A PBL-3 Project Report on

**ISL CONNECT**

Submitted to Manipal University Jaipur

Towards the partial fulfillment for the Award of the Degree of

**BACHELOR OF TECHNOLOGY**

In Computers Science and Engineering (AIML)

*Aanya Mittal 229310435*

*Manasvi Garg 229310259*

*Shubhdeep Singh 2293*



(Height:0.9”, width: 4.56”)

Under the guidance of

**Dr. PREETI NAROOKA**

Name of the Supervisor:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature of Supervisor (12pt.)

**Department of Artificial Intelligence and Machine Learning School of Computer Science and Engineering Manipal University Jaipur , Jaipur, Rajasthan**

**CERTIFICATE**

This is to certify that the project report (PBL-3) entitled **ISL CONNECT** submitted by **AANYA MITTAL (229310435)** ,Department of Artificial Intelligence and Machine Learning (AIML), School of Computer Science and Engineering Manipal University Jaipur, Rajasthan for the award of the degree of *Bachelor of Technology* is a record of bonafide work carried out by him/her under my supervision, as per the code of academic and research ethics of Manipal University Jaipur, Rajasthan.

The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university. The report fulfills the requirements and regulations of the University and in my opinion, meets the necessary standards for submission.

Place: Manipal University Jaipur Signature

Date: 26/11/2024 **Dr.Preeti Narooka**

**CERTIFICATE**

This is to certify that the project report (PBL-3) entitled **ISL CONNECT** submitted by **MANASVI GARG (229)** ,Department of Artificial Intelligence and Machine Learning (AIML), School of Computer Science and Engineering Manipal University Jaipur, Rajasthan for the award of the degree of *Bachelor of Technology* is a record of bonafide work carried out by him/her under my supervision, as per the code of academic and research ethics of Manipal University Jaipur, Rajasthan.

The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university. The report fulfills the requirements and regulations of the University and in my opinion, meets the necessary standards for submission.

Place: Manipal University Jaipur Signature

Date: 26/11/2024 **Dr.Preeti Narooka**

**CERTIFICATE**

This is to certify that the project report (PBL-3) entitled **ISL CONNECT** submitted by **SHUBHDEEP SINGH (2293)** ,Department of Artificial Intelligence and Machine Learning (AIML), School of Computer Science and Engineering Manipal University Jaipur, Rajasthan for the award of the degree of *Bachelor of Technology* is a record of bonafide work carried out by him/her under my supervision, as per the code of academic and research ethics of Manipal University Jaipur, Rajasthan.

The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university. The report fulfills the requirements and regulations of the University and in my opinion, meets the necessary standards for submission.

Place: Manipal University Jaipur Signature

Date: 26/11/2024 **Dr.Preeti Narooka**

**ABSTRACT**

Traditional communication platforms often lack inclusivity for deaf, blind, and mute individuals, creating a significant gap in their ability to access basic information and communicate and express ideas fully. Our model addresses this challenge by utilizing existing tools such as speech recognition, computer vision, natural language processing (NLP), and integrating Generative AI (GenAI) to enable seamless communication and response generation in a form that can be easily processed by the Deaf and blind community. Inputs whether in text, speech, or Indian Sign Language (ISL) gestures are analyzed, converted into a unified text format, and transformed into the desired output format. GenAI enhances communication quality by generating accurate and adaptive responses, providing contextual understanding and personalized outputs. This platform aims to bridge the communication gap, fostering inclusivity and enabling broader participation in research and everyday interactions for individuals with sensory and communication disabilities.

**TABLE OF CONTENTS**

1. CERTIFICATE
2. ABSTRACT
3. ABBREVIATIONS
4. LIST OF FIGURES
5. INTRODUCTION
6. OBJECTIVE
7. LITERATURE SURVEY
8. SOFTWARE REQUIREMENT / REQUIREMENT ENGINEERING
9. CODING
10. BIBLOGRAPGHY

* **Abbreviations**
* NLP: Natural Language Processing enables machines to process and analyze human language.
* NLTK: Natural Language Toolkit is a Python library for language data analysis.
* SpeechRecognition: A Python library for converting spoken words into text.
* YOLO: You Only Look Once is a real-time object detection algorithm.
* Ultralytics: A framework for advanced AI models like YOLOv5 and YOLOv8.
* CNN: Convolutional Neural Networks are used for image and video processing tasks.
* ML: Machine Learning allows systems to learn and improve from data.
* DL: Deep Learning leverages neural networks for complex pattern recognition.
* ChatGPT API: An API for integrating OpenAI's ChatGPT conversational AI into applications.
* **List of Figures**

1. Figure 1 - Workflow Diagram
2. Figure 2 – Use Case Diagram
3. Figure 3 – Audio Input
4. Figure 4 – Visual Input
5. Figure 5 – Text Output for visual or audio input

* INTRODUCTION

1. ISL Connect is developed to facilitate communication of help for ease of education and learning for hearing / visually impaired citizens.
2. It can also be developed in a way so that it can be integrated to our existing public spaces such as railway stations where hard hearing/ visually impaired individuals can interact easily with the staff.
3. To provide more inclusivity providing equal access to all types of services for disabled people and help them avail the use of growing technology as by the other people .

* OBJECTIVE

1. People with disability who are restraint from using regular technology : Frequent use of Audio to Visual or Visual to audio module for real time translations .
2. For Example : Railway staff(Or any other staff working at a public space) : Frequent use to communicate with primary users and understand their needs.
3. **Real-Time Interaction**: The invention incorporates real-time gesture recognition and Generative AI capabilities, enabling fluid communication for users who rely on ISL. This contrasts with existing technologies that may offer delayed responses or lack interactivity

* **LITERATURE SURVEY**

The invention addresses the significant communication barriers faced by deaf, blind, and mute individuals. Traditional communication tools often fail to accommodate the diverse needs of these users, limiting their ability to interact effectively in various social and professional contexts. By developing a multi-modal communication platform that integrates text, speech, and Indian Sign Language (ISL), this invention aims to enhance accessibility and facilitate real-time interaction, thus empowering individuals with disabilities.

* **Sign Language Recognition Application Systems**: The article "Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output" (ResearchGate, 2018) provides an overview of various sign language recognition systems. However, most of these systems primarily focus on English sign language and do not incorporate other forms of communication, such as speech or text. This limits their applicability for users who may not be proficient in English or who use multiple languages or modes of communication.[1]
* **Gesture Recognition and Communication:** The paper "A Survey on Gesture Recognition Techniques" highlights advancements in gesture recognition technology. While it discusses various methodologies, it lacks a comprehensive approach that integrates gesture recognition with other communication modalities. This results in a fragmented user experience, as users often have to switch between different systems rather than having a unified platform.[2]
* **IoT in Communication Systems**: The IEEE paper "IoT-Based Communication Systems for the Deaf and Hard of Hearing" explores the use of IoT technologies to facilitate communication. However, it primarily focuses on English and does not provide a multi-lingual or multi-modal approach. Additionally, it does not address the needs of blind and mute individuals, which is a significant gap in the current research landscape.[3]
* **Limitations of Existing Technologies**: The MIT paper "A Framework for Human-Computer Interaction" discusses the integration of AI in enhancing communication tools. However, it does not explore the specific needs of users with multiple disabilities or the potential for Generative AI to facilitate real-time, context-aware communication across various input and output methods.[4]
* PLANNING OF WORK

**4.1** **SOFTWARE REUQUIREMENT**

# 1. Introduction

**1.1 Document Conventions**

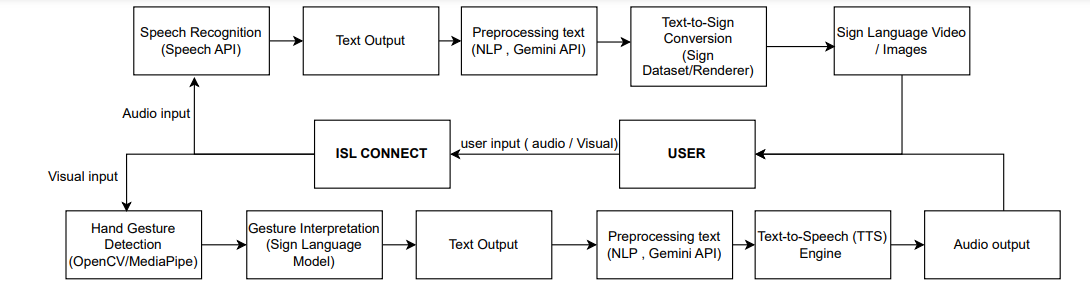
This SRS is following the conventional IEEE format which is specified on their official format available.

**1.2 Product Scope**

ISL Connect is public communication software developed to ease communication for disabled individuals. The main function involves the user convenience like communication based on user preferences that is audio and visual input and output.

# 2. Overall Description

## 2.1 Product Perspective

**The system being developed is a system of its own consisting of the modules stated above.

## 

## 2.2 Product Functions

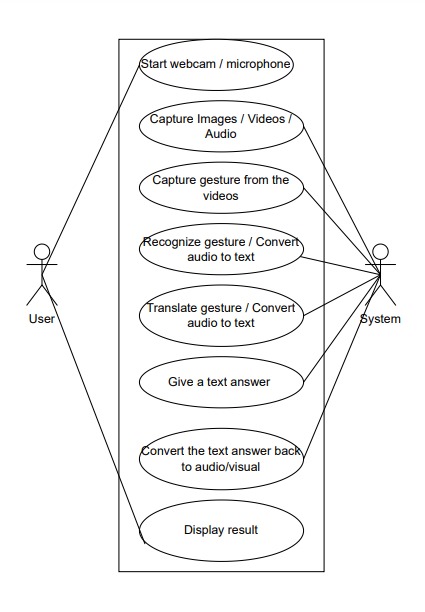
* **Audio to Text**: Converts spoken language into written text.
* **Text to Audio**: Transforms written text into audible speech.
* **Visual to Text**: Recognizes Indian Sign Language gestures and converts them into text.
* **Text to Visual**: Converts text into Indian Sign Language videos or animations.

## 

## 2.3 User Classes and Characteristics

1. Primary Users (People with disability who are restraint from using regular technology) : Frequent use of Audio to Visual or Visual to audio module for real time translations .
2. For Example : Railway staff(Or any other staff working at a public space) : Frequent use to communicate with primary users and understand their needs.

## 2.6 User Documentation



**Fig 2 : Use Case Diagram**

## Assumptions and Dependencies

Third party libraries being used are -

* **gTTS**: For text-to-speech conversions.
* **SpeechRecognition**: For converting audio input to text.
* **OpenCV**: For visual input processing and gesture tracking.
* **MediaPipe / Detectron2**: For hand tracking in sign language gesture recognition.
* **Nltk AND Spacy** : for analyzing and processing text.
* **Python((FLASK/TKINTER)** - interface

# 3. External Interface Requirements

## 3.1 User Interfaces

Input :

* Microphone/Video Icon to take in audio/visual input from the user .
* Options to choose preferred mode of output display

Output : Real time display of processed text in audio / visual form.

## 3.2 Hardware Interfaces

* Camera for image detection
* Microphone for audio input
* Speakers for audio output
* Interactive kiosks for visual output

## 3.3 Software Interfaces

**1. Software Components and Versions**

* **Speech Recognition API**:
  + **Name**: Google Cloud Speech-to-Text
* **Computer Vision Library**:
  + **Name**: OpenCV and Mediapipe
* **Natural Language Processing Library**:
  + **Name**: NLTK (Natural Language Toolkit)

# 4. System Features

## Feature: Audio Input Processing

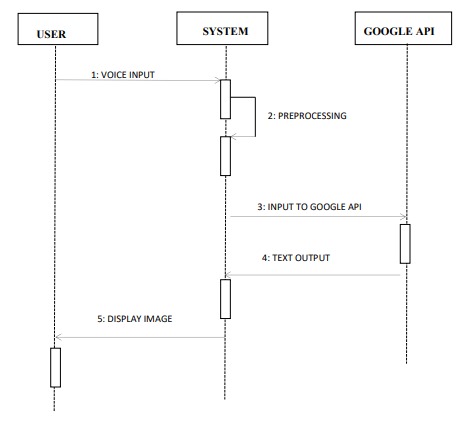


Fig 3: Audio Input

Input from the user is sent to the system , where appropriate modules are selected for processing the request. Once processed , output is displayed to the user.

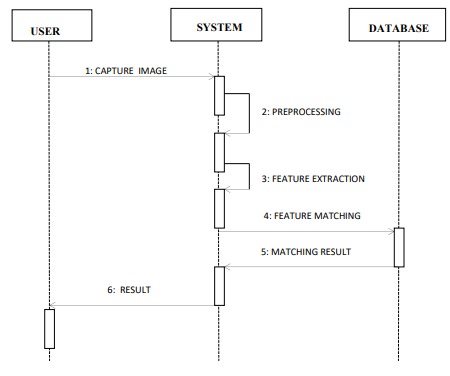
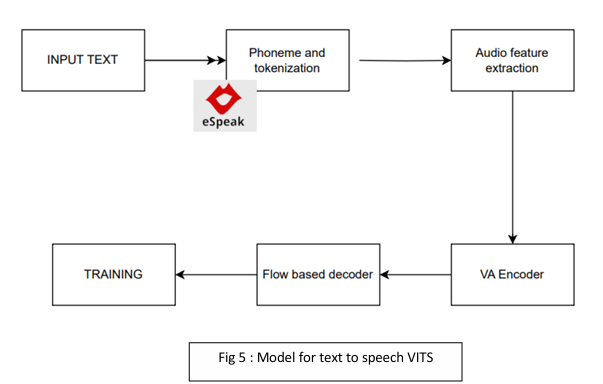
* 1. **Feature: Visual Input Processing**

Fig 4 : Visual Input

******

**5. Other Nonfunctional Requirements**

## 5.1 Performance Requirements

The software should perform real-time processing of inputs (audio, visual, text) and provide accurate output within a short time frame (preferably under 2 seconds). This is particularly important in use cases such as emergency communications or live educational sessions where immediate responses are required. The system must maintain at least 95% accuracy in transcriptions and conversions, including support for various accents and regional dialects in the speech-to-text conversion. The visual recognition component must detect and interpret gestures from ISL accurately under varying lighting and environmental conditions.

## 5.2 Safety Requirements

The software must ensure that no user data is mishandled or exposed to unauthorized access. Since the software may be used in sensitive environments like healthcare or public spaces, it must adhere to local safety standards, including ensuring that the visual outputs do not cause harm or discomfort to visually impaired users. The application should fail gracefully, with proper notifications to the user, in case of any operational errors. Safety certifications, including ADA (Americans with Disabilities Act) or regional equivalents, must be met when used for assistive technology purposes.

## 5.3 Security Requirements

The system should incorporate strong user authentication protocols to protect against unauthorized access, especially in healthcare or other sensitive environments. It should ensure that personal data, such as recorded audio or video for conversion, is encrypted both in storage and in transit. Adherence to global privacy standards like GDPR or India’s PDP (Personal Data Protection) Bill is mandatory. Additionally, the system should include role-based access controls to limit sensitive functions to authorized users.

* **CODING [5]**
* **Bibliography/References**

1. [**https://openurl.ebsco.com/EPDB%3Agcd%3A10%3A26295082/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A177050355&crl=c&link\_origin=scholar.google.com**](https://openurl.ebsco.com/EPDB%3Agcd%3A10%3A26295082/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A177050355&crl=c&link_origin=scholar.google.com)
2. [**https://www.researchgate.net/publication/320402323\_Sign\_Language\_Recognition\_Application\_Systems\_for\_Deaf-Mute\_People\_A\_Review\_Based\_on\_Input-Process-Output**](https://www.researchgate.net/publication/320402323_Sign_Language_Recognition_Application_Systems_for_Deaf-Mute_People_A_Review_Based_on_Input-Process-Output)
3. [**https://dspace.mit.edu/handle/1721.1/138512**](https://dspace.mit.edu/handle/1721.1/138512)
4. [**https://ieeexplore.ieee.org/abstract/document/10593656**](https://ieeexplore.ieee.org/abstract/document/10593656)
5. [**https://github.com/aanxieee/isl\_connect**](https://github.com/aanxieee/isl_connect)