**JSS MAHAVIDYAPEETHA**
**JSS Science and Technology University**


# Audio to ISL Conveter

**Mini project final report submitted in partial fulfilment of curriculum prescribed for the Machine Learning(22CS610) course for the award of the degree of**

## Bachelor of Engineering

**In**

## Computer Science and Engineering

**By**


| USN | NAME |
|---|---|
| 01JST22UCS068 | Kundhan A |
| 02JST22UCS128 | Vaibhav v Ron |
| 02JST22UCS138 | Sankalp Nadiger |
| 02JST22UCS105 | Shrihari A S |

***Under the guidance of***
**Prof. Shalini K C**
Assistant Professor
Dept. of CS&E
SJCE, JSS STU, Mysuru


**Department of Computer Science & Engineering**

2024-2025

**JSS MAHAVIDYAPEETHA**
**JSS Science and Technology University**


# CERTIFICATE

This is to certify that the work entitled "**Audio to ISL Converter**" is a Bonafide work carried out by **Kundhan A, Vaibhav v Ron, Sankalp Nadiger, Shrihari A S** in partial fulfilment of the award of the degree of Bachelor of Engineering in Computer Science and Engineering of JSS Science and Technology, Mysuru during the year 2025. It is certified that all corrections/ suggestions indicated during CIE have been incorporated into the report. The mini-project report has been approved as it satisfies the academic requirements in respect of the mini-project work prescribed for the Machine Learning (22CS610) course.

**Course in Charge**

**Prof. Shalini K C**
**Assistant Professor, Dept of CS & E,**
**JSS STU Mysuru**

Place: Mysuru Date: 24 July 2025

# Contents

Chapter 1

# Abstract

Communication is a fundamental human right, yet millions of individuals with hearing impairments face significant barriers in interacting with the hearing population. In India, Indian Sign Language (ISL) serves as a primary mode of communication for many in the deaf community. However, ISL is not widely understood by the general public, leading to a persistent communication gap. To address this, an **Audio to ISL Converter** system is proposed, which aims to bridge spoken language and ISL using modern technologies such as speech recognition, natural language processing (NLP), and sign language visualization.

This system functions by capturing spoken audio through a microphone, converting the audio to text using advanced speech-to-text algorithms, and then translating the resulting text into ISL. The ISL output is rendered either through a 3D animated avatar or a sequence of video-based sign gestures. The core challenge lies in accurately mapping spoken language characterized by grammar, slang, and regional dialects—to ISL, which has its own grammar and structural nuances.

By leveraging artificial intelligence and rule-based language models, the system can effectively parse and simplify spoken sentences into ISL gloss (a simplified textual representation used in sign language translation), which is then visualized for the user.

This converter can be especially beneficial in environments where real-time, inclusive communication is essential, such as educational institutions, hospitals, government offices, and customer service platforms.

In conclusion, the Audio to ISL Converter offers a significant step toward inclusive communication, helping bridge the accessibility gap between the hearing and non-hearing populations. The system not only promotes equality and social integration but also paves the way for further innovations in assistive technology, particularly for Indian linguistic and cultural contexts.

Chapter 2

# Introduction

In today's digital age, communication barriers still exist between the Deaf and Hard of Hearing community and the hearing world — especially in public spaces, classrooms, and digital media. While traditional methods like human interpreters are helpful, they are often limited by cost, availability, and scalability.

To address this challenge, we developed an Audio-to-Indian Sign Language (ISL) Converter. This system automatically translates spoken language into Indian Sign Language gestures in real time, helping bridge the gap between speech and sign. It combines speech recognition, natural language processing (NLP), and real-time 3D avatar animation into a single, web-based application called AurisVue.

The frontend is designed to be clean, modern, and user-friendly, accessible directly through browsers on different devices. Users can choose modes such as real-time translation for public use, interactive classroom mode, browser plugin for video content, or a training mode to practice ISL.

Key features of our system include:

- Accurate speech-to-text conversion using advanced AI models.
- Real-time ISL gesture rendering with a 3D avatar.
- Interactive modules for learning, practice, and live translation.
- Secure, scalable web architecture with user authentication and progress tracking.

By merging AI, web technology, and inclusive design, this project aims to empower Deaf and Hard of Hearing individuals, making spoken information instantly accessible and supporting broader social inclusion.

Chapter 3

# Tools and Technologies Used

1. Speech Recognition
    - Uses advanced models like OpenAI Whisper and Google Speech-to-Text.
    - Converts live or recorded audio into accurate text form.

2. Text Processing & Natural Language Processing (NLP)
    - Simplifies and reorders recognized text into ISL grammar.
    - Uses NLP libraries to tokenize, clean, and structure text.
    - Maps words to ISL gestures via a structured sign dictionary.
    - Falls back to finger spelling for unknown words.

3. 3D Avatar Animation
    - Renders real-time ISL signs using Three.js and GLTFLoader.
    - Uses rigged animations created with Mixamo for smooth gesture playback.

4. Webcam-Based Practice Module
    - Uses MediaPipe to track users' hand gestures through the webcam.
    - Provides automated visual feedback to help users practice and correct their ISL signs.

5. Secure and Scalable Architecture
    - Implements API rate limiting with caching to handle high user loads.
    - Uses JWT-based authentication to keep sessions secure.
    - Supports cloud deployment using Docker, Vercel, or Render for scalability and reliability.

6. Frontend and User Experience
    - Fully web-based interface that runs directly in browsers.
    - Minimal hardware requirements make it widely accessible.
    - Designed to be responsive and user-friendly across devices.

Chapter 4

# Dataset and Preprocessing

1. Audio Data Collection
   - Captures real-time speech input directly from a microphone.
   - Allows users to upload pre-recorded audio files for conversion.
   - Supports multiple file formats (e.g., WAV, MP3).

2. Speech-to-Text Transformation
   - Uses powerful AI models like OpenAI Whisper and Google Speech-to- Text.
   - Accurately transcribes spoken language into text, including punctuation.
   - Handles background noise and varying accents to improve recognition accuracy.

3. Text Normalization & Simplification
   - Cleans raw transcribed text by removing special symbols and numbers.
   - Converts text to lowercase for uniform processing.
   - Tokenizes sentences into individual words for detailed analysis.
   - Uses NLP to reorder words according to ISL grammar rules.

4. Sign Dictionary & Finger Spelling
   - Matches cleaned and reordered words to available ISL signs stored in a dictionary.
   - Uses finger spelling as a backup for words not found in the dictionary.
   - Keeps a separate list of frequently used words to prioritize common gestures.

5. 3D Sign Models
   - Uses GLTF format for lightweight, browser-friendly models.
   - Models are animated using rigging tools like Mixamo for natural movement.
   - Optimizes animations to reduce latency and improve playback on low-end devices.

6. User Input Data
   - Captures webcam-based hand gesture data during practice sessions.

- Records user performance: gesture speed, accuracy, and consistency.
- Stores historical data to track improvement over time.

7. Preprocessing Efficiency
   - Implements caching and serialized storage to avoid repeated preprocessing.
   - Uses Joblib or similar tools to save NLP-processed text and sign mappings.

8. Handling Missing or Complex Data
   - Detects and flags ambiguous words or code-mixed speech (e.g., English- Hindi mix).
   - Uses fallback translation layers to standardize multilingual input.
   - Plans for future dataset expansion by adding crowd-sourced signs and collaborating with ISLRTC.

Chapter 5

# Methodology

1. Audio Capture and Input
    - The system captures speech input live from the user's microphone.
    - Supports uploading of pre-recorded audio files for ofline processing.
    - Accepts multiple file formats (e.g., MP3, WAV) for flexibility.

2. Speech-to-Text Conversion
    - Uses robust models like OpenAI Whisper or Google Speech-to-Text.
    - Converts spoken language into text accurately, even in noisy environments.
    - Preserves important features like punctuation to help with natural text structure.

3. Text Cleaning and Normalization
    - Processes raw text to remove unwanted symbols, numbers, and special characters.
    - Converts all text to lowercase for consistency.
    - Tokenizes text into individual words for detailed NLP processing.
    - Filters out filler words and very short words that don't contribute meaning.

4. Grammar Simplification and Reordering
    - Uses Natural Language Processing (NLP) to rearrange text into ISL-compatible grammar.
    - Removes redundant words and simplifies complex sentence structures.
    - Ensures that the text aligns with the visual and structural needs of ISL signing.

5. Sign Mapping
    - Matches processed words and phrases to corresponding ISL signs stored in a structured dictionary.
    - Uses fallback finger spelling to handle words that aren't directly mapped to signs.
    - Keeps track of frequently used words to optimize real-time performance.

6. 3D Animation Rendering
    - Loads GLTF 3D models and animations rigged via Mixamo.
    - Uses Three.js and GLTFLoader to render ISL gestures in the browser.
    - Ensures animations run smoothly, even on devices with limited graphics power.

7. Practice and Feedback Module
    - Uses MediaPipe to track user hand gestures during practice.
    - Provides automated visual feedback, comparing user gestures with standard ISL signs.
    - Helps users identify mistakes and improve signing technique.

8. User Progress Tracking
    - Records key metrics such as:
    - Gesture accuracy and speed
    - Quiz scores and comprehension test results
    - Practice session frequency and duration
    - Stores data locally or in the cloud to allow users to review and track improvement over time.

9. Integration and Security
    - Connects all modules (speech recognition, NLP, 3D rendering, practice) into a unified web interface.
    - Uses JWT-based authentication to protect user data and sessions.
    - Applies API rate limiting with caching to handle multiple users simultaneously.
    - Deploys using Docker containers on cloud platforms (e.g., Vercel, Render) for scalability.

Chapter 6

# Web Interface

1. Frontend Architecture
    - Built as a web-based application accessible via modern browsers.
    - Uses lightweight design so it can run efficiently on both desktop and mobile devices.
    - Supports cross-platform access without requiring installation.

2. User Registration and Login
    - Allows users to create an account with email and password.
    - Implements JWT-based authentication to protect sessions and secure user data.
    - Ensures that only logged-in users can access learning and practice modules.

3. Homepage and Navigation
    - Displays project introduction and quick links to modules like “Learn”, “Practice”, and “Translate”.
    - Provides an intuitive navigation bar that updates based on login status (e.g., shows “Profile” or “Logout” after login).

4. Live Translation Module
    - Captures real-time speech input and instantly displays corresponding ISL signs.
    - Shows animated 3D avatar performing the recognized signs.
    - Offers fallback finger spelling when no sign is available.

5. Learn Mode
    - Divides ISL learning into categorized lessons (e.g., greetings, numbers, daily objects).
    - Includes step-by-step tutorials with visual cues and avatar

demonstrations.

- Designed to help beginners build vocabulary gradually.

6. Practice Module (Webcam-Based)
    - Uses MediaPipe to track user hand gestures through a webcam.
    - Compares user gestures with standard ISL signs and provides:
    - Real-time feedback on accuracy and movement
    - Suggestions to improve hand shapes or motion
    - Helps users self-correct and gain confidence.

7. Virtual Assistant
    - Integrated chatbot that answers user questions about:
    - ISL grammar and sign meanings
    - How to use different modules
    - General help and troubleshooting
    - Improves user engagement and accessibility.

8. User Progress Tracking Dashboard
    - Displays historical data like:
    - Quiz scores and comprehension levels
    - Practice session counts and duration
    - Gesture recognition speed and accuracy trends
    - Encourages users by showing visual progress over time.

9. Responsive Design and Accessibility
    - Developed to work smoothly across different screen sizes (mobile, tablet, desktop).
    - Follows modern accessibility standards, e.g., clear fonts, contrast, and keyboard navigation support.
    - Makes the tool inclusive for both hearing and hearing-impaired users.

10. Integration of Backend Services
    - Connects speech recognition, NLP, sign mapping, and animation services in a seamless flow.
    - Uses API rate limiting and caching to optimize performance under load.

- Designed to be modular, allowing easy updates or future integration with real-time ISL datasets.

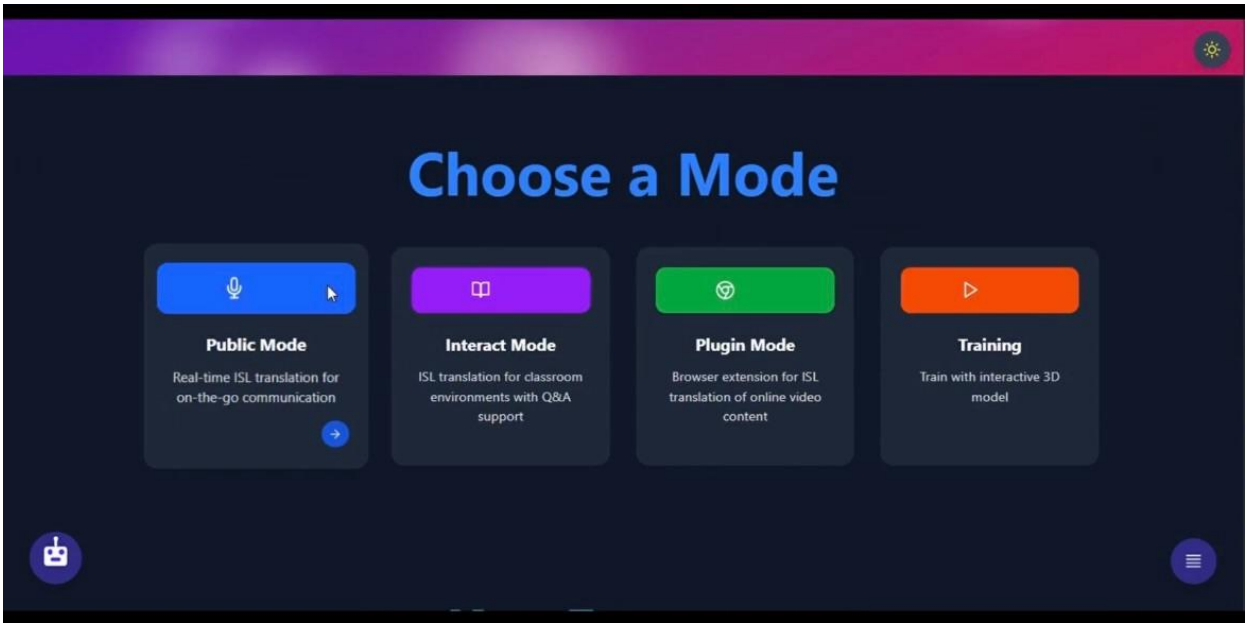
11. Cloud-Based Deployment
    - Deployed using Docker containers on platforms like Vercel and Render.
    - Enables scalability to serve more users simultaneously.
    - Reduces latency by hosting close to users geographically.

Thus the web interface is the heart of the Audio-to-Indian Sign Language (ISL) Converter system.
While powerful AI models and NLP algorithms handle speech recognition and sign translation in the background, it is the web interface that brings all these technologies to life for the user.

Chapter 7

# Results and Observations

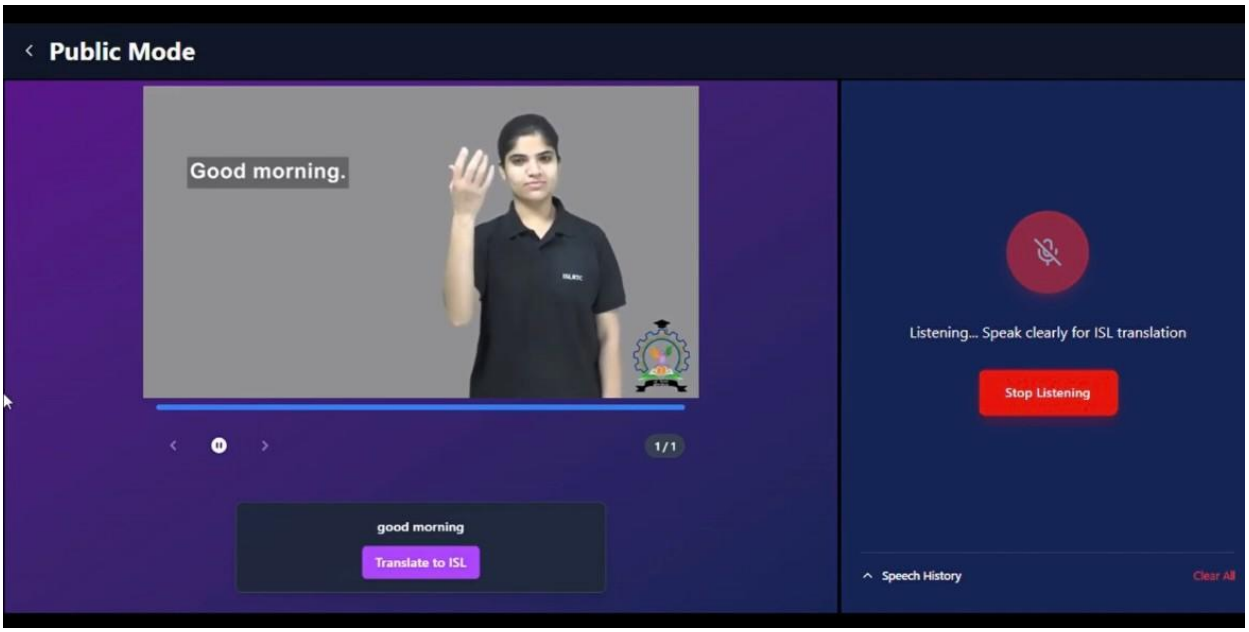

This fig 1.1 shows the welcome and introduction screen of the AurisVue system. It displays the title “Create Amazing Digital Experiences” and briefly explains the project’s purpose: an inclusive Audio-to-Indian Sign Language converter designed for the Deaf and Hard of Hearing community. The background uses calming purple and pink gradients, and on the right, a logo with signing hands symbolizes accessibility and digital communication.



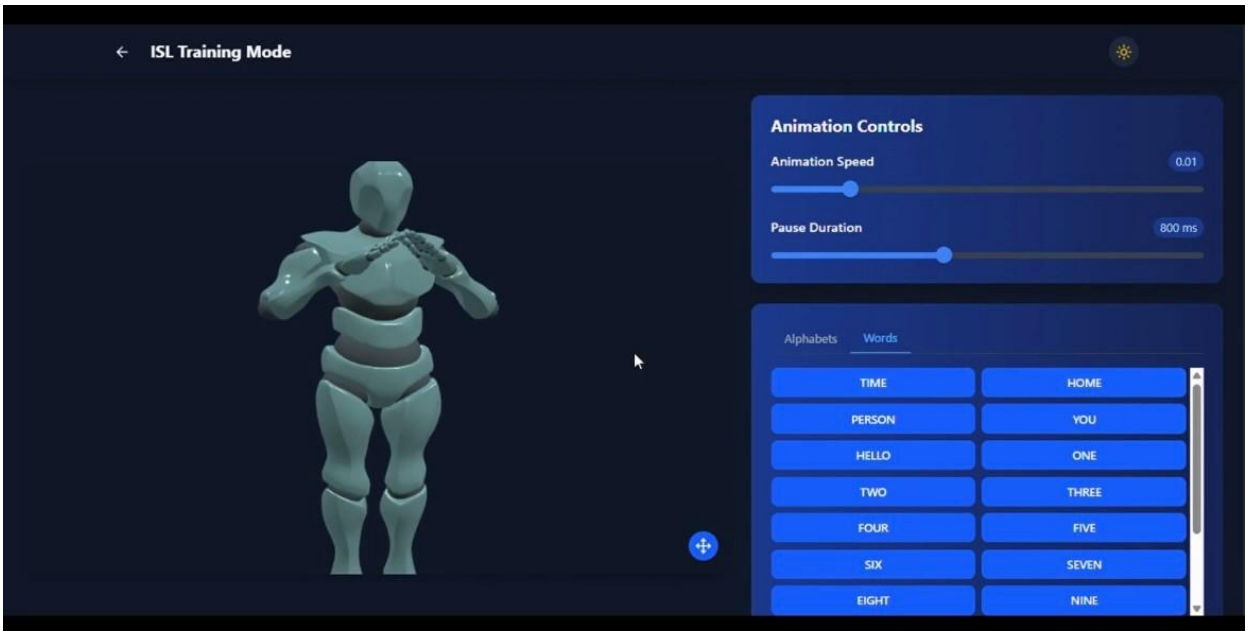

This fig 1.2 presents the “Choose a Mode” interface where users select different ways to use the system:

- Public Mode: Real-time ISL translation for live conversations.
- Interact Mode: Classroom translation with question–answer support.
- Plugin Mode: Browser extension to translate online video content into ISL
- Training Mode: Practice signs using an interactive 3D avatar.

The design is modern and uses bright colors to guide users easily.




This fig 1.3 illustrates the Public Mode in use. The screen shows a presenter signing “Good morning” in Indian Sign Language (ISL). On the right, a voice input panel actively listens for user speech. The system transcribes spoken text and then shows the matching ISL sign, supporting real-time and natural communication.




This fig 1.4 shows the ISL Training Mode with a 3D animated avatar demonstrating sign language gestures. On the right side, users can control animation speed, pause duration, and select from a list of words or alphabets to practice. This mode helps users learn by visually observing and repeating the correct ISL movements interactively.

Chapter 8

# Conclusion and Future Work

## Conclusion

The Audio to Indian Sign Language (ISL) Converter is a promising step toward bridging the communication gap between hearing and hearing-impaired individuals. By converting spoken language into ISL gestures, it promotes inclusive communication in essential sectors like education, healthcare, and public services. While challenges remain in accuracy and language complexity, continued advancements in AI and sign language research will help refine and expand this technology for real-world use.

## Future Work

The current version can be improved in the following ways:

- **Expansion to Multiple Indian Languages**: Incorporate regional languages (Hindi, Tamil, Bengali, etc.) in addition to English, making the system more inclusive across India's linguistically diverse population.
- **Reverse Translation**: Develop a system to convert ISL gestures back to spoken language or text, enabling two-way communication between deaf and hearing individuals.
- **Facial Expression and Emotion Recognition**: Enhance the avatar's realism by including facial expressions, which are a vital part of sign language communication.
- **Real-Time Performance Optimization**: Improve processing speed and reduce latency to ensure seamless live interactions.
- **Personalization and User Adaptability**: Allow users to customize gesture speed, avatar appearance, or language preferences for better accessibility.
- **Integration with Smart Devices**: Adapt the system for mobile phones, tablets, kiosks, and wearable devices for wider usability in different environments.
- **Dataset Enrichment**: Create or contribute to an open-source ISL dataset that includes a wide range of vocabulary, regional variations, and real-life sentence structures.