IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14388

Intelligent Sign Language Video Generation Using Seq2Seq and NLP Techniques

Talla Nikhil Babu[1], S. Aharon Kumar[2], K. Sai Rakesh[3], M. Yenosh Kumar[4], B Avinash[5]

UG Student, Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Peddakakani Mandal, Nambur, Guntur- 522508 Andhra Pradesh, India[1,2,3,4,]

Assistant Professor, Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Peddakakani Mandal, Nambur, Guntur- 522508 Andhra Pradesh, India[5]

Abstract: This system fills the communication gap between the deaf and hearing population by interpreting oral language into Indian Sign Language (ISL) video streams, making it more accessible. It begins with Whisper ASR, a transformer-based automatic speech recognition algorithm that transcribes speech to accurate text. The transcribed text is then processed with Natural Language Processing (NLP) methods, such as tokenization, part-of-speech tagging, stop-word elimination, and lemmatization, to prepare the structure for ISL translation. For additional enhancement of compatibility with ISL grammar, a Sequence-to-Sequence (Seq2Seq) model with Recurrent Neural Networks (RNNs) is employed for restructuring sentences to produce fluent and natural translations. The optimized text is then translated into pre-recorded ISL video clips, and MoviePy performs the seamless stitching and synchronization of sign segments. A web interface based on Flask offers users a simple platform to upload audio files, input text, and create ISL videos in real-time. The system is optimized for efficiency and ease of use, and it can be useful in education, healthcare, government services, and customer support. The future developments will emphasis on developing the ISL vocabulary dataset, real-time processing optimization, mobile compatibility, and cloud platforms deployment for improved scalability. These enhancements will make the system more efficient, accurate, and accessible, further improving communication for the deaf and hard-of-hearing.

Keywords: Speech-to-Sign Language, Indian Sign Language (ISL), Seq2Seq Model, Natural Language Processing (NLP), Audio-to-Video Conversion, Gesture Recognition, Whisper ASR, Deep Learning.

I. INTRODUCTION

In human interaction, communication is pivotal to bridge the gaps among people, convey information, and express emotions. For the hearing-impaired and speech-impaired individuals, communication is blocked with the general lack of information and awareness about Indian Sign Language (ISL). The mute and deaf population in India utilize ISL as their primary mode of expression. But low levels of adoption, lack of knowledge, and the lack of organized learning materials greatly impede their participation in everyday activities, access to basic services, involvement in workplaces, and quality education in schools.

Despite major advances in assistive technology, most contemporary speech-to-text and sign language translation systems are severely constrained. They comprise language constraints, poor accuracy in noisy environments, no contextual understanding, and lack of usefulness in real-time domains. Most current systems also cater mainly to English speakers, and this reduces their applicability in India's multilingual culture, where several regional languages are used and preferred for communication.

To address these challenges, this research introduces an advanced ISL video-to-speech and speech-to-ISL video translation system. The proposed framework integrates Whisper ASR for highly accurate speech-to-text conversion, Natural Language Processing (NLP) for refining textual output, and a Seq2Seq model for restructuring sentences to align with ISL vocabulary and grammar conventions. Unlike conventional systems that rely on 3D animated avatars, this approach maps optimized text to pre-recorded ISL video clips, using MoviePy to assemble them into cohesive and seamless sign language videos. This approach guarantees that translations sound natural, contextually correct, and simple to comprehend by ISL users, making the entire experience of communication more intuitive and user-friendly.

One of the most compelling aspects of this framework is its support for multiple Indian languages, enabling users to give voice inputs in Indian languages and receive ISL video translations in real-time. This is one of the biggest advantages this system has over current systems, which mostly support English users and neglect India's linguistic diversity.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14388

By becoming more inclusive of sign language translation, the suggested solution promotes social communication, makes governmental services more accessible, enhances healthcare interactions, makes deaf education more accessible, and simplifies customer service interactions.

To further enhance efficiency, accuracy, and usability, future development efforts will focus on expanding the ISL vocabulary dataset, optimizing real-time processing capabilities, improving gesture alignment, ensuring mobile compatibility, and deploying the system through cloud-based platforms for large-scale accessibility. This research presents a robust, efficient, and user-friendly solution for bridging the communication gap between hearing individuals and the deaf community, ultimately fostering greater inclusivity, accessibility, and widespread adoption across various domains while significantly improving real-world communication experiences.

II. LITERATURE REVIEW

To enhance communication accessibility for the deaf and mute population, various scholars have explored various methods for sign language and speech-to-text translation. Youhao Yu (2012) examined the application of hidden Markov models (HMMs) for voice recognition in different environments, including cars, phones, and classrooms. Many contemporary voice-to-text applications are based on his research, which showed how HMM-based systems may increase speech recognition accuracy [1]. Similarly, to recognize Indian Sign Language (ISL) numbers, Madhuri Sharma, Ranjna Pal, and Ashok Kumar (2014) used neural networks and k- Nearest-Neighbor (kNN) classifiers. Their method demonstrated the efficacy of machine learning approaches in ISL recognition with an astounding 97.10% accuracy [2].

Purva C. Badhe and Vaishali Kulkarni (2015) advanced the field further by introducing a gesture detection system that tracks hand motions utilizing the YCbCr color model and Fourier Descriptors. Their approach demonstrated the importance of extraction of features and template matching in enhancing gesture-based language translation by effectively translating ISL motions into text in English with a 97.5% accuracy rate [3]. Neha Poddar et al. (2015) looked into machine vision-based sign language identification in a different study, noted the need for real-time processing and scalability in sign language translation systems [4]. In spite of these advances, most of the existing systems are plagued with their inability to process data in real-time, language dependency, and coverage of vocabulary.

Amit Kumar Shinde and Ramesh Khagalkar (2015) worked on computer vision-based sign language-to-text and vice versa recognition.Their work focused on the importance of better voice recognition accuracy in noisy environments and on the importance of multi-language support [5]. In addition, Mahendra S. Naik and Sulabha M. Naik (2016) advocated AI-based translating platforms to extend accessibility and recognized the role of assistive technology in rehabilitating children with hearing impairments [6]. The paper [7] suggests a two-way communication system, but the application of CNN models by the authors only provides an accuracy rate of 99.78% in converting 26 alphabets and three characters. For converting voice to sign language, the authors simply suggest that more research should be conducted in the field of natural language processing.

The authors of the paper [8] propose a method that converts sign language into Malayalam and English. The authors of the paper suggest the use of an Arduino Uno, which identifies the signs with gloves and translates them from ISL to the preferred language. The system's capacity to detect motion and two-hand signals makes it worthwhile. In another study, the authors convert ISL signals to text with the HSR model. Though the HSR model is better than RGB-based models, its performance ranges from 30% to 100% depending on factors such as light, position of hands and fingers [9]. A system with the capability of recognizing 26 ASL signals and converting them into English text is suggested by the authors of the [10] paper. They use MATLAB to locate signs using principal component analysis. The technology employs VRML avatars with the ASL to sign languages composition tool that plays them in a BAP player. The chief problem of the technology is that the existing VRML avatars cannot execute very complex maneuvers. For example, the current mechanism disallows the extension of the hand towards any part of the body [11]. A video-based system for translating signals between ISL, BSL, and ASL attained a general accuracy of 92.4%, per another study referenced in [12]. The program uses CNN and RNN to identify dynamic cues in real time. Upon translating the signs into text, the system offers the user an audio response through a text-speech API. The camera of Microsoft Kinect 360 is first used by the authors of another study to capture the motion of the ISL indicators.

The authors' Blender 3D animation is presented through a Unity engine. The technology can efficiently convert words into sign language, but it cannot convert sentences or lots of phrases into ISL[13]. Another bilateral sign language system is that of the authors of [14]. The technology has the ability to translate sign language into text or audio with a 97% accuracy rate. Once it gets the keywords from the input, the system applies the Unity engine to generate a 3D figure. Authors use the Google API to convert voice to text [15].

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025
DOI: 10.17148/IJARCCE.2025.14388

The authors of the paper [16] also propose another technology that converts Malayalam text and gives a 3D animated avatar in sign language. As HamNoSys notation is the main framework of the signs, it is implemented by the system [17]. Semantic analysis techniques are employed in a new Russian text to Russian sign language [18] system to convert text into sign language. It gives priority to lexical meanings of terms. Expanding on these frameworks, our research creates a multilingual speech-to-Indian Sign Language (ISL) translating system through the integration of state-of-the-art and Natural Language Processing (NLP) methods. The approach enhances communication as a more seamless and accessible process by employing an ML-based text optimization model that reformulates input sentences for optimal utilization of ISL word mapping available, providing efficient sign language generation.

III. METHODOLOGY

A. DATASET

The quality and diversity of the datasets used for training and assessment have a significant impact on the efficacy of any speech-to-sign translation system. We use a large dataset in this work that includes text transcriptions, multilingual voice samples, and related Indian Sign Language (ISL) movement films. To provide comprehensive vocabulary coverage, the collection contains pre-recorded ISL video clips for popular words, alphabets, and phrases. Multilingual support is enhanced by adding custom-recorded Indian language voice datasets to open-source speech samples from LibriSpeech and Mozilla Common Voice. The ISL gesture dataset ensures precise mapping between textual and visual representations by hand-annotating video sequences from existing sign language corpora. In order to improve the resilience of the model, the dataset was supplemented with noise injection, speed fluctuations, and pitch modulation. This allowed the system to function effectively in noisy real-world settings.

B. OBJECTIVE

The goal of this project is to create an *Indian Sign Language (ISL) Translator that translates ISL gestures accurately into English text based on state-of-the-art gesture recognition. The system intends to increase deaf-mute people's access to communication by creating a hassle-free means to translate sign language into a common language. To obtain high precision, the project integrates strong gesture recognition

algorithms based on methods like Fourier Descriptors, motion tracking, skin color recognition, and template matching. The system utilizes a self-designed ISL gesture database to ensure efficient recognition to facilitate the processing of various gestures with low error rates. Moreover, real-time processing is also incorporated to support rapid and effective translations, which makes the system applicable for everyday life. The natural interface is designed to allow for easy input and interpretation of gestures so that it will be usable by all. Besides, the project is scalable such that future improvements such as the inclusion of deep learning models and an expanded gesture vocabulary can be implemented to make it more accurate and flexible. Overall, this project is beneficial in assistive technology in terms of bridging the communication gap between society and the deaf-mute community, enhancing inclusiveness and appropriate interaction.

C.PROPOSED SYSTEM

The Proposed System introduces an advanced Speech-to-Sign Language Translation System that enables improved communication between the deaf-mute population and non-sign language users. The system addresses the shortcomings of existing solutions using Whisper ASR for speech recognition, Natural Language Processing (NLP) for sentence optimization, and Indian Sign Language (ISL) video generation to ensure accurate and real-time translation. Unlike typical systems that deal mostly with English, this system gives support to more than one language, enhancing access for non-native English speakers.

The system architecture comprises the following main components:

1. Input Module:

The input module provides users with the ability to enter data either as text or audio. When the user inputs an audio, it is converted from speech-to-text using Whisper ASR, a powerful automatic speech recognition engine. The system handles the input to ensure that it is processed well, whether through direct text typing or audio typing, before proceeding to the next phase.

2. Speech-to-Text Processing:

This module is responsible for extracting meaningful features from audio input and transcribing speech into structured text output. The system employs Whisper ASR, which leverages deep learning-based acoustic and language models to maximize accuracy and support various speech variations. The decoder transcribes audio features extracted with the help of a lexicon and language model, offering high-accuracy transcription even in noisy environments.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14388

3. Text Preprocessing with NLP:
Once it receives the transcribed text, the NLP module preprocesses it to conform to Indian Sign Language (ISL) grammar. The process entails tokenization (division of text into words), POS tagging (word category identification), stop-word removal (removal of unnecessary words such as "is" and "the"), and lemmatization (reduction of words to their base form). Because ISL has a different grammatical structure than English, this module ensures correct sign representation through linguistically aligned translated text.

4. ISL Video Generation:
After the optimized text is prepared, this module translates the words into pre-recorded ISL video clips. With the help of MoviePy, a powerful video editing library, the system synchronizes and stitches the sign language gestures into a coherent video output. This module is critical in providing smooth and contextually correct sign translations, which makes communication easy for ISL users.

5. Output Module:
The processed ISL video is finally output to the user via the output module. The avatar-based sign language interpretation provides a visually appealing and accessible translation. The module provides a seamless display of the generated sign language video so that non-signers and the deaf-mute community can communicate better.

The system greatly enhances convenience for the deaf-mute community by providing a fast and precise speech-to-sign language translation. Utilizing ASR, NLP, and video synthesis, the solution ensures that oral communication can be translated into ISL without difficulty, bringing down communication hurdles. Possible future advancements may involve using deep learning-based sign synthesis and real-time avatar animation in order to make the translation process even more natural.

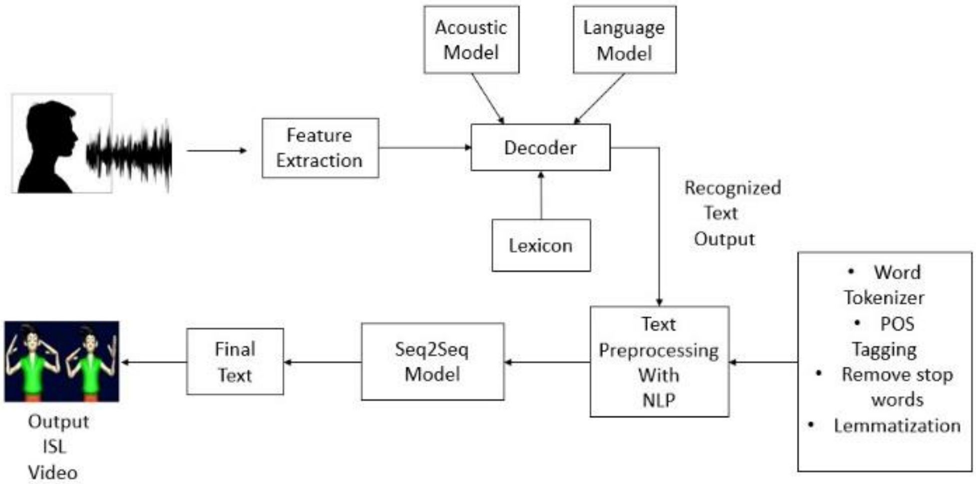

Fig- 1 Proposed System

D. AlGORITHM STACK

1. Audio Extraction & Speech Recognition (Whisper ASR & Feature Extraction)
The initial step in the system workflow is audio extraction from user input, either a direct speech recording or an uploaded audio file. Whisper ASR is utilized for speech-to-text conversion, using a transformer-based deep learning model that improves transcription accuracy even in noisy conditions. This module guarantees that speech is properly recorded and salient features are obtained to enhance recognition. Lexicon and language models also help fine-tune the transcription by error correction and ensuring contextual consistency to provide an accurate and meaningful output..

2. Natural Language Processing (NLP) for Text Optimization
After converting the spoken input into text, NLP methods are used to make the text optimal so that it is consistent with Indian Sign Language (ISL) grammar. This module carries out tokenization, part-of-speech (POS) tagging, removal of stop-words, and lemmatization through NLTK and special NLP algorithms to reshape sentences. Because ISL employs a different grammatical structure than spoken English, unnecessary words like auxiliary verbs are omitted and sentences rearranged to follow ISL conventions, promoting smooth and natural sign translations.

DOI: 10.17148/IJARCCE.2025.14388

3. Seq2Seq Model for Sentence Reorganization

To further refine the optimized text and align it more effectively with ISL grammar, a Sequence-to-Sequence (Seq2Seq) model is implemented. This deep learning-based approach is trained on English-to-ISL sentence transformations and helps restructure input text while preserving meaning. The model consists of an encoder-decoder architecture, where the encoder captures the semantic context of the input sentence, and the decoder generates an ISL-compatible sentence with proper word order. By learning from ISL sentence structures, the Seq2Seq model improves fluency, ensuring that translated sentences are both grammatically correct and contextually accurate for ISL.

4. Sign Language Video Synthesis (MoviePy & Pre-Recorded ISL Clips):

Once the text is processed, the system creates a sign language representation by translating words into pre-recorded ISL video clips. Utilizing MoviePy, a powerful video processing library, individual sign clips for every word are cut out, synchronized, and joined together to create a cohesive ISL video sequence. The system achieves smooth transitions between signs, with natural flow and contextual appropriateness, so the produced video properly expresses the intended message.

5. Avatar-Based Output Generation:

The last step of the process is the presentation of the synthesized ISL video via an avatar-based interface. The system provides for the seamless display of the translated sign language video to the user, offering an accessible and intuitive means of communication for people who depend on ISL. Through the incorporation of sophisticated video assembly mechanisms and sign language synthesis, this module fills the gap in communication between spoken language users and the deaf or hard-of-hearing population.

IV. IMPLEMENTATION

1. Tools and Technologies

The creation and implementation of the suggested speech-to-sign language translation system need a mixture of sophisticated tools and technologies. Python 3 serves as the central programming language because of its enormous library ecosystem applicable to speech recognition, natural language processing (NLP), and video processing. The Flask web framework is used for developing a light and scalable backend so that communication among modules can be smooth. Whisper ASR is used to perform accurate speech-to-text conversion, whereas NLTK (Natural Language Toolkit) is used to preprocess and optimize text to conform to Indian Sign Language (ISL) grammar. MoviePy is utilized for video processing and combining sign language videos from pre-recorded ISL clips. The system also includes MySQL for database management to efficiently manage user input and processed information. The development environment is mostly based on Visual Studio Code, which offers a strong coding, debugging, and testing interface.

2. Dataset and Preprocessing

The system employs a structured dataset of pre-recorded ISL video clips to produce sign language translations. Preprocessing involves extraction of audio, noise elimination, normalization of text, and ISL-text restructuring aligned to ISL to increase precision. The speech-to-text process is enhanced so as to exclude unrequired words and auxiliary verbs to maintain coherence in the sentence. The system then aligns identified words with their matching ISL video clips so as to achieve natural and flowing sign language depiction. The data is formatted and optimized for clean word transitions and grammatical consistency in ISL.

3. Machine Learning Model

The system relies on Transformer-based deep neural models for language understanding through speech-to-text and text optimization. Whisper ASR, which is multilingual and noise-robust, provides accurate speech-to-text transcription even in poor audio conditions. The identified text is then processed by NLTK-based NLP methods, reformulating the sentences to adhere to ISL grammar rules. Additionally, a Seq2Seq model is employed to further refine the sentence structure, ensuring optimal reorganization for ISL grammar by leveraging an encoder-decoder architecture. The video generation module utilizes MoviePy to splice pre-recorded ISL video clips, synchronizing them to reflect the optimized text output. This AI-based method allows for a contextually accurate and fluent sign language translation system.

4. System Deployment and User Interaction

The system is hosted locally with the Flask framework and tested, where efficient processing and smooth user experience are guaranteed. The web-based interface offers an easy-to-use experience, whereby users can:

- **Upload text or audio inputs** for translation into Indian Sign Language (ISL).
- **Preview the ISL-generated video**, which reflects the given spoken or written input accurately.
- **Download** the translated ISL video for further use.

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⫶ Peer-reviewed & Refereed journal ⫶ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14388

By combining AI-based automation, the Speech-to-Sign Language Translation System enables real-time communication between speakers of spoken languages and the deaf or hard-of-hearing population. The system offers correct, contextually relevant, and fluid translations that enhance access to sign language material and effectively bridge the communication gap.

V. RESULTS

A. System Performance and Evaluation

The Speech-to-Sign Language Translation System was thoroughly tested with varied speech inputs such as deviating pronunciation, noise in the background, and varying sentence constructions. The system was tested on various performance measures to ensure accuracy and fluency of speech recognition, text optimization, and generation of Indian Sign Language (ISL) videos. Proper testing ensured the generation of contextually correct ISL videos by the system, ensuring linguistic consistency and correct sign representation.

B. System Efficiency and Processing Time

In order to determine the efficiency of the system, a number of important factors were examined:

- **Processing Speed:** The processing time taken to transcribe speech into text, optimize the sentence for ISL, and produce the resultant sign language video was measured. The system exhibited effective real-time processing, with a considerable reduction in the amount of manual effort needed for ISL translation..
- **Accuracy of ISL Videos through Generation:** The ISL videos generated were assessed for accuracy and grammatical consistency with ISL syntax, so they could properly deliver the intended message.
- **Scalability:** The system was subjected to different sentence lengths and levels of complexity, proving its capability to process varied inputs while upholding performance and delivering smooth, synchronized sign language animation.
- **User Experience:** The usability of the online interface was tested, providing smooth interaction for users submitting speech or text inputs and accessing the resulting ISL videos. The system provides an accessible, user-friendly, and reliable platform for closing gaps in communication among users of spoken language and the deaf or hard-of-hearing population.

C. Output Screens

1. Home Page of the Web Application

The Indian Sign Language Video Generator home page offers an easy, easy-to-use, and intuitive user interface to generate ISL videos easily. Users can upload an audio file directly from their local machine to generate a corresponding Indian Sign Language (ISL) video. The interface has a simple "Choose File" button for choosing the desired audio file and then there is a "Generate Video" button to trigger the processing. Upon uploading the file, the system performs speech-to-text conversion, applies text optimization, and generates the ISL video using pre-recorded ISL video clips. The backend ensures accurate transcription, smooth word alignment, and proper synchronization of sign gestures. After the process is finished, the ISL video is shown on the same page so that users can preview and download it. The organized but simple design allows for a smooth workflow, and the generation of sign language videos is both user-friendly and highly efficient.

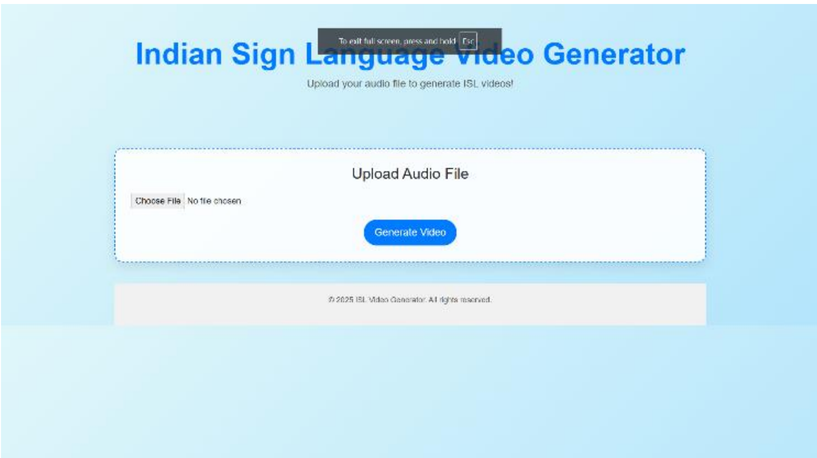

Fig- 2 Home Page

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14388

2. Text File or Audio File Upload and Processing

Users can upload a local audio file to produce an Indian Sign Language (ISL) video with ease. After the audio file is uploaded, the system takes care of processing it by extracting the speech from it, converting it to text using Whisper ASR, and optimizing the text to conform to ISL grammar. This optimization is taken care of so that the translated text remains well-structured and meaningful in ISL. The backend handles correct transcription, word alignment, and smooth video synthesis with MoviePy so that individual sign clips are properly placed to create a smooth ISL translation. Users are provided with real-time status updates during the process, which informs them of every processing stage. After the ISL video is created, it is shown on the same page for users to preview before downloading to ensure accuracy. This simplified process removes redundant steps, making it easier to create sign language videos, more efficient, accessible, and user-friendly .

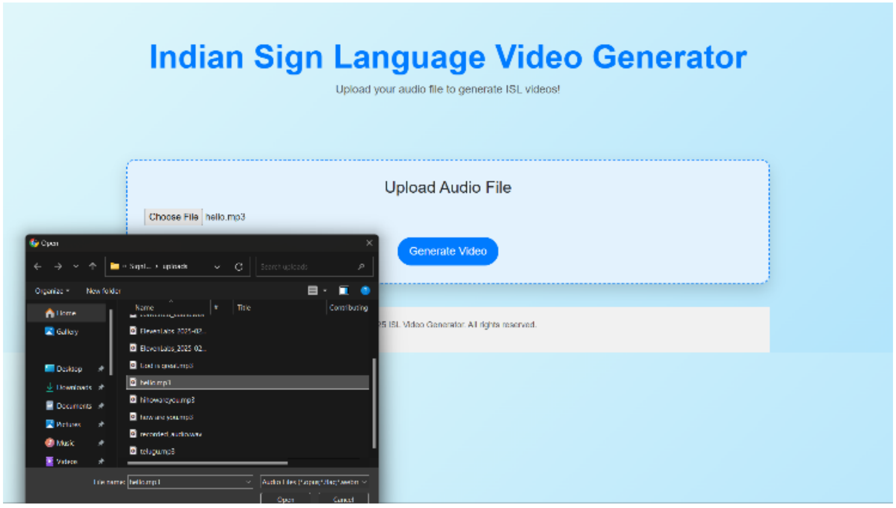

Fig- 3 Audio or Text file Upload Page

3. ISL Video Generation Output

After processing, the system generates an Indian Sign Language (ISL) video equivalent to the uploaded audio file, ensuring an accurate and accessible translation. The translated video is displayed on the same webpage, allowing users to preview the ISL gestures before downloading. The system ensures precise speech-to-text conversion, proper word alignment, and seamless video synthesis using pre-recorded ISL clips. Advanced processing methodologies aid in retaining even transitions among signs, ascertaining fluent communications. Final delivery ensures that video quality and synchronizing accuracy for gestures remain very high and it can thus fit in to provide solutions within applications such as accessibility, learning, and communication by and among deaf and hard-of-hearing.

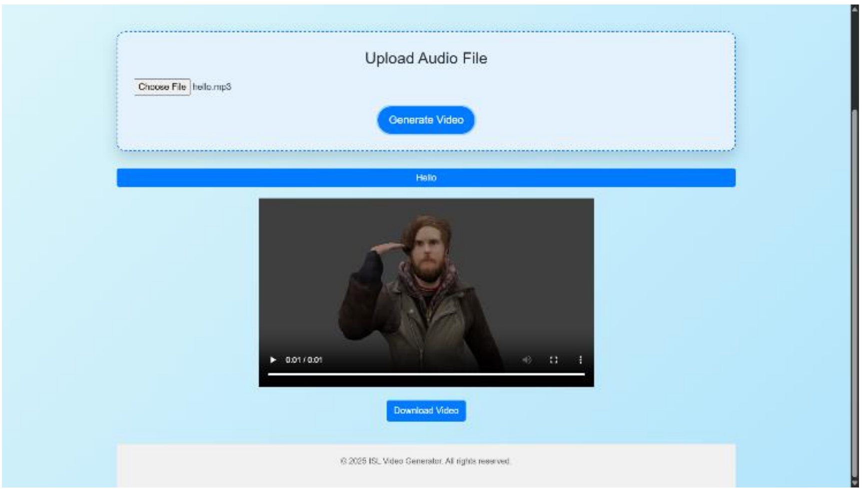

Fig- 4 Generated ISL Video Output Page

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14388

D. Comparison with Existing Systems

Table 1 Analysis Result

Feature	**Manual Dubbing**	**Commercial AI Tools**	**Proposed System**
Processing Time	High	Medium	Low
Cost	Expensive	Costly	Affordable
Automation Level	Low	Medium	High
Video Quality Retention	High	Medium	High
Translation Accuracy	High	Medium	High
Voice Synchronizatin	High	Medium	High

VI. CONCLUSION

The suggested speech-to-Indian Sign Language (ISL) translating system fills the communication gap among the hearing and deaf-mute populations by synergizing deep learning, natural language processing (NLP), and speech recognition strategies. In contrast to conventional systems with restricted languages and fixed representation, this platform offers real-time processing, language support for various languages, and dynamic ISL video creation for inclusivity and accessibility.

The system utilizes Whisper ASR for accurate speech-to-text translation even in noisy settings, while a Seq2Seq model rearranges sentences to match ISL grammar. Rather than 3D avatars, the system uses pre-recorded ISL video clips, which are dynamically constructed with MoviePy to provide natural and expressive sign language expression. A Flask-based web interface offers an easy-to-use platform for users to enter text or audio and get real-time ISL translations, improving usability in education, public services, and social interactions. With its scalable and modular design, this system can be further developed to cover more ISL vocabulary, handle more languages, and add cloud-based deployment for enhanced accessibility. Real-time gesture recognition and more comprehensive dataset coverage are future prospects for refining the accuracy of translations.

In total, this project is a huge leap toward an inclusive world, giving power to the speech and hearing impaired by facilitating more communication through assistive technology. By developing speech-to-sign language translation, this system encourages better accessibility and inclusivity, allowing ISL users to communicate freely with the world.

VII. FUTURE WORK

To better enhance the Indian Sign Language (ISL) Video Generator, future advancements will be towards broadening its functionality and enhancing translation precision:

1. **Enlarged ISL Dataset**– Building the vocabulary through the addition of more words, phrases, and technical terms in order to cover various translation requirements.
2. **Advanced Deep Learning Models** – Utilizing sophisticated AI methods for context-sensitive sentence reordering, providing improved ISL grammar compatibility and enhanced translation precision.
3. **Real-Time Sign Language Recognition** – Adding gesture input recognition in addition to speech-to-ISL conversion, facilitating smooth two-way communication.
4. **Mobile Optimization** – Optimizing the system for mobile platforms with ow-latency processing, providing real-time interaction across platforms.
5. **Cloud-Based Deployment** – Scaling the system using cloud platforms like AWS or Google Cloud, ensuring real-time accessibility and executing multiple requests in parallel.
6. **Public Service Integration**– Increasing the usages of the system in education, healthcare, and public services, making assistive communication tools widely available.
7. **Cooperation with ISL Experts**– Collaboration with sign language experts and linguists to normalize translations and improve naturalness and contextual appropriateness**.**

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 3, March 2025

DOI: 10.17148/IJARCCE.2025.14388

REFERENCES

[1]. Amit kumar shinde and Ramesh Khagalkar "sign language to text and vice versa recoganization using computer vision in Marathi" International journal of computer Application (0975-8887) National conference on advanced on computing (NCAC 2015).

[2]. Sulabha M Naik Mahendra S Naik Akriti Sharma " Rehabilitation of hearing impaired children in India"International Journal of Advanced Research in Computer and Communication Engineering.

[3]. Neha Poddar, Shrushti Rao, Shruti Sawant, Vrushali Somavanshi, Prof. Sumita Chandak "Study of Sign Language Translation using Gesture Recognition" International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 2, February 2015.

[4]. Kunjumon, J.; Megalingam, R.K. Hand gesture recognition system for translating indian sign language into text and speech. In Proceedings of the 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 27–29 November 2019; pp. 14–18.

[5]. Gangadia, D.; Chamaria, V.; Doshi, V.; Gandhi, J. Indian Sign Language Interpretation and Sentence Formation. In Proceedings of the 2020 IEEE Pune Section International Conference, PuneCon 2020, Pune, India, 16–18 December 2020; pp. 71–76. [CrossRef]

[6]. Shangeetha, R.K.; Valliammai, V.; Padmavathi, S. Computer vision based approach for Indian sign language character recognition.

[7]. In Proceedings of the 2012 International Conference on Machine Vision and Image Processing, MVIP, Coimbatore, India, 14–15 December 2012; pp. 181–184. [CrossRef]

[8]. Sawant, S.N.; Kumbhar, M.S. Real time sign language recognition using PCA. In Proceedings of the 2014 IEEE International Conference on Advanced Communication, Control and Computing Technologies, ICACCCT 2014, Ramanathapuram, India, 8–10 May 2014; pp. 1412–1415. [CrossRef]

[9]. Papadogiorgaki, M.; Grammalidis, N.; Tzovaras, D.; Strintzis, M.G. Text-to-sign language synthesis tool; Text-to-sign language synthesis tool. In Proceedings of the 2005 13th European Signal Processing Conference, Antalya, Turkey, 4–8 September 2005.

[10]. Sonare, B.; Padgal, A.; Gaikwad, Y.; Patil, A. Video-based sign language translation system using machine learning. In Proceedings of the 2021 2nd International Conference for Emerging Technology, INCET, Belagavi, Inida, 21 May 2021. [CrossRef]

[11]. Kanvinde, A.; Revadekar, A.; Tamse, M.; Kalbande, D.R.; Bakereywala, N. Bidirectional Sign Language Translation. In Proceedings of the International Conference on Communication, Information and Computing Technology, ICCICT, Mumbai, India, 25–27 June 2021. [CrossRef]

[12]. Qi, J.; Wang, D.; Jiang, Y.; Liu, R. Auditory features based on gammatone filters for robust speech recognition. In Proceedings of the 2013 IEEE International Symposium on Circuits and Systems (ISCAS), Beijing, China, 19–23 May 2013.

[13]. Nair, M.S.; Nimitha, A.P.; Idicula, S.M. Conversion of Malayalam text to Indian sign language using synthetic animation. In Proceedings of the 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), Kottayam, India, 1–3 September 2016.

[14]. Qi, J.; Wang, D.; Xu, J.; Tejedor Noguerales, J. Bottleneck features based on gammatone frequency cepstral coefficients. In Proceedings of the International Speech Communication Association, Lyon, France, 5–29 August 2013.

[15]. Grif, M.; Manueva, Y. Semantic analyses of text to translate to Russian sign language. In Proceedings of the 2016 11th Inter-national Forum on Strategic Technology, IFOST, Novosibirsk, Russia, 1–3 June 2016; pp. 286–289.

[16]. Dhanjal, A.S.; Singh, W. Comparative Analysis of Sign Language Notation Systems for Indian Sign Language. In Proceedings of the 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Sikkim, India, 25–28 February 2019.

[17]. Varghese, M.; Nambiar, S.K. English to SiGML conversion for sign language generation. In Proceedings of the 2018 Interna-tional Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), Kottayam, India, 21–22 December 2018.

[18]. Raghavan, R.J.; Prasad, K.A.; Muraleedharan, R. Animation System for Indian Sign Language Communication using LOTS Notation. In Proceedings of the 2013 International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications (C2SPCA), Bangalore, India, 10–11 October 2013).