SRES SANJIVANI COLLEGE OF ENGINEERING

(An Autonomous Institute)

DEPARTMENT OF COMPUTER ENGINEERING

Final Year B.Tech Project Synopsis

Academic Year 2025-202

**Group ID: 28**

**Project Title: “Voice Bridge”**

1. **Project Definition**

The Voice Bridge project aims to develop an AI-powered communication application that bridges the gap between deaf and mute individuals and the hearing community. The system will utilize speech-to-text (STT), text-to-speech (TTS), and gesture recognition technologies to facilitate two-way communication. When a deaf or mute user performs gestures or signs, the application will capture and process these actions through computer vision and machine learning models, converting them into readable text on a mobile device. Conversely, when the other person inputs text, the application will translate it into sign animations, GIFs, or emojis, enabling deaf or mute individuals to understand and respond easily. This project primarily focuses on Machine Learning and Data Science techniques to process speech, text, and gestures, making the application accurate, efficient, and user-friendly. The ultimate goal is to build an accessible, inclusive platform that simplifies daily communication for people with hearing and speech disabilities.

# Scope

# Develop a mobile-based AI application for real-time communication between deaf-mute individuals and the hearing community.

# Implement speech-to-text (STT) and text-to-speech (TTS) functionalities for smooth voice and text interaction.

# Integrate gesture/sign recognition using computer vision to convert signs or actions into text.

* + Enable text-to-sign/GIF/emoji conversion to help deaf-mute individuals understand responses visually.
  + Ensure the application is lightweight, user-friendly, and accessible to people of all age groups.

# Objectives

# To design and develop an AI-powered application for effective two-way communication.

# To implement speech recognition models for accurate speech-to-text conversion.

# To build a gesture recognition system to convert signs into text using machine learning algorithms.

# To create a text-to-sign module that converts typed responses into signs, GIFs, or emojis.

# To ensure real-time processing for faster and more natural interactions.

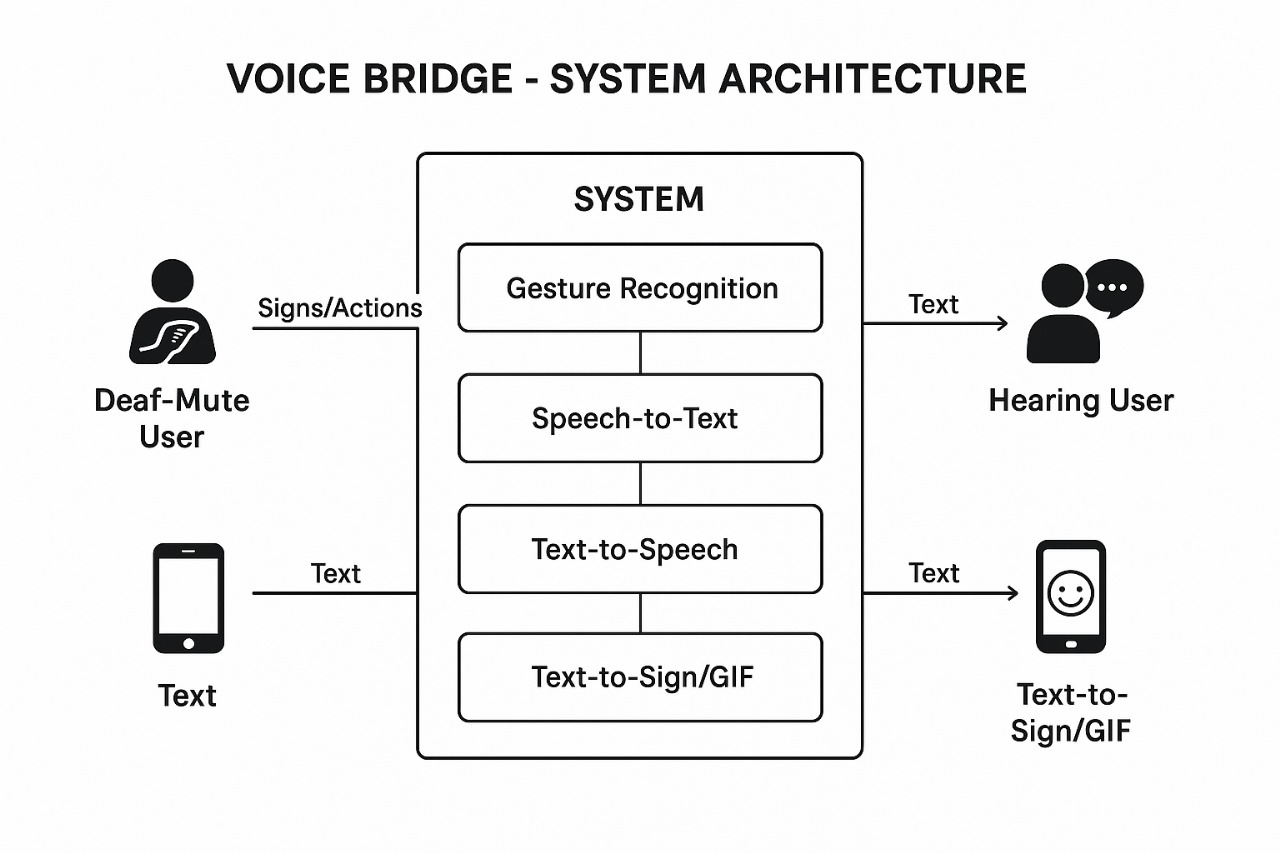
# To improve accuracy and efficiency of gesture and speech recognition through continuous model training.

# To contribute towards digital accessibility and inclusivity for individuals with hearing or speech impairments.

# Review of Conference/Journal Papers and Relevant Theory

* 1. **“A Survey of Deep Learning-Based Sign Language Recognition” (2024)**  
     Reviews CNN, Random Forest, and deep neural models for accurate SLR, achieving up to **94.3% accuracy.**
  2. **“Deep Learning for Sign Language Recognition: A Comparative Review” (2024)**  
     Analyzes **140+ studies,** highlighting CNN, LSTM, and Transformer models and real-time deployment challenges**.**
  3. **“Sign Language to Text and Speech Conversion” (2025)** Use**s CNN and MediaPipe** for gesture detection, converting signs into **text and speech** in real time
  4. **“Real-Time Sign Language Gestures to Speech Transcription” (2025)**  
     Demonstrates **webcam-based CNN systems** for instant and accurate **gesture-to-speech conversion.**

# System Architecture



*Fig 5.1:-System Architecture*

# System Requirements

**6.1 Hardware Requirements:**

**For Development:**

* Processor: Intel i5 or higher / AMD equivalent
* RAM: Minimum 8 GB (16 GB recommended for ML model training)
* Storage: 256 GB SSD or higher
* GPU: NVIDIA GPU (e.g., GTX 1650 or higher) for model training and testing
* Camera: High-resolution webcam (for testing gesture recognition)

**For End Users (Mobile App):**

* Device: Android (8.0 and above) / iOS (13 and above) smartphone
* RAM: 3 GB or higher
* Storage: At least 1 GB free space
* Functional Camera and Microphone

**6.2 Software Requirements:**

**Operating System:** Windows 10/11, macOS, or Android/iOS for the mobile app

**Programming Languages:** Python, Java/Kotlin (for mobile), JavaScript (for frontend if web-based)

**Frameworks & Libraries:**

* TensorFlow / PyTorch – for machine learning models
* OpenCV – for gesture recognition and image processing
* Speech Recognition, Google API – for STT and TTS modules
* Flask/Django – for backend API development
* React Native / Flutter – for cross-platform mobile app development
* Database: MySQL / MongoDB for data storage

**IDE/Tools:**

* PyCharm / VS Code for coding
* Jupyter Notebook for model training and testing
* Android Studio (for mobile application build)
* Version Control: Git & GitHub

#### **The proposed system has the following requirements:**

The proposed system will use machine learning and data science techniques to enable real-time communication between deaf-mute individuals and others. It will integrate:

1. Speech-to-Text (STT) for converting spoken words into readable text.
2. Text-to-Speech (TTS) for converting written text into speech.
3. Gesture and Sign Recognition to interpret signs or actions and display them as text.
4. Text-to-Sign/GIF Conversion to visually represent typed responses as signs, emojis, or GIFs.
5. Real-time processing to ensure smooth and efficient communication.
6. User-friendly interface to make the application accessible for all age groups.
7. Scalability to support multiple languages and additional accessibility features in the future.

# Conclusion

# The Voice Bridge project is an innovative step toward enhancing communication for deaf and mute individuals using AI and Machine Learning technologies. By integrating speech-to-text, text-to-speech, and gesture recognition, the system creates a two-way communication bridge that is both efficient and user-friendly. It not only promotes inclusivity and accessibility but also demonstrates how modern technology can transform lives. With continuous improvements and scalability, Voice Bridge has the potential to become a reliable and impactful solution for real-world communication challenges faced by the specially-abled community.

# Reference

1. https://docs.pytorch.org/audio/stable/tutorials/speech\_recognition\_pipeline\_tutorial.html?utm\_source=chatgpt.com
2. https://medium.com/%40tharindumadhusanka99/real-time-iot-dashboard-with-next-js-esp32-firebase-and-aws-1b4e0db452b6Database Programming with JDBC and Java by O'Reilly
3. https://medium.com/@tharindumadhusanka99/real-time-iot-dashboard-with-next-js-esp32-firebase-and-aws-1b4e0db452b6
4. https://divjoy.com/guide/next-firebase-stripe?utm\_source=chatgpt.comhttp[s://www.w3schools.com/django/django\_intro.php](http://www.w3schools.com/django/django_intro.php)
5. https://www.geeksforgeeks.org/nlp/wav2vec2-self-a-supervised-learning-technique-for-speech-representations/?utm\_source=chatgpt.com

**Group Members**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | **Roll No** | **Student Name** | **Signature** |
| 1 | 2 | Unde Ashwini Somnath |  |
| 2 | 20 | Bhoge Nikita Kalyan |  |
| 3 | 33 | Deshmukh Ishwari Chintaman |  |
| 4 | 37 | Dethe Aaditi Ravindra |  |

|  |  |  |
| --- | --- | --- |
| Prof. P.B. Dhanwate | Dr. S. R. Deshmukh | Dr. M. A. Jawale |
| (Project Guide) | (Project Coordinator) | (HOD, Computer Engg Dept.) |