

Sign Speak: Speech To Sign Language Conversion System

Dr. Narendra Kumar S¹, Nisarga G², Nishmitha N Pai², Nuthana A M², Pallavi Naik²

¹Assistant Professor, Department of Computer Science and Engineering, JNNCE, Shivamogga, Karnataka, India

²UG Students, Department of Computer Science and Engineering, JNNCE, Shivamogga, Karnataka, India

Abstract— Communication remains a significant barrier for individuals who are deaf, particularly in scenarios where sign language is not widely understood. These barriers often prevent full participation in essential activities such as daily conversations, video conferencing, gaming, or professional meetings, resulting in social and informational isolation. To address this challenge, our project proposes an NLP-based system designed to convert oral English into Indian Sign Language (ISL). The model uses speech recognition to capture audio input, translates it into text, and then translate the text into animated ISL representations. This enables real-time communication by displaying both textual and visual sign outputs, bridging the gap between hearing individuals and the deaf community. In addition to supporting the deaf community, the system incorporates functionalities that assist users with visual impairments, further widening its impact. By combining Natural Language Processing, speech recognition, and an intuitive user interface, this system represents a forward-thinking approach to inclusive, accessible communication technologies.

Index Terms— *Indian Sign Language (ISL), Real Time Translation, Sign Language Translation, Speech to Text, Natural Language Processing(NLP).*

I. INTRODUCTION

Communication is a problem for a person who are deaf, particularly in scenarios where sign language is not widely recognized. These barriers often prevent full participation in essential activities such as daily conversations, video conferencing, gaming, or professional meetings, resulting in social and informational exclusion.

To address this challenge, our project proposes an NLP-based system designed to convert oral language into Indian Sign Language (ISL). The model allows to use the speech recognition to capture audio input, translates it into text, and then converts that text into animated ISL representations. This enables real-time communication by displaying both textual and visual sign outputs, bridging the gap between hearing individuals and the deaf community.

In addition to supporting the deaf community, the system incorporates functionalities that assist users with visual impairments, further widening its impact. By combining NLP, speech recognition, and an intuitive user interface, this system represents a forward-thinking approach to inclusive, accessible communication technologies.

II. LITERATURE SURVEY

The area of automated conversational system application in converting speech, text into sign language has been adequately researched in an attempt to bridge the gap that exists between the deaf or non-verbal community and the rest of the population. This section aims to discuss a few of those studies, prominent in their ways, and how they can be beneficial to recover and make sense of the research being done on the Machine Learning.

[1] Sign Language Transformers: Joint End-to-end sign Language Recognition and Translation

N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden have done a study on “sign language Transformers that use end-to-end sign language recognition and translation”. The system in the paper centers on a unified, end-to-end transformer-based architecture that works on the concept of Continuous Sign Language Recognition (CSLR) and Sign Language Translation (SLT)[1]. The input sign language video is first processed with the help of pretrained Convolutional Neural Network to extract embeddings of each frame. These embeddings are sequentially encoded and it is fed into the SLRT, a transformer encoder that learns spatio-temporal features from the video sequence. To enable recognition of sign glosses, a Connectionist Temporal Classification (CTC) loss is applied to the output of SLRT, allowing the model to align sequences without needing precise frame-level annotations. The encoder’s output representations are then passed to the SLTT, an autoregressive transformer decoder that generates spoken language sentences one word at a time.

[2] Progressive Transformers for End-to-End Sign Language Production

B. Saunders, N. C. Camgoz, and R. Bowden proposed the “Progressive Transformers of End-to-End Sign language production” [2] which focuses on translating spoken language text directly into 3D sign poses sequences. The model consists of two main components: one that translates text to sign pose and another which uses an intermediate gloss representation. The concept used here is the counter decoding mechanisms, which enables the model to generate continuous output of variable length by predicting a counter value that helps to

signifies the progress through the output sequence. Therefore, it eliminates the need for fixed output sizes or discrete end-of-sequence tokens. It also incorporates the data augmentation techniques such as future frame prediction and Gaussian noise to combat prediction drift during sequence generation. Analysis of model output is done using a novel back translation approach, where the generated sign sequences are translated back into spoken language using a pre-trained sign language translation model to access accuracy and intelligibility.

[3] Real time conversion of sign language to speech and prediction of gestures using Artificial Neural Network

A. Kumar, P. Maheshwari, and S. Raj use the technique of Artificial Neural Network for the people with hearing-impaired.[3].The system involves both hardware and software components to make real-time conversion of sign gestures into speech. It has a glove which is embedded with four flex sensors that captures hand gestures by detecting the degree of bending in each finger. Sensor readings are processed by an Arduino Uno microcontroller, by connecting to a SIM900A GSM module. The GSM module transmits the gesture data to both an Android application and a cloud platform. These Android apps receive the corresponding text messages and convert them into audible speech using text-to-speech functionality which enables communication between mute individuals and others. In parallel, the gesture data collected over time is fed into a back-propagation neural network developed in MATLAB. This neural network is trained using sensor values and corresponding time-of-day inputs to predict future gestures or needs of the user. By learning these individual patterns and behaviors, the system aims to become more adaptive and personalized over time as it improves both communication accuracy and predictive reliability.

[4] Advancing human-computer interaction: AI-driven translation of American Sign Language to Nepali using convolutional neural networks and text-to-speech conversion application

The work of Biplov Paneru, Bishwash Paneru and Khem Narayan Poudyal has done a significant work to propose an assistive system to translate American Sign Language to Nepali text and speech using deep learning and computer vision techniques[4]. It uses the pre-trained Convolutional Neural Networks- ResNet50 and VGG16 for classifying signs from a dataset and images are preprocessed through normalization, resizing, and stratified sampling. The GUI for the system was developed using Python's Tkinter, and the Nepali text output was converted to speech using Google's Text-to-Speech API.

[5] Towards Automatic Speech to Sign Language generation

"Towards Automatic Speech to Sign language Generation" was proposed by Parul Kapoor, Rudrabha Mukhopadhyay, Sindhu B. Hegde, Vinay Namboodiri, and C.V.Jawahar [5] in which they used three components:(1) a speech encoder which extracts feature from input melspectrograms, (2) a pose decoder that generates signer pose sequences, and (3) a text decoder that performs auxiliary speech-to-text translation. Along with these, a cross-modal discriminator is utilized to enhance the quality

and semantic alignment of generated pose sequences with speech input. The model is trained using a fusion of regression, cross-entropy and adversarial losses.

[6] Translating Speech to Indian Sign Language Using Natural Language Processing

The paper is proposed by R.K Reddy, K.R.R.Reddy ,R.R.Reddy and K.R.Reddy about the processing of actionable language[6]. The Natural Language Processing to convert spoken or typed language into Indian Sign Language videos. The process starts with audio-to-text conversion, followed by NLP-based text processing using segmentation, trimming, parsing, lemmatization, and part-of-speech tagging. The system attempts to match the processed input with a pre-recorded ISL video database. If it has no direct match then it shows signs letter by letter, ensuring that no content is lost in translation. Additionally, the system can identify common phrases and output a single sign video for them if available by improving translation fluency.

[7] Conversion of Sign Language to Text and Speech Using Machine Learning Techniques

The work done by V.A. Adewale and A.O. Olamiti[7] indicate a system that understands sign language gestures and converts them into both text and speech. Kinect sensor is applied to capture sign language images followed by image segmentation for the translation of the images into a binary format suitable for analysis. There are some of the features such as edges and corners are extracted from the Regions of Interest (ROIs), particularly focusing on the hands and face, using the FAST (Features from Accelerated Segment Test) and SURF (Speeded-Up Robust Features) algorithms. The characteristics are classified using the K-Nearest Neighbor (KNN) algorithm, includes both supervised and unsupervised learning techniques to improve recognition accuracy.

[8] Real Time Sign language to Speech Converter Using OpenCV and MediaPipe

The paper proposes "a real-time sign language to speech converter using OpenCV and MediaPipe", proposed by A. Pathak, A. Jadhav, N. Patil, S. Rukhande, P. Padhy and L.Gadhikar focusing on dynamic hand gesture recognition without relying on extensive gesture dataset [8]. The system uses a camera to capture gestures, processes the images using OpenCV, and tracks hand landmarks via MediaPipe. Logical conditions identify signs based on finger positions, which are converted to text and then to speech using a text-to-speech API. This system includes image capturing, preprocessing, gesture recognition and speech synthesis.

[9] An Efficient Speech to Sign Language Conversion and Text Recognition through Live Gesture

M. Kowsigan, R. Dhawan, and A. Kundu proposed the "An Efficient Speech to Sign Language Conversion and Text Recognition through Live Gesture" helps to share information with the people who are hard of hearing. Over 5% people globally affected by the problem of communication with deaf people.

The old methods such as converting the speech into text and text into speech fails to facilitate seamless communication among these peoples. The proposed system in this paper has speech-to-sign language conversion and live gesture-to-text recognition, using audio and visual inputs to enable real-time, bidirectional conversation among people. By using advanced concepts like NLP and Machine Learning, the system offers users to choose between American Sign Language and Indian Sign Language (ISL) which ensuring the system would work well for all users, regardless of their cultural background. This model uses advanced speech recognition techniques combined with Natural Language Processing (NLP) to translate the oral words into text, Machine Learning models are developed using Python which are trained on a custom datasets that features are personalized hand symbols for both American Sign Language (ASL) and Indian Sign Language (ISL), enhance the validity of speech-to-sign language conversion and gesture detection and also Convolutional Neural Networks (CNN) are implemented for real-time gesture-to-text recognition, which allows the system to interpret hand gestures accurately. Sign language plays a vital role across the globe in reducing the communication barrier between people with hearing disabilities or who have a speech impairment.

[10] Enhancing Communication for the Hearing Impaired: A Real-Time Speech to Sign Language Converter

The paper “Enhancing Communication for the Hearing Impaired: A Real-Time Speech to Sign Language Converter” which is proposed by A. Deshmukh, A. Machindar, S. Lale and P. Kasambe [10] introduces, SignLingo system that is designed for people having hearing and speech impairment. The SignLingo application is implemented using several advanced tools and technologies that enables real time speech is converted to sign language. It uses Natural Language Processing (NLP) techniques which include the Natural Language Toolkit (NLTK) and the JavaScript Web Speech API for live speech recognition, the WordNet lexical database that helps to extract meaningful information from the recognized text. The Indian Sign Language gestures are visualized using 3D animations and corresponding signs are matched based on keywords. In this model, the ReactJS is used for frontend which provides a smooth and user-friendly interface and also Flask is used for backend which manages data processing and system integration. The system enhances real-time communication and promotes accessibility for the person with hearing and speech impairments.

[11] A Novel Speech to Sign Communication Model for Gujarati Language

N. Aasofwala, S. Verma and K. Patel proposes a model in a Gujarati language [11] that is used to converts spoken Gujarati speech into actionable language system. It mainly focuses on bridging the gap by connecting the normal people and hearing-impaired individuals in Gujarat language. The author says about difficulties in communication between deaf peoples and the

normal person. There will be no middle persons to translate the sign to speech and vice-versa. To manage these challenges, they have come up with as model as solution that converts Speech to Sign for Gujarati Language. This model takes spoken Gujarati language through microphone, also removes the background noise and then changes speech into written Gujarati using speech recognition. The written Gujarati text is converted into sign language using HamNoSys, which shows hand shapes, directions and movements. Then, this is changed into SiGML, a special format that helps to show the signs visually using the 3D animated avatar. paper also uses Artificial Intelligence, natural language processing (NLP), and 3D animation to make communication is clearer. The authors feel that, speech-to-sign system can help the students in deaf schools, people with hearing loss in daily life, and also supports building a GSL dictionary.

[12] A Communication Platform Between Bangla and Sign Language

Good communication helps people to understand and connect with each other. But for those who unable to hear and have speech impairments, it might be hard to connect with common people using traditional methods. Hence R. Shahriar, A. G. M. Zaman, T. Ahmed, S. M. Khan, and H. M. Maruf proposed the A Communication Platform Between Bangla and Sign Language[12] to overcome the difficulties faced by these people especially people of Bangla in their daily life, this research paper introduces system which is smartphone app that sign language. Sphinx base provides the environment for speech recognition, and Sphinx train helps to train the models. Pocket sphinx is a smaller version that runs efficiently on mobile devices. A Bangla language model helps in processing common sentences, and a Bangla acoustic model uses recorded audio to improve the model that hears and understands speech. Also, Google Translates server is used to convert Bangla text into spoken words. The author used these tools, the app achieved good results of about 84.71% accuracy and also shows strong potential to help people with speech and hearing impairments. These technologies improve interaction among the people and also strengthen the community connections for all.

III. SUMMARY

The Conversion of oral to Sign language is a technology which translates spoken language into visual sign language using a combination of speech recognition. The model provides efficient communication between hearing individuals and the deaf or hard-of-hearing community by converting speech into sign language. It enhances accessibility in various field like education, healthcare. The technology aims to promote inclusivity by bridging the communication gap and supporting equal access to information.

Reference	Dataset	Tools Used	Machine Learning Algorithms Used	Evaluation Metrics	Outcomes/Performance
Camgoz et al. [1]	PHOENIX14T	CNN+ Transformer	Transformer (SLRT + SLTT), CTC Loss	End-to-end accuracy	High accuracy, scalable, data-intensive
Saunders et al. [2]	RWTH-PHOENIX-Weather 2014T	Transformer, Gloss-Free Input	Progressive Transformer with counter decoding	Back Translation Score	Robust, dynamic-length signs, gloss-free
Kumar et al. [3]	Glove Sensor Data	MATLAB, Arduino, GSM	Backpropagation Neural Network	Real-time, predictive accuracy	Affordable, wearable, gesture-to-speech
Biplov Paneru et al. [4]	68,400 ASL images	ResNet50, VGG16, OpenCV, Tkinter, gTTS	CNN (ResNet50, VGG16)	Classification accuracy, real-time response	99.99% accuracy (ResNet50), real-time output, no sensors/gloves needed; limitations: camera dependency, limited gesture complexity, scalability issues
Kapoor et al. [5]	Custom Indian Sign Dataset	PyTorch, OpenPose	Multi-task Transformer, Cross-modal Discriminator	DTW, PCK	Real-time ISL generation, speech direct input
Reddy et al. [6]	ISL Video Database	NLTK, WordNet, Custom Video Mapping	NLP (POS tagging, Lemmatization)	Translation Accuracy	Real-time ISL, fallback letter-by-letter
Adewale & Olamiti [7]	ASL Images	Kinect, MATLAB	FAST, SURF, KNN	Supervised vs. Unsupervised Accuracy	92% (supervised), 78% (unsupervised)
Pathak et al. [8]	Custom Gestures	OpenCV, MediaPipe	Rule-based on hand landmark logic	Gesture Recognition Accuracy	Real-time, cost-effective
Kowsigan et al. [9]	Custom Dataset	CNN, Gaussian Filter	CNN + NLP + Speech Recognition	Overall accuracy	>85%, bi-directional (speech <-> sign)
Deshmukh et al. [10]	Custom Speech & ISL Animations	Blender, Flask, Web Speech API	Keyword Extraction + NLP	Keyword mapping accuracy	Real-time ISL, animated, ReactJS interface
Shahriar et al. [11]	Custom Bangla Audio + Sign Image DB	CMU Sphinx, Google TTS, Avro/Bijoy Bangla	MFCC, Euclidean Distance, HMM	Accuracy in varied environments	84.71% avg. accuracy; 92.5% (home), 72.22% (noisy)
Aasofwala et al. [12]	Not available	NLP Tools, Avatar Generator, HamNoSys, SiGML	CNN for sign recognition, NLP for structure	Conceptual model	Proposed system with avatar display; focuses on Gujarati speech-to-sign model

IV. REFERENCES

- [1] N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden, "Sign Language Transformers: Joint End-to-end Sign Language Recognition and Translation,"
- [2] B. Saunders, N. C. Camgoz, and R. Bowden, "Progressive Transformers for End-to- End Sign Language Production,"
- [3] A. Kumar, P. Maheshwari, and S. Raj, "Real-time conversion of sign language to speech," *Int. J. Sci. Res. Eng. Dev.*, vol. 2, no. 2, pp. 278–283, Mar. 2019.
- [4] <https://www.sciencedirect.com/science/article/pii/S2772941924000942>
- [5] Parul Kapoor, Rudrabha Mukhopadhyay, Sindhu B. Hegde, Vinay Namboodiri, and C. V. Jawahar, "Towards Automatic Speech to Sign Language Generation," *arXiv preprint arXiv:2106.12790*, Jun. 2021.
- [6] R. K. Reddy, K. R. R. Reddy, R. R. Reddy, and K. R. Reddy, "Translating Speech to Indian Sign Language Using Natural Language Processing," *Future Internet*, vol. 14, no. 9, Art. no. 253, Sep. 2022.
- [7] V. A. Adewale and A. O. Olamiti, "Conversion of Sign Language to Text and Speech Using Machine Learning Techniques," *Journal of Research and Review in Science*, vol. 5, no. 1, pp. 58–65, Dec. 2018.
- [8] A. Pathak, A. Jadhav, N. Patil, S. Rukhande, P. Padhy, and L. Gadhikar, "Real-Time Sign Language to Speech Converter Using OpenCV and MediaPipe," 2025 International Conference on Emerging Systems and Intelligent Computing (ESIC), 2025.
- [9] M. Kowsigan, R. Dhawan, and A. Kundu, "An Efficient Speech to Sign Language Conversion and Text Recognition through Live Gesture," 2024 International Conference on Smart Power Control and Renewable Energy (ICSPCRE), 2024.
- [10] A. Deshmukh, A. Machindar, S. Lale and P. Kasambe, "Enhancing Communication for the Hearing Impaired: A Real-Time Speech to Sign Language Converter," 2024 27th International Symposium on Wireless Personal Multimedia Communications (WPMC), 2024.
- [11] N. Aasofwala, S. Verma and K. Patel, "A Novel Speech to Sign Communication Model for Gujarati Language," *2021 3rd International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2021, pp. 20–25, doi: 0.1109/ICIRCA51532.2021.9544635.
- [12] R. Shahriar, A. G. M. Zaman, T. Ahmed, S. M. Khan, and H. M. Maruf, "A Communication Platform Between Bangla and Sign Language," *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Dhaka, Bangladesh, 2017, pp. 1–4, doi: 10.1109/R10-HT.