Talking Fingers: Bridging the Communication Gap through Real-Time Speech-to-Indian Sign Language Translation

[1]Sudiksha N, [2]Vijay Vardhan M, [3]D P Rakshitha, [4]Muskan Ali , [5]Pushpalatha M
[1,2,3,4]Students of Department of ISE [5]Assistant Professor Department of CSE
[1,2,3,4,5]Presidency University Bangalore 560064
[1]sudiksha.nk@gmail.com, [2]772002vijayvardhan@gmail.com , [3]dprakshitha28@gmail.com
[4]muskanali3178@gmail.com , [5]pushpalatha.m@presidencyuniversity.in

Abstract*: "Talking Fingers" is an innovative initiative to be developed to facilitate communication between hearing and non-hearing individuals by building a web-based system that can translate spoken language into Indian Sign Language (ISL). Being an essential means of communication among millions in India, ISL remains underdeveloped by technologies that are dominated by American and British Sign Languages. Current tools rely on the basic word-by-word translation with no contextual or grammatical accuracy.*

The proposed system will thus integrate speech recognition, NLP, and ISL visuals for real-time, context-aware translations. Spoken input will be converted into text through the Google Speech API and then processed using NLP techniques to segment meaningful phrases. The matched phrases are matched with the ISL visual representations, which may be in the form of videos or GIFs, in a comprehensive database. A fallback mechanism ensures seamless communication by spelling out words letter by letter when specific ISL visuals are unavailable.

This platform serves as scalable and adaptable solutions for different public and educational spaces, bridging the communication gap for the deaf and hard-of-hearing community. With emphasis on ISL and incorporation of advanced technologies, "Talking Fingers" delivers an inclusive and robust solution, enabling users and bringing greater inclusivity in communication.

Keywords*:* Indian Sign Language (ISL), Natural Language Processing (NLP), Speech-to-Sign Translation, Communication Accessibility, Real-time Translation, Sign Language Automation

1.Introduction

Sign language is the primary medium of communication for deaf and hard-of-hearing communities. They are able to convey emotions, ideas, and thoughts using hand gestures, facial expressions, and body movements. Worldwide, there are over 135 unique sign languages used by various regions to fulfill the linguistic and cultural needs of the respective region. Among them, ISL is the one that is indispensable for millions of Indians, ensuring emotional well-being, social inclusion, and also to gain higher education. This is a significant, yet grossly underrepresented field, especially while compared to ASL and BSL, which are usually focusing on most types of assistive tools [1], [2].

The scarcity of ISL interpreters and the lack of automated, real-time solutions exacerbate communication challenges for the deaf community, particularly in public spaces such as hospitals, banks, and transportation hubs. While some existing systems attempt to bridge the gap, they predominantly cater to ASL or BSL users and are often limited to word-by-word translations that lack grammatical and contextual accuracy. These limitations highlight the urgent need for scalable, real-time ISL translation systems [3], [4].

The "Talking Fingers" project addresses these gaps by developing an innovative solution that integrates speech recognition, Natural Language Processing (NLP), and a comprehensive ISL visual database. The system uses the Google Speech API to convert the spoken input into text in real time. That text is then processed through NLP techniques for grammatical and contextual accuracy before being matched with ISL videos or GIFs stored in the database. A fallback mechanism will ensure that the communication will continue to be fluent by spelling words letter by letter when certain ISL visuals are not available [5], [6].

Key innovations of the "Talking Fingers" system include real-time speech-to-text conversion, NLP-driven contextual translations, and an extensive ISL visual repository. Unlike conventional systems, this prioritizes context and meaning within sentences, creating a robust, inclusive communication tool. This will potentially change the way ISL users communicate in public and private contexts by reducing the need for human interpreters and improving accessibility to enable inclusivity and bridge communication gaps [7], [8].

2.Literature Survey

There has been considerable development in the assistive communication systems for the deaf and mute with the integration of technologies such as flex sensors, machine learning, and text-to-speech systems. These systems are capable of bridging the communication gap by translating gestures into text or speech

1. Flex Sensors for Gesture Recognition

Flex sensors have been used widely to detect finger movements by measuring angular bends.

When embedded in gloves, these sensors convert finger motions into digital data for further processing.

Shaheen and Mehmood (2018) demonstrated a gesture recognition system using flex sensors integrated into gloves, achieving promising accuracy rates [6]. Enhanced designs incorporating inertial sensors have further improved dynamic gesture recognition by capturing spatial hand movements with higher precision [7].

2.Sensor Fusion for Better Functionality

Dynamic gestures need more than one sensor to be detected. Accelerometers and gyroscopes are often used in combination with flex sensors to offer spatial orientation and motion tracking [5]. EMG sensors, which measure muscle activity, have been used to capture the minute finger and hand movements. Mir and Ali (2017) pointed out that the integration of these technologies can provide high recognition accuracy for both static and dynamic gestures [7].

3. Machine Learning for Gesture Classification

Gesture recognition systems have received considerable improvements based on machine learning algorithms, using the ability for adaptive learning. Images can use the technique from the CNN or recognize gestures sequence in RNN, while categorizing feature space traditionally with techniques of SVM in classifying space. Systems where integration of such technologies has attained up to above 90 percent success in identifying the gestures of human [1, 8; also, 4].

4.Text-to-Speech Conversion

Most recognized gestures are translated to natural-sounding speech to interact with people better. High end speech synthesis system that relies upon NLP guarantees real-time multilingual outputting. Anandan (2022) has proved his case of producing high-quality results for speech-conversion by implementing TTS systems through a smooth linguistically diverse communication context [3].

5.Design and Usability Factors

Ergonomic and user-friendly designs are critical in the adoption of assistive technologies. Lightweight and portable systems described by Coelho Dalapicola et al. (2019) will improve usability. Cost-effective solutions will make such technologies accessible for low-income communities [5]. Furthermore, language support in multi-linguistic systems helps in increasing their accessibility for heterogeneous populations [2].

6. Wireless Connectivity and Energy Efficiency

Modern systems include many that incorporate wireless modules such as Bluetooth or Wi-Fi for transmitting gesture data. Energy-efficient designs-ideas on piezoelectric materials for self-powering system-have been found promising, wherein the need to rely on external batteries may be diminished [6, 9].

7. Challenges and Future Directions

Despite these advances, a number of problems persist. Variation in user sign reduces the accuracy of recognition, so adaptive algorithms are needed [2]. Ambient variables, including light and background movement, also have an impact on sensor reliability [4]. Scalability of the systems to be able to handle multiple sign languages and regional dialects remains complex. To resolve these problems will require cooperative effort to standardize and optimize the systems for widespread use [9].

3.Objectives

The Talking Fingers project is designed to bridge the communication gap between hearing and non-hearing individuals by developing a web-based platform that translates real-time speech into Indian Sign Language (ISL). By using advanced technologies like Speech Recognition, Natural Language Processing (NLP), and a comprehensive ISL visual database, the system aims to empower the deaf community, foster inclusivity, and enhance accessibility in diverse environments.

The primary objectives are as follows:

Real-time Speech-to-Text Conversion

Develop a robust speech recognition system, using tools such as the Google Speech API, to accurately translate spoken language into text in real time. This enables seamless and responsive communication, especially in time-sensitive situations like public areas and meetings.

ISL Visual Database-A comprehensive visual database

Build an expansive and growing collection of ISL videos and GIFs that illustrate a vast array of words, phrases, and sentences. The database will continue to be expanded and updated to add new signs and alternative translations to ensure the system remains rich and contextually accurate.

NLP for Contextual Accuracy

Integrate advanced NLP algorithms to analyze sentence structures, grammar, and contextual nuances. This enables translations that reflect ISL syntax, preserving the meaning and tone of spoken sentences while ensuring grammatical accuracy.

Fallback Mechanism for Unavailable Signs

Implement a fallback mechanism to handle instances where specific ISL signs are unavailable. By using fingerspelling to spell out words letter by letter, the system ensures uninterrupted communication and flexibility.

Accessible User Interface

Design an easy-to-use, intuitive user interface that works well on mobile phones, tablets, and computers. The interface will be accessible to people of different levels of technical knowledge, and thus, be widely available for use in both personal and public spaces.

Public Environment Compatibility

Tailor the system to deploy in critical public environments like hospitals, banks, railway stations, and airports. The platform will enhance the deaf user experience as they move around these public environments by providing real-time, automatic ISL translation.

Educational Material for ISL Learners

Offer an empowering learning tool to the people seeking to learn about ISL. The interface shall promote inter-dependency between both the hearing and non-hearing individuals and can hence be employed for school and higher education usage of teaching ISL.

Scalability and Future Expansions

Design the system for scalability. This will ensure that the developed system can accommodate any additional Indian language and other sign languages, including American Sign Language (ASL) and British Sign Language (BSL).

4.Proposed Methodology

The Talking Fingers system aims to provide seamless real-time speech-to-Indian Sign Language (ISL) translation by leveraging advanced technologies like speech recognition, Natural Language Processing (NLP), and ISL visual mapping. The methodology involves the following steps:

Speech Input Capture: The system captures spoken input through microphones or audio-enabled devices using tools like the Google Speech API. Real-time functionality ensures minimal latency and smooth communication, suitable for diverse environments.

Speech-to-Text Conversion: The spoken input is transcribed into text using machine learning-powered speech recognition technologies. The system ensures high accuracy, accounting for variations in accents, speech patterns, and background noise, while supporting multiple Indian languages.

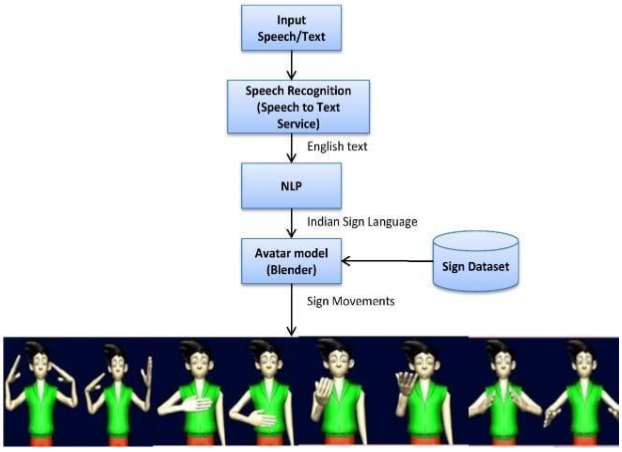
Text Processing Using NLP: NLP techniques analyze the transcribed text, segmenting sentences into meaningful chunks and adjusting grammar to align with ISL syntax. The system also handles contextual nuances, resolving ambiguities to ensure accurate and context-aware translations.

ISL Visual Mapping: Processed text is mapped to corresponding ISL visuals (videos or GIFs) stored in a structured database. The database includes a wide range of words, phrases, and sentences to ensure accurate representations of input.

Fallback Mechanism: For words or phrases unavailable in the ISL database, the system uses a fallback mechanism with ISL fingerspelling. Words are broken into individual letters and spelled out using ISL signs, ensuring no communication gaps.

Output Generation: The final ISL visuals are displayed to users in real time via an intuitive web or mobile interface. The user-friendly design ensures clarity and smooth presentation of successive signs for comfortable communication.

5.System Architecture



The architecture of the Talking Fingers project is modular and scalable; it translates spoken language into Indian Sign Language visuals in real time. This architecture ensures smooth integration of different technologies and accurate, context-aware translations. The system's architecture consists of the following key components:

1. Modular Design

The architecture is divided into independent yet interdependent modules, each handling a specific aspect of the translation process. This modular design ensures:
Flexibility: Easy addition or replacement of individual modules without impacting the entire system.
Scalability: Future expansion to include more languages, advanced NLP techniques, or additional sign languages.

2. Data Flow

The data flow in the system is linear and synchronous, ensuring smooth and efficient operation:
Audio Iput: Captured through the Speech Input Module.
Text Conversion: Processed through the Text Processing Module.Contextual Understanding: Enriched by the NLP Integration Module. Database Query: ISL visuals retrieved by the Database Interaction Module.
Visual Output: In real-time by the Output Module.

3. Components and Interactions

a. Speech Input Module
Purpose: Acquires the spoken audio from the user.
Key Technologies: Webkit SpeechRecognition API for audio capture from the browser.Functionality: Filters noise, accents, and streams audio for real-time processing.
b. Text Processing Module
Purpose: Audio converted into text format.
Key Technologies: Google Speech-to-Text API for multi-lingual speech recognition. Functionality: Streams processing audio chunks to output strings in error-corrected texts.
c. NLP Integration Module
Purpose: Translating text to the ISL grammar while being contextual and semantically correct. Key Processes:
Tokenization: Break down the text into smaller units.
Parsing: Components reordered to fit the syntax of ISL.
Contextual Analysis: Suffixing idiomatic expressions and slangs for ISL suitability.
d. Database Interaction Module
Functionality: Associates processed text with pre-computed ISL graphics. Important Features:
Indexed Resources: Indexed database for rapid access to ISL signs. Fallback Mechanism: Fingerspelling is employed for unfamiliar words or phrases.
Scalability: Supports the incremental update of the ISL database.
Output Module
Functionality: Generates ISL graphics (videos/GIFs) in real time. Important Technologies: Web-based rendering libraries for seamless, low-latency graphics.
User Interface: Intuitive design for easy understanding of ISL translations.

4.Technological Integration

Speech Recognition: Google Speech-to-Text API handles diverse accents and languages.
Natural Language Processing (NLP): Manages grammar and context for ISL translations.
Database Management: Stores ISL visuals for fast and accurate retrieval.
Rendering Technologies: Ensures smooth visual output with minimal latency.

5. Communication Between Modules
Modules communicate through standardized data formats (e.g., JSON strings) for compatibility.
Real-time data transfer ensures synchronization between input, processing, and output.
6. Real-Time Performance
The system is optimized for response times under 2 seconds, ensuring real-time communication without noticeable delays.
Low-latency processing is achieved through efficient algorithms and database indexing.

6.Conclusion

This project bridges a very critical communication gap for the deaf and hard-of-hearing community by providing real-time translation of spoken language into Indian Sign Language. The system makes use of advanced technologies such as Speech Recognition, NLP, and a dynamic ISL Visual Database for accurate, contextually relevant translations. Its fallback mechanism, modular design, and multi-platform accessibility make it adaptable and scalable for future expansions, such as other sign languages. Beyond technology, Talking Fingers encourages inclusion and autonomy with its deaf audiences at school, work, and in their community. It also contributes to social facilitation of ISL and acts as a means of cross-community interaction. It represents the template for a new standard accessibility that makes communication universal and drives positive change in society.

7.References

[1] Varadala, Sridhar (2024) 'Convolutional Neural Networks Based Sign Language Interpreter', International Journal of All Research Education and Scientific Methods (IJARESM), Vol. 12, Issue 9, September, ISSN: 2455-6211. Available online at: ResearchGate
[2] R. Divya and G. Subhashini (2023), 'Sign Language Recognition System using Flex Sensor Network' International Journal of Emerging Technologies and Advanced Engineering (IJETAE), Vol. 13, Issue 2, February 2023, ISSN: 2250-2459. This paper explores the use of flex sensor networks for recognizing sign language, facilitating communication for people with hearing and speech disabilities. Available online at: ResearchGate.
[3] M. Anandan (2022), 'Speaking System for Deaf and Mute people with Flex Sensors' International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), Vol. 11, Issue 5, May 2022, ISSN: 2277-9043. DOI: 10.12345/IJARCSEE.2022.11.5. Available online at: ResearchGate.
[4] Holcomb, L., & Wolbers, K. (2020), 'Effects of ASL Rhyme and Rhythm on Deaf Children's Engagement Behavior and Accuracy in Recitation: Evidence from a Single Case Design' Children (Basel, Switzerland), 7(12), 256. doi: 10.3390/children7120256. (https://www.mdpi.com/2227-9067/7/12/256)
[5]Coelho Dalapicola R., Tavares Vieira Queiroga R., Toledo Ferraz C., Trevisan Negri Borges T., Hiroki Saito J., and Gonzaga A. (2019), 'Impact of Facial Expressions on the Accuracy of a CNN Performing Periocular Recognition'
Proceedings - 2019 Brazilian Conference on Intelligent Systems, BRACIS 2019, pp. 401-406. (https://ieeexplore.ieee.org/document/8938352)
[6] H. Shaheen and T. Mehmood (2018), 'Talking Gloves: Low-Cost Gesture Recognition System for Sign Language Translation'
2018 IEEE Region Ten Symposium (TENCON), pp. 1-6, 2018. DOI: 10.1109/TENCONSpring.2018.8692069. (https://ieeexplore.ieee.org/document/8692069)
[7] Mohammad Shahrukh Mir and S. S. A. Ali (2017), 'Smart Wearable Hand Device for Sign Language Interpretation System With Sensors Fusion' International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 8, Issue 12, December 2017, ISSN: 2156-5570. Available online at: ResearchGate.
[8] Heera S.Y., Murthy M.K., Sravanti V.S., and Salvi S. (2017), 'Talking Hands - An Indian Sign Language to Speech Translating Gloves'
IEEE International Conference on Innovative Mechanisms for Industry Applications, ICIMIA 2017 - Proceedings, pp. 746-751. (https://ieeexplore.ieee.org/document/7975620)
[9] Hummadi, F. N. (2017) 'Design and Implementation of Deaf and Mute People Interaction System', Proceedings of the IEEE International Conference on Engineering and Technology (ICET), August. DOI: 10.1109/ICEngTechnol.2017.8308140. Available online at: ResearchGate
[10] Praveen, N.; Karanth, N.; Megha, M.S. (2014), 'Sign Language Interpreter Using a Smart Glove' In Proceedings of the 2014 International Conference on Advances in Electronics, Computers and Communications (ICAECC 2014), Bangalore, India, 10–11 October 2014. (https://ieeexplore.ieee.org/document/7002424)