, reported accuracies above 80%, demonstrating potential but also revealing dataset and scalability challenges. On the other hand, some works, such as the Sign Language Translator with Speech Recognition Integration, emphasize practical deployment issues like background noise, which continue to limit real-world usability despite promising results in experimental conditions.

2.2. Figures

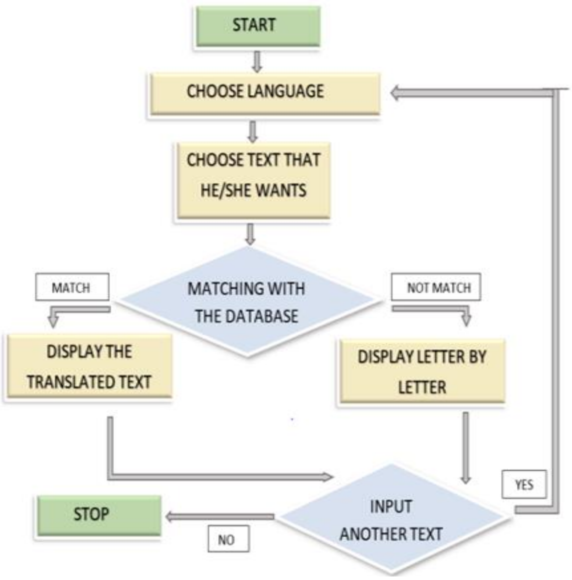

Figure 1 **Text-to-Sign flowchart**

Figure 1 The Text-to-Sign flowchart [5] illustrates the process for a text-to-sign language conversion system. The process begins with the user selecting a language. The user then inputs the text they want to be translated. The system attempts to match this input text with entries in its database. If a match is found, it displays the corresponding translated text. If a match is not found, the system displays the text letter by letter. After either displaying the translated text or the letters, the user is prompted to input another text. If the user chooses to input more text, the process loops back to the beginning; if not, the process stops.

This flow demonstrates a direct, database-driven approach to text-to-sign conversion. FIGURE 2. The Sign-to-Text flowchart [5] outlines the process for a sign-to-text conversion system. The process begins with the system scanning sign language gestures through a camera. It then attempts to match the recognized gesture with entries in its database. If a match is found, the corresponding text is displayed. If a match is not found, the system checks the sign letter by letter. After either outcome, the user is prompted to input another sign. If the user chooses to continue, the process loops back to scanning the sign language through the camera; if not, the process stops. This flowchart demonstrates a real-time, gesture-to-text translation workflow.

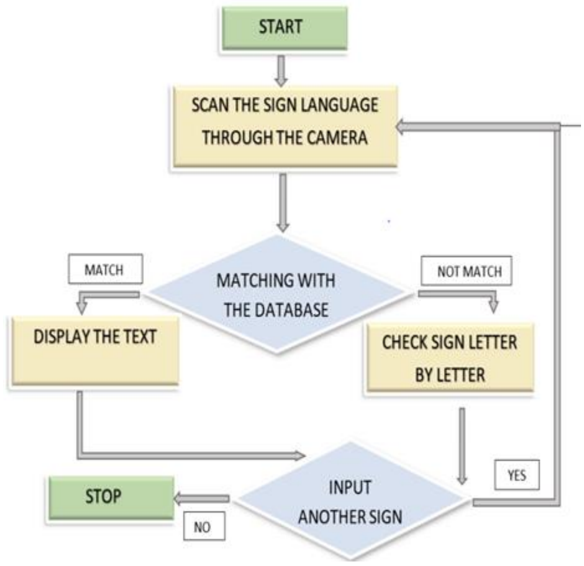

Figure 2 **Sign-to-Text Flowchart**

3. Results and Discussion

3.1. Results

The review of existing works highlights that most current systems perform well in controlled environments but face challenges in real-world use due to dataset scarcity, signer variability, and environmental noise. CNN-based models are effective for static gestures such as alphabets and digits, while video-based datasets and hybrid architectures offer better performance for continuous signing. However, real-time translation remains limited by computational complexity and the lack of large multilingual corpora. Studies also show that avatar-based sign synthesis often lacks naturalness, particularly in facial expressions and smooth motion, leading to mixed user acceptance. Sensor-based systems improve robustness but reduce usability due to cost and intrusiveness. Overall, the results suggest that while promising progress has been made, scalable multilingual speech-to-sign translation systems are still in an early developmental stage.

3.2. Discussion

The findings underline that while existing models demonstrate encouraging accuracy, practical deployment remains hindered by real-world variability in signing styles, lighting, and background conditions. A major limitation lies in the lack of large-scale, diverse, and multilingual datasets that can generalize across regions and sign languages. The reliance on high-computer architectures also raises concerns about real-time usability on low-resource devices, especially in developing contexts. Sensor-based approaches, though accurate, reduce accessibility due to cost and intrusiveness, highlighting a trade-off between precision and practicality. Similarly, avatar-based synthesis, while bridging the communication gap, often fails to capture expressive elements such as facial cues and smooth transitions, which are vital for natural communication. These gaps suggest a need for hybrid solutions that combine vision-based models with lightweight architectures and improved synthesis techniques. Moreover, participatory design involving Deaf communities is essential to ensure cultural and linguistic appropriateness. Future research should focus on creating standardized benchmarks and open datasets that encourage cross-lingual adaptability. With advancements in deep learning efficiency and multimodal integration, truly scalable speech-to-sign translation systems can become a reality in the near future.

Conclusion

This study shows that while notable progress has been made in speech-to-sign translation, current systems remain limited in scalability, accuracy, and real-world usability. Vision-based models perform well for static and isolated signs, but continuous signing and multilingual adaptation still pose challenges. Sensor-based and avatar-driven solutions add value but face issues of accessibility, naturalness, and cost. Addressing these gaps will require larger, standardized datasets, lightweight yet robust architectures, and closer involvement of Deaf communities. With continued advances in deep learning, multimodal integration, and inclusive design, future systems have strong potential to deliver real-time, cross-lingual sign translation, significantly improving accessibility for the Deaf and Hard of Hearing.

References

[1]. L. Anusha and Y. U. Devi, "Implementation of gesture-based voice and language translator for dumb people," 2016 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2016, pp. 1-4, doi: 10.1109/CESYS.2016.7889990.

[2]. Sudhan.M.B, S. Kumar K, U. Bala Varanasi, G. Nirmala, A. G. R and G. Karthikeyan, "Real-Time Sign Language Translator with IoT and Flex Sensors for Multilingual Support," 2025 7th International Conference on Inventive Material Science and Applications (ICIMA), Namakkal, India, 2025, pp. 268-273, doi: 10.1109/ICIMA64861.2025.11074077.

[3]. J. Peguda, V. S. S. Santosh, Y. Vijayalata, A. D. R. N and V. Mounish, "Speech to Sign Language Translation for Indian Languages," 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2022, pp. 1131-1135, doi: 10.1109/ICACCS54159.2022.9784996.

[4]. S. M. Antad, S. Chakrabarty, S. Bhat, S. Bisen and S. Jain, "Sign Language Translation Across Multiple Languages," 2024 International Conference on Emerging Systems and Intelligent Computing (ESIC),

Bhubaneswar, India, 2024, pp. 741-746, doi: 10.1109/ESIC60604.2024.10481626.

[5]. K. S. Sindhu, Mehnaaz, B. Nikitha, P. L. Varma and C. Uddagiri, "Sign Language Recognition and Translation Systems for Enhanced Communication for the Hearing Impaired," 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU), Bhubaneswar, India, 2024, pp. 1-6, doi: 10.1109/IC-CGU58078.2024.10530832.

[6]. Prof.(Dr.) Kaushika D Patel, Kasak Vaghasiya, and Radhika Savaliya, “Sign Language Translator with Speech Recognition Integration: Bridging the Communication Gap”, Int J Sci Res Sci Eng Technol, vol. 12, no. 2, pp. 741–749, Apr. 2025, doi: 10.32628/IJSRSET25122201.

[7]. Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur, Dr. Dayananda P, 2020, Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCAIT – 2020 (Volume 8 – Issue 15), DOI: 10.17577/IJERTCONV8IS15042

[8]. A. Yadav, R. Saxena, B. Saini, V. K. Verma and V. Srivastava, "Audio to Sign Language Translator Web Application," 2021 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 2021, pp. 321-326, doi: 10.1109/ComPE53109.2021.9751857.

[9]. F. M. Najib, "A multi-lingual sign language recognition system using machine learning," Multimedia Tools and Applications, vol. 84, no. 23, pp. 27987-28011, 2025, doi: 10.1007/s11042-024-20165-3.

[10]. M. Alaghband, H. R. Maghroor and I. Garibay, "A survey on sign language literature," Machine Learning with Applications, vol. 14, p. 100504, 2023, doi: 10.1016/j.mlwa.2023.100504.