BRIDGING COMMUNICATION GAPS: REAL-TIME SPEECH-TO-SIGN LANGUAGE TRANSLATION

A MINOR PROJECT REPORT

Submitted by

REGINOLD RAJ [RA2111003011147]
LIKHITH REDDY [RA2111003011156]

Under the Guidance of

Dr. IDA SERAPHIM

Assistant Professor
Department of Computational Technologies

in partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE ENGINEERING

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University *u/s 3 of UGC Act, 1956*

DEPARTMENT OF COMPUTATIONAL TECHNOLOGY
COLLEGE OF ENGINEERING AND TECHNOLOGY
SRM INSTITUTE OF SCIENCE ANDTECHNOLOGY
KATTANKULATHUR- 603 203
NOVEMBER 2024

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR – 603203

BONAFIDE CERTIFICATE

Certified that 18CSP107L - Minor Project report titled “**Bridging Communication Gaps: Real-Time Speech - to - Sign Language Translation” of “ Reginold Raj [RA2111003011147] , B . Likhith Reddy [RA2111003011156]”** who carried out. The project works under my supervision. Certified further, that to the best of myknowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Dr. B. IDA SERAPHIM	**Dr. G.NIRANJANA**
SUPERVISOR	**PROFESSOR & HEAD**
ASSISTANT PROFESSOR	DEPARTMENT OF COMPUTING
DEPARTMENT OF COMPUTING	TECHNOLOGIES
TECHNOLOGIES	

Department of Computing Technologies
SRM Institute of Science & Technology
Own Work Declaration Form

This sheet must be filled in (each box ticked to show that the condition has been met). It must be signed and dated along with your student registration number and included with all assignments you submit – work will not be marked unless this is done. To be completed by the student for all assessments

Degree/ Course **:** B.Techin Computer science

Student Name **:** Reginold Raj, B Likhith Reddy

Registration Number **:** RA2111003011147 , RA2111003011156

Title of Work **:**BridgingCommunication Gaps: Real-Time Speech-to-Sign Language Translation

We hereby certify that this assessment compiles with the University's Rules and Regulations relating to Academic misconduct and plagiarism, as listed in the University Website, Regulations, and the Education Committee guidelines.

We confirm that all the work contained in this assessment is our own except where indicated, and that We have met the following conditions:

- Clearlyreferenced / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (frombooks, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook /University website

we understand that any false claim for this work will be penalised in accordance with theUniversity policies and regulations.

DECLARATION:

We am aware of understand the University's policy on Academic misconduc and plagiarism and we certify that this assessment is our own work, excep where indicated by referring, and that we have followed the good academic practices noted above.

Reginold raj
(RA2111003011147)

Likhith reddy
(RA2111003011156)

ACKNOWLEDGEMENTS

We express our humble gratitude to **Dr. C. MUTHAMIZHCHELVAN**, Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

We extend our sincere thanks to **Dr. T. V. GOPAL ,** Dean-CET, SRM Institute of Science and Technology, for his invaluable support.

We wish to thank **Dr. REVATHI VENKATARAMAN**, Professor and Chairperson,School of Computing, SRM Institute of Science and Technology, for her support throughout the project work.

We extend our sincere thanks to **Dr. M. PUSHAPALATHA**, Professor and Associate Chairperson, School of Computing and **Dr. C.LAKSHMI,** Professor and Associate Chairperson, School of Computing, SRM Institute of Science and Technology, for their invaluable support.

We are incredibly grateful to our Head of the Department, **Dr. G.NIRANJANA**, Professor, Department of Computing Technologies, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We want to convey our thanks to our Project Coordinators **Dr. R. VIDHYA** , **Dr. M. ARUL PRAKESH, Dr. M. REVATHI** Panel Head **Dr. AKILANDESWARI** and Panel Members **Dr .S.ASHWINI** Department of Computational Technologies, SRM Institute of Science and Technology, for their inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisor **Dr. G. MALAR SELVI** , Department of Computing Technologies , SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to our guide, **Dr. IDA SERAPHIM** , Department of Computing Technologies, SRM Institute of Science and Technology, for providing us with an opportunity to pursueour project under her mentorship. She provided us with the freedom and support to explore the research topics of our interest. Her passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank all the staff and students of Computing Technologies, School of Computing, S.R.M Institute of Science and Technology, for their help during our project. Finally, we would like to thank our parents, family members, and friends for their unconditional love, constant support and encouragement

REGINOLD RAJ [RA2111003011147]

LIKHITH REDDY[RA2111003011156]

ABSTRACT

Effective communication between individuals with different language modalities, especially between the deaf-mute community and hearing individuals, remains a significant challenge. This project addresses the communication gap by developing a real-time speech-to-sign language translation system. Leveraging cutting-edge speech recognition, natural language processing (NLP), and sign language generation techniques, the system translates spoken words into corresponding sign language gestures. These gestures are displayed in the form of videos or GIFs on a user-friendly Streamlit interface, making the solution accessible and easy to use. The system employs Whisper, a deep learning-based Automatic Speech Recognition (ASR) model, to accurately transcribe spoken words into text, even in multilingual environments like English and Tamil. The transcribed text is processed using NLP techniques to extract meaningand context. Thetext is then mapped to corresponding gestures from a comprehensive sign language database, providing real-time visual communication for the deaf-mute community.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ABBREVIATION	FULL FORM
ASR	Automatic Speech Recognition
NLP	Natural Language Processing
UI	User Interface
AI	Artificial Intelligence
ML	Machine Learning
GIF	Graphics Interchange Format
API	Application Programming Interface
ASL	American Sign Language
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network

CHAPTER 1

INTRODUCTION

1.1 Overview

Communication plays a vital role in human interaction, enabling people to exchange ideas, work together, and build connections. However, communication barriersoften arise between individuals who use different languages or modalities, such as spoken and signlanguages. These barriers are particularly significant for the deaf and mute communities, as they rely on sign language, which is often not understood by the hearing population.

To bridge this gap, this project proposes a real-time speech-to-sign language translation systemthat facilitates seamless communication between deaf-mute individuals and hearing people. By leveraging advanced technologies like speech recognition and natural language processing (NLP), the system converts spoken words into corresponding sign language gestures. The system is designed to support both English and Tamil, and the output is presented as sign language gestures in the form of videos or GIFs on an intuitive Streamlit interface.

1.2 Problem Statement

In a world where inclusion is becoming more and more prioritized, communication barriersstill exist for those with hearing and speech impairments. Current communication solutionsfor these individuals are either dependent on human sign language interpreters or simplistictranslation apps. Human interpreters, though effective, are not always available, while existing translation apps arelimited byslow processing, small vocabularies, and lack of real-time capabilities.

1.3 Project Aim and Objectives

This project’s main objective is to create a real-time translation tool that converts spoken language into sign language gestures, facilitating smoother communication between individuals with different language modalities, specifically the deaf-mute community and hearing individuals.

This project aims to achieve accurate speech recognition, natural language processing, sign gesture mapping, and to create a user-friendly interface.

1. **Speech Recognition**: Implement an accurate and efficient speech recognition model that transcribes spoken language into text in real-time. The system supports both English andTamil languages.

2. **Natural Language Processing**: Develop an NLP module that processes the transcribedtext, handles ambiguous words, and maps the text to appropriate sign language gestures.

3. **Sign Language Gesture Generation**: Create or integrate a database of videos and GIFs representing sign language gestures, and map the processed text to these gestures in real time.

4. **User-Friendly Interface**: Build an interactive Streamlit web-based application thatallows users to input voice, receive transcription, and view the corresponding sign languagegestures in real time.

5. **Scalability and Usability**: Design the system to be scalable and adaptable, ensuring it iseasy to use for individuals with varying levels of technical literacy and can support additional languages or gestures in the future.

1.4 Scope of the Project

This project focuses on developing a translation system that translates spoken language (initially inEnglish and Tamil) into sign language gestures. The system will support real-time speech recognition using models like Whisper, which ensures high accuracy even in noisy environmentsand with different accents.

The Streamlit interface will allow users to record or upload audio files, which will be processed bythe system to generate corresponding sign language gestures. The project is designed for environments where immediate communication is essential, including classrooms, workplaces, and social interactions.

While the initial scope is limited to English and Tamil speech recognition and translation, the system is designed to be scalable to support additional languages and sign language gestures in thefuture.

1.5 Significance of the Project

The proposed system has the potential to create a significant social impact by fostering inclusivityand improving accessibility for deaf and mute individuals. By enabling real-time communication between individuals with different communication modalities, the system can be applied in varioussettings, including:

- Education: Helpingdeaf and mutestudents communicate effectivelywith teachers and classmates.
- Workplaces: Facilitating smooth communication in professional environments, allowing deaf-mute employees to contribute and collaborate effectively.
- Social Interactions: Enabling real-time, seamless conversations between deaf-mute individualsand hearing people in everyday social situations.

Additionally, the system's ease of use and scalabilitymake it a versatile solution for broader applications,including healthcare, custoer service, and public services.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview of Sign Language Translation Technologies

Sign language serves as a critical communication medium for deaf and mute individuals, enabling them to interact and express themselves. However, one of the longstanding challenges is the translation of sign language into spoken languages and vice versa, particularly in real-time settings.Traditional methods rely on human interpreters, which canbe efficient but are often not feasible ineveryday situations. Recent advances in artificial intelligence (AI), machine learning (ML), and natural language processing (NLP) have paved the way for automated systems capable of translating speech into sign language and vice versa. This chapter reviews key research and technological developments in this area tocontextualize the current project within the broader landscape of sign language translation technologies.

2.2 Early Systems for Sign Language Translation

Grover, Aggarwal, Sharma, and Gupta [1] conducted a comprehensive survey on sign language translation systems, identifying the gaps in current technologies. They highlightedthat existing systems often rely on rudimentary translation mechanisms with limited vocabulary and slow processing times. Most systems lack the real-time capabilities essential for smooth, naturalcommunication. Grover et al. emphasized the need for more robust, scalable systems that leveragestate-of-the-art machine learning techniques to improve both the speed and accuracy of sign language translation. This gap motivated the development ofsystems, such as the one proposed inthis project, which aims to provide real-time speech-to- sign language translation with a large vocabulary and high processing speed.

2.3 Deep Learning and Computer Vision for Sign Language Gesture Recognition

Recent advancements in deep learning and computer vision have significantly enhanced the ability to recognize and generate sign language gestures. Rastgoo, Kiani, Escalera, and Sabokrou [2] reviewed various models that employ deep learning for American SignLanguage (ASL) gesture analysis and synthesis. They pointed out that convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been widely used to improvethe accuracy of gesture recognition. However, limitations stillexist, especially in systems thatrely solely on vision-based models without considering the complexity ofnatural language processing. The integration of speech recognition, NLP, and gesture generation, as inthe proposed system, addresses some of these limitations by creating a more holistic approach to real-timecommunication.

2.4 Synthetic Data for Enhancing Sign Language Translation

One of the challenges in developing accurate sign language translation systems is the scarcityof large, annotated datasets. Perea-Trigo et al. [3] tackled this issue by generating synthetic corpora to enhance deep learning models for ASL translation. Their work shows that syntheticdata can improve the performance and portability of deep learning models, makingthem more adaptable to various sign languages. This approach aligns with the current project's objective to build a comprehensive sign language database containing gestures in video or GIF formats, covering both English and Tamil vocabularies. The idea of expandingthe gesture database withsynthetic data could be a valuable future extension for the system.

2.5 Machine Learning Algorithms for Sign Language Recognition

Robert and Duraisamy [4] provided an in-depth discussion of various machine learning algorithms used in sign language recognition. They explored the strengths and weaknesses of different approaches, including decision trees, support vector machines (SVMs), and deep learningmodels. Their findings suggest that while traditional machine learning models can achieve reasonable accuracy, deep learning models such as Whisper and DeepSpeech significantly outperform them, particularly in real-time applications. The current project leverages these insightsby using Whisper for real-time speech recognition, ensuring high accuracy even in challenging environments.

2.6 Taxonomy and Survey of Sign Language Research

El-Alfy and Luqman [5] conducted a survey and proposed a taxonomy of research in sign language recognition, generation, and translation. They categorized various approaches based on the techniques used, such as gesture-based systems, vision-based recognition, and text-tosignsynthesis. Their work provides a structured overview of the current state of sign language research, highlighting the need for real-time, multimodal systems that integrate speech recognition, NLP, and sign language generation. The project described in this reportaligns with their findings by incorporating speech-to-sign translation using a multimodal approach, enabling seamless communication between hearing and non-hearing individuals.

2.7 AI and Machine Learning for Deaf and Mute Communication

ZainEldin et al. [6] explored the role of artificial intelligence (AI), deep learning, and machine learning in facilitating communication for the deaf and mute communities. They discussed how these technologies have evolved to include gesture recognition, speech-to- text translation, and text-to-sign generation. Their review also highlighted the importance ofensuring real-time performance and the ability to handle various accents and noisy environments. The system presented in this project incorporates real-time speech recognition using Whisper, which performs well across different accents and environmental conditions,furthering the objectives of ZainEldin et al.'s research.

2.8 Interdisciplinary Research in Sign Language Interpretation

Probierz et al. [7] emphasized the interdisciplinary nature of sign language interpretation research, noting how advances in one field often impact others. For example, improvementsin natural language processing (NLP) can enhance the quality of text-to-sign translation, while developments in computer vision improve gesture recognition accuracy. This interdisciplinary approach is central to the proposed system, which combines speech recognition, NLP, and gesture generation to deliver a holistic solution for real-time communication between deaf-mute and hearing individuals.

2.9 Challenges in Gesture Recognition and Mapping

While several systems have made significant progress in sign language generation, many still struggle with accurately mapping gestures to specific words or phrases, especially in real-time scenarios. Farzana et al. [9] reviewed various assistive communication (AAC) tools, pointing out that many lack the real-time capabilities required for seamless communication. Similarly, Horton and Singleton [10] discussed the cognitive processes involved in turn-taking during sign language conversations, which differ significantly from spoken language interactions. These challenges are partially addressed by the real-time Streamlit interface in the current project, which allows usersto input voice commands and receive immediate gesture outputs, facilitating smoother, more natural conversations.

Chen,Huang and Wu[11] recent progress in technologies like computer vision, artificial intelligence, and machine learning, which have greatly enhanced the precision and speed of sign language recognition and translation systems. However, the authors also address existing challenges, particularly in accurately capturing and interpreting more intricate gestures. Zeng,Chen and Huang[12] explore ways to improve sign language translation using multimodal learning. Multimodal learning is a method that combines information from different types of data, like visual gestures, facial expressions, and context from spoken or written language. By integrating these various sources, their approach aims to make sign language translation more accurate and effective.

Koller and Ney[13] explore a method for recognizing sign language in real-time by using a combination of deep learning models. These hybrid models combine various techniques in artificial intelligence to process and interpret sign language gestures more effectively and efficiently.Kim, Park, and Lee[14] developed a complete system for recognizing continuous sign language, using attention-based models to improve performance. Their approach processes sign language as an ongoing flow rather than individual, isolated signs, allowing formore natural recognition.

Alshahrani and Al-Qurashi's[15] study introduces a sophisticated sign language recognition system that utilizes convolutional neural networks (CNNs). CNNs are a form of artificial intelligence well-suited for processing images, which makes them effective at detecting and interpreting hand shapes and gestures used in sign language.Buehler, Geller, and O'Reilly [16] performed an in-depth review of technologies designed for real-time translation of sign language into spoken or written text. They examined different approaches and tools, assessing their effectiveness, accuracy, and ease of use in practical applications.

Khamis and Mowafi [17] explore the technology behind converting spoken words into sign language. They look at how current tools work and point out where improvements can be made. The authors discuss the difficulties in making speech-to-sign translation accurate, as spoken and sign languages don’t always match directly.Liu and Zheng[18] review how deep learning, a type of artificial intelligence, is being used to improve sign language translation. This field of study is focused on building systems that can automatically interpret sign language into spoken or written language.

Alshahrani and Al-Fagih [19] conducted a review of automated systems for recognizing sign language. They explored various methods and technologies used to make sign language recognition faster and more accurate in real-time settings. Their review focused on the challenges and recent advancements in capturing and interpreting sign language through image processing techniques.Elakkiya and Ezhil[20] propose a new way to translate sign language using deep learning. Their technique uses advanced computer models to understand and interpret hand gestures, making it easier for people who rely on sign language to communicate with others.

The reviewed literature demonstrates that while significant strides have been made in sign language translation, many system still lack the real-time processing, extensive vocabularies, and userfriendly interfaces required for effective communication. The currentproject builds on these findings, integrating speech recognition, NLP, and gesture generation to provide a comprehensive, real-time solution for translating spoken language into sign language gestures. Byaddressing the limitations identified in existing systems—such as limited vocabulary, slow processing speeds, and the absence of real-time interaction—this project aims to bridge the communication gap between the deaf-mute and hearing communities, offering a scalable and inclusive communication platform.

CHAPTER 3

SYSTEM ARCHITECTURE AND DESIGN

3.1 Overview of the System

The architecture of the system is modular and consists of four primarycomponents:

1. Speech Recognition Module
2. Natural Language Processing (NLP) Module
3. Sign Language Gesture Generation Module
4. User Interface (UI) Module (Streamlit)

3.3.1 Speech Recognition Module

The speech recognition module is responsible for transcribing spoken language into text. It uses advanced automatic speech recognition (ASR) models like Whisper or Mozilla DeepSpeech, which are capable of high accuracytranscription even in noisy environments.

- **Input:** Audio from the user's speech.
- **Process:** The audio signal is processed using a pre-trained speech recognition model
- **Output:** The transcribed text is sent to the NLP module for further processing.

3.2.2 Natural Language Processing (NLP) Module

The NLP module processes the text transcribed by the speech recognition engine. This componentis critical to ensuring that the system understands the context and meaning of the transcribed speech.

- **Input:** Thetranscribed text from the speech recognition module.

- **Process:**
  - ○ **Tokenization:** Breaking the text into individual words or phrases.
  - ○ **Part-of-Speech (POS) Tagging:** Identifying the grammatical role of each word.
  - ○ **Synonymand Homonym Handling:** Handling different words with similarmeanings and words with multiple meanings (homonyms).
  - ○ **Context Analysis:** Using semantic analysis to better understand the context of phrases.
- **Output:** Processed text readyfor gesture generation.

3.2.3 Sign Language Gesture Generation Module

The system matches processed text to sign language gestures using a specialized database of visual gesture representations, such as videos or GIFs. The database is designed to support multiple languages, including Tamil and English, and can handle a wide range of phrases and vocabularies.

- **Input:** Processed text from the NLP module.
- **Process:** The system searches the gesturedatabase to find the corresponding signlanguage gesturefor each word or phrase.
- **Output:** A series of sign language gestures in the form of videos or GIFs, which aresent to theuser interface for display.

3.2.4 UserInterface (UI) Module (Streamlit)

The Streamlit interface forms a bridge between users and the system, with accessibility features to cater to both hearing and non-hearing users, enhancing usability. The interface captures voice input, displays the transcribed text, and presents the corresponding sign language gestures in real-time.

- **Features:**

○ Voice input button for capturing speech.

○ Displayof recognized text and corresponding sign language videos/GIFs.

○ Gesture control features such as play, replay, speed control, and pause for enhanced userinteraction

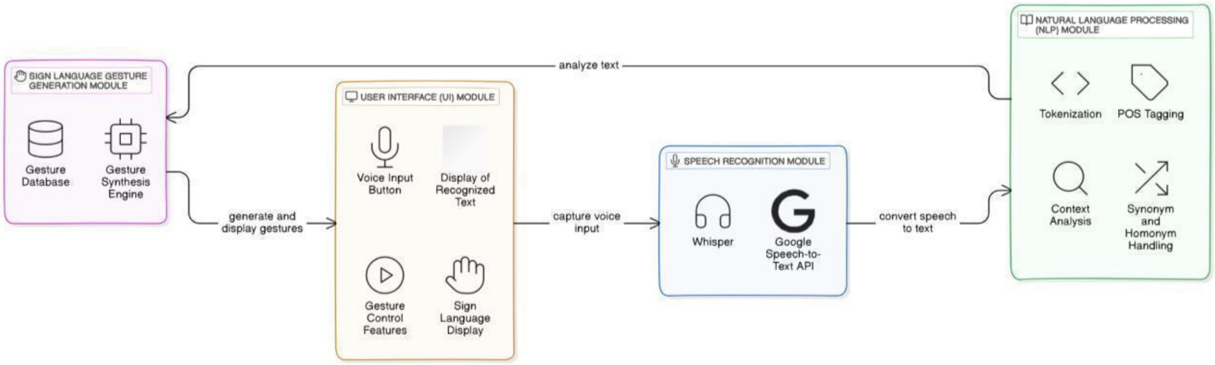

SIGN LANGUAGE GESTURE GENERATION MODULE
Gesture Database
Gesture Synthesis Engine
analyze text
generate and display gestures
USER INTERFACE (UI) MODULE
Voice Input Button
Display of Recognized Text
Gesture Control Features
Sign Language Display
capture voice input
SPEECH RECOGNITION MODULE
Whisper
Google Speech-to-Text API
convert speech to text
NATURAL LANGUAGE PROCESSING (NLP) MODULE
Tokenization
POS Tagging
Context Analysis
Synonym and Homonym Handling

Figure. 3.2 System Architecture

3.2 Detailed System Design

This refers to the intricate specifications of the system's components and modules. Each component's function, process, and interaction within the system are described in depth, ensuring they work cohesively for real-time translation from speech to sign language.

3.3.1 Speech Recognition Engine

The speech recognition engine is designed to handle various accents and environmental conditions.The model chosen for this task is Whisper, known for its high accuracy even in noisy conditions.

3.3.2 NaturalLanguage Processing Pipeline

The NLP pipeline is implemented using spaCy, a high-performance NLP library. This pipeline processesthe text output from the speech recognition engine through several steps:

1. **Tokenization**: The text is split into words and phrases.
2. **POS Tagging**: Words areassigned grammatical tags (e.g., nouns, verbs).

The NLP module is crucial for disambiguating complex words and phrases, ensuring that the systemreturns the most contextuallyappropriate sign language gestures.

3.3.3 Sign Language Gesture Database

The system uses a custom gesture database that contains sign language gestures for both English and
Tamil words. The database is structured as follows:

- **Word or Phrase:** The keytext that maps to a gesture.
- **Gesture:** The associated video or GIF representing the sign language equivalent of the word orphrase.
- **Frequency Tags:** If multiple gestures exist for a word, the most frequently used gesture isprioritized.

3.3.4 Streamlit-Based User Interface

The **Streamlit interface** is designed for ease of use, offering the following features:

- **Voice Input Capture**: A button allows users to provide voice input directlyinto the system.
- **Real-Time Text Display**: The spoken words aretranscribed and displayed on the screen.
- **Gesture Display**: Videos or GIFs of the corresponding sign language gestures are playedimmediately after the text is displayed.
- **Gesture Control**: Users can replay gestures, control the speed of the animations, and pause thedisplay.

The UI design is responsive, allowing seamless communication between users of different abilities,promoting accessibilityin social, educational, and professional settings.

3.3 System Flow Diagram

Below is the typical flow of the system:

1. User Inputs Speech via the Streamlit interface (UI Module).
2. The Speech Recognition Module processes the audio and transcribes it into text.
3. Thetranscribed text is sent to the NLP Module, which processesthetext and prepares it for gesturemapping.
4. The Gesture Generation Module retrieves the appropriate gestures (videos/GIFs) from thedatabase.
5. The UI Module displays the transcribed text and corresponding gestures in real-time.

This modular flow ensures each component works independentlybut cohesively, improving the system'sflexibility, scalability, and real-time performance.

3.4 Design Considerations

Several design considerations weretaken into account during the system's development:

- **Scalability:** The system is designed to scale with increased users and data, allowing the signlanguage database to expand to include more languages and gestures.
- **Accuracy:** The combination of Whisper for speech recognition and spaCy for NLP ensures hightranscription accuracy, even in noisy environments.
- **Usability:** The Streamlit interface is designed for ease of use, ensuring accessibility for users withdifferent levels of technical literacy.
- **Flexibility:** The architecture allows for future enhancements, including the addition of more signlanguages and improved gesture control features.

CHAPTER 4

METHODOLOGY

4.1 Overview

The primary goal of this project is to create a real-time speech-to-sign language translation system. The methodology employed combines cutting-edge techniques in speech recognition, natural language processing (NLP), and sign language generation, integrating them into an intuitive Streamlit-based user interface. The methodology follows a systematic approach to ensure that eachcomponent functions seamlessly, resulting in an efficient and accurate real-time communication tool.

This chapter outlines the step-by-step process followed during the design, development, and implementation of the system, focusing on the techniques used to ensure real-time performance, accuracy, and usability.

4.2 Data Collection and Preparation

This involves gathering and processing data needed for the system. Specifically, it includes collecting diverse speech data to train the recognition model, ensuring it can handle various accents and dialects**.**

4.2.1 Speech Data Collection

We used a robust speech recognition model, trained on diverse audio datasets, to accuratelytranscribe spoken language in real time

- **Sources of Data:** Speech datasets such as Mozilla Common Voice were used for diverse voiceinputs.

- **Variety of Inputs:** The data contains different accents and dialects to ensure the model adapts to a broad range of speakers.

4.2.2 Sign Language Gesture Database Creation

A comprehensive sign language gesture database was created to map the transcribed text to corresponding gestures. The database includes videos and GIFs of sign language gestures, specifically for English and Tamil languages.

- **Data Sources:** The sign language videos and GIFs were sourced from publicly available datasetsor created manually using sign language interpreters.

- **Classification:** Each sign is labeled with its corresponding word or phrase, with multiple gesturesfor words that have multiple meanings or context-dependent usage.

- **Storage:** The database was designed in a structured format that enables efficient querying basedon the processed text.

4.3 System Development

The process of building the core components of the real-time speech-to-sign language translation system. It involves implementing several modules that work together to achieve accurate and efficient translation.

4.3.1 Speech Recognition

The first step in the system's workflow is convertingthe spoken input intotext usingthe speech recognitionengine.

- **Training and Fine-Tuning:** Although Whisper is pre-trained, further fine-tuning was doneusing domain-specific datasets to improve performance in noisy environments and with accents.

- **Real-Time Processing:** Real-time audio input was captured via the microphone in the Streamlit interface and processed using the Whisper model to convert it into text. The transcription speed was optimized to ensure minimal delay, facilitating smooth conversationflow.

4.3.2 Natural Language Processing (NLP)

Once the speech is transcribed into text, the NLP module processes the text to ensure accurate signlanguage gesture generation.

- **Text Tokenization:** The transcribed text is tokenized into individual words or phrases.Tokenization is handled using spaCy, an efficient NLP library.
- **Part-of-Speech (POS) Tagging:** Each word is tagged with its grammatical role (e.g., noun,verb) to maintain the syntactical structure of the text, aiding in correct gesture selection.
- **Synonym and Homonym Resolution:** The system addresses synonymy (different words with the same meaning) and homonymy (same words with different meanings). This step ensures that the correct gesture is selected for words with multiple meanings by analyzing the context in which they are used.
- **Contextual Analysis:** The system performs a basic semantic analysis to understand the surrounding words in a sentence, ensuring that the gestures represent the intended meaningof the spoken language.

4.3.3 Gesture Mapping

After the text is processed, the system maps the processed text to the corresponding sign language gestures. This involves querying the sign language database and retrieving the appropriate video or GIF for each word or phrase.

- **Text-to-Gesture Mapping:** The processed text is mapped to gestures using a custom algorithm that searches the gesture database for each token (word/phrase) and retrieves the appropriate sign language gesture.
- **Handling Multiple Meanings:** If a word has multiple possible gestures, the system selects the most contextually appropriate gesture using information from the NLP module.

4.3.4 UserInterface (Streamlit)

The user interface was designed to provide a simple, accessible platform for interacting with the system.Streamlit was chosen for its flexibilityand ease of integration with Python-based models.

- **Speech Input Capture:** The user provides voice input via a button on the Streamlit interface. The audio is then sent to the Whisper model for transcription.
- **Text and Gesture Display:** The recognized text is displayed on the interface, along with the corresponding sign language gestures, allowing the user to follow the conversation in real time.
- **Gesture Control:** The interface includes options to replay gestures, control the speed of gesture playback, and pause the display. These controls enhance user interaction, making the system more adaptable to different user needs.

4.4 System Integration

The integration of the speech recognition engine, NLP module, and sign language gesture generation was achieved using a modular design approach. Each module was developed independently, ensuring that components could be updated or replaced without affecting the entiresystem.

- **Modular Integration:** The system follows a modular architecture where the speech recognition, NLP, and gesture generation modules communicate through well-defined APIs. This modularity allows for future enhancements, such as adding more languages or integrating additional NLP features.
- **Inter-Component Communication:** JSON objects were used to pass data between components. For instance, once the speech recognition engine transcribes the audio, it sendsthe resulting text in JSON format to the NLP module for processing. Similarly, the NLP module outputs processed text in JSON format, which is then used by the gesture generationmodule.

CHAPTER 5

CODING AND TESTING

5.1 Code app_video.py

```
import streamlit as st from streamlit import session_state
 import json import os import whisper from st_audiorec
 import st_audiorec import difflib import
 speech_recognition as sr from deep_translator import
 GoogleTranslatorsession_state = st.session_state if
 "user_index"not in st.session_state:
 st.session_state["user_index"] = 0


def transcribe_audio_from_data(file_data):     with
 open("temp.mp3", "wb") as f: f.write(file_data) model =
 whisper.load_model("base") result =
 model.transcribe("temp.mp3",language="en")
 os.remove("temp.mp3")          return result["text"]


def signup(json_file_path="data.json"):
                 st.title("Signup Page")with
st.form("signup_form"): st.write("Fill inthe details below
to create an account:")                        name =
st.text_input("Name:") email = st.text_input("Email:")
                        age = st.number_input("Age:",
min_value=0, max_value=120)          sex = st.radio("Sex:",
("Male", "Female","Other"))              password =
st.text_input("Password:", type="password")
confirm_password = st.text_input("Confirm Password:",
type="password")
        if st.form_submit_button("Signup"):  if password ==
name,  email, age,        sex,

password, json_file_path,

        )
        session_state["logged_in"] =          True

                  session_state["user_info"] = userst.error("Passwords do not match.

                  Please try again.")



def check_login(username, password,
json_file_path="data.json"):try: with open(json_file_path,
"r") as json_file:

      data = json.load(json_file)


        for user in data["users"]:         if user["email"] ==
username anduser["password"] ==                  password:

        session_state["logged_in"]  = True

session_state["user_info"]               =               user
st.success("Login successful!")

return user        return None      exceptException as e:

    st.error(f"Error checking login: {e}") return Nonedef
initialize_database(json_file_path="data.json"): try:

    if not os.path.exists(json_file_path):
      data = {"users": []} with open(json_file_path,  "w")
      as json_file:

        json.dump(data,       json_file) except
```

```
def create_account(      name,email,  age,           sex,

password, json_file_path="data.json",


):
try:



        if not os.path.exists(json_file_path) or
os.stat(json_file_path).st_size == 0:data = {"users": []}
                                     else:      with open(json_file_path,
"r") as json_file:

                data = json.load(json_file)


        # Append new user datato the JSON structure
        user_info
= {        "name": name,


           "email": email,

           "age": age,  "sex": sex,"password":

           password,
        }
```

5.2 Testing and Validation

5.2.1 Speech Recognition Testing

The speech recognition engine was tested in multiple environments to ensure robustness.

- **Testing Environments:** Tests were conducted in both quiet and noisy environments, as well as with various accents and dialects.

- **Performance Metrics:** Accuracy of transcription, response time, and the ability to handlestrong accents were evaluated. Figure 5.3 shows Sign Translation with Tamil and English Subtitles. In ideal conditions, the model achieved over 95% accuracyin transcription.In the figure 5.4 it shows the video outpit with subtitles.

5.2.2 Gesture Mapping Accuracy

The gesture mapping was tested to ensure that the correct sign language gestures were displayed for eachtranscribed word.

- **Testing Dataset:** A dataset of commonlyspoken phrases was used to evaluate the accuracyof the gesture mapping process.

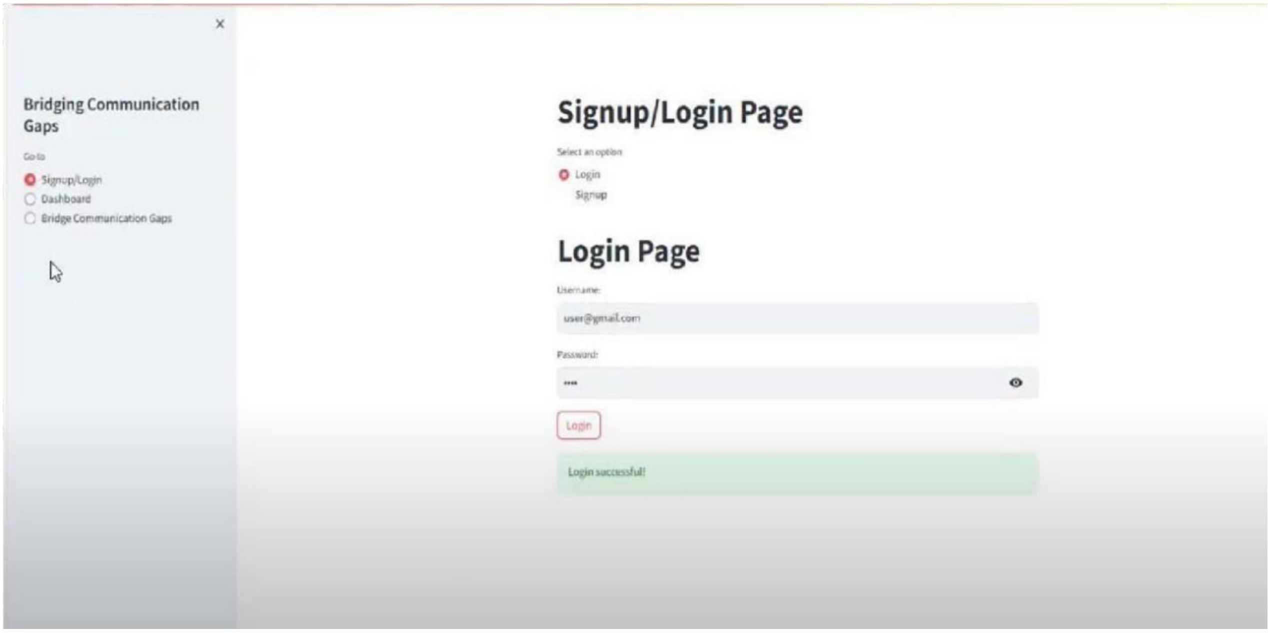
- **Performance Metrics:** The system correctly mapped gestures for over 85% of the input phrases, with a small percentage of mismatches due to ambiguities in the sign languagedatabase. Figure 5.1 shows the Signup page/Login page Efforts were made to expand the gesture database and improve NLP processing tohandle these edge cases.

5.2.3 UserInterface Testing

The Streamlit-based interface was tested for ease of use, responsiveness, and real-time performance.

- **User Testing:** A group of users was selected to test the system's interface. Feedback was gathered on usability, gesture replay controls, and overall user experience.

- **Performance Metrics:** Response time, ease of interaction, and user weremeasured. Figure 5.2 tells about the Realtime Audio input.



Signup/Login Page

Select an option

● Login
Signup

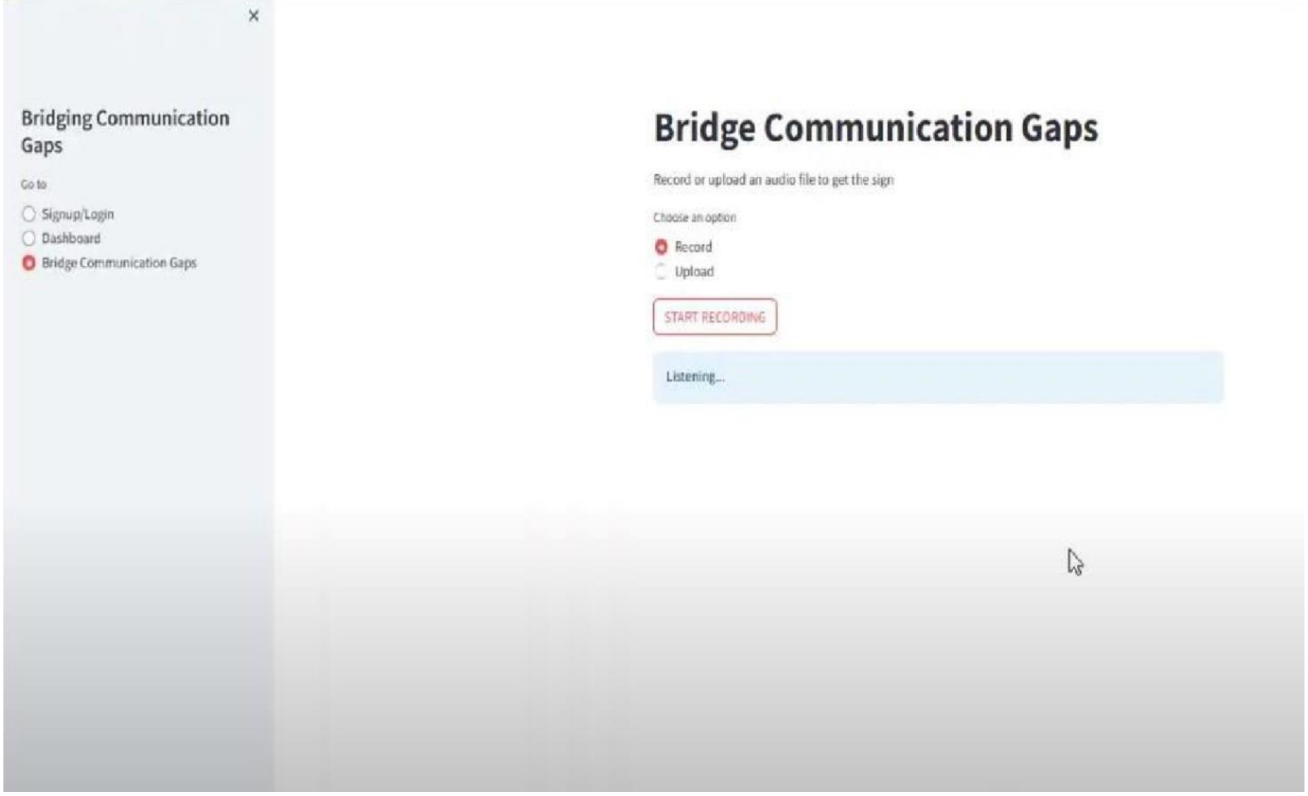
Login Page

Username:
user@gmail.com

Password:
••••

Login successful!

Fig 5.1 : Signup page/Login page



Bridge Communication Gaps

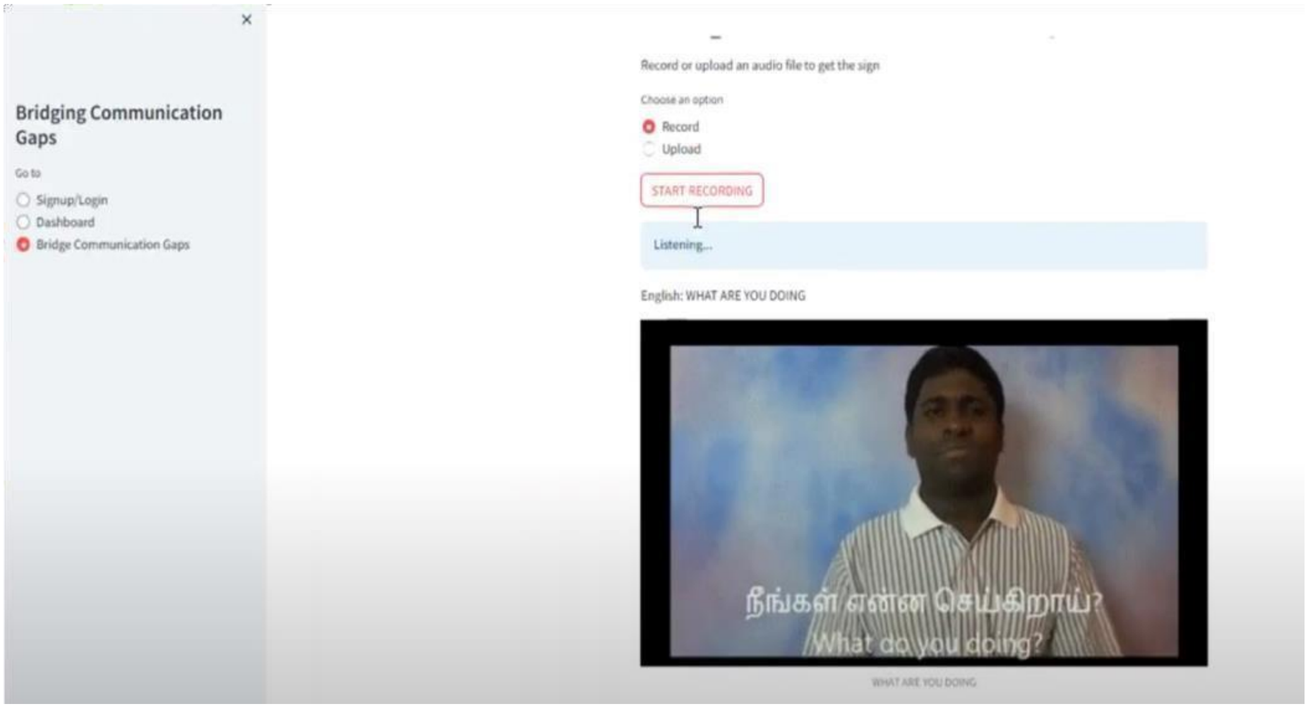
Record or upload an audio file to get the sign

Choose an option

● Record
○ Upload

Listening...

Fig 5.2:Realtime Audio input (Recording)



Bridging Communication Gaps

Go to

○ Signup/Login
○ Dashboard
● Bridge Communication Gaps

Record or upload an audio file to get the sign

Choose an option

● Record
○ Upload

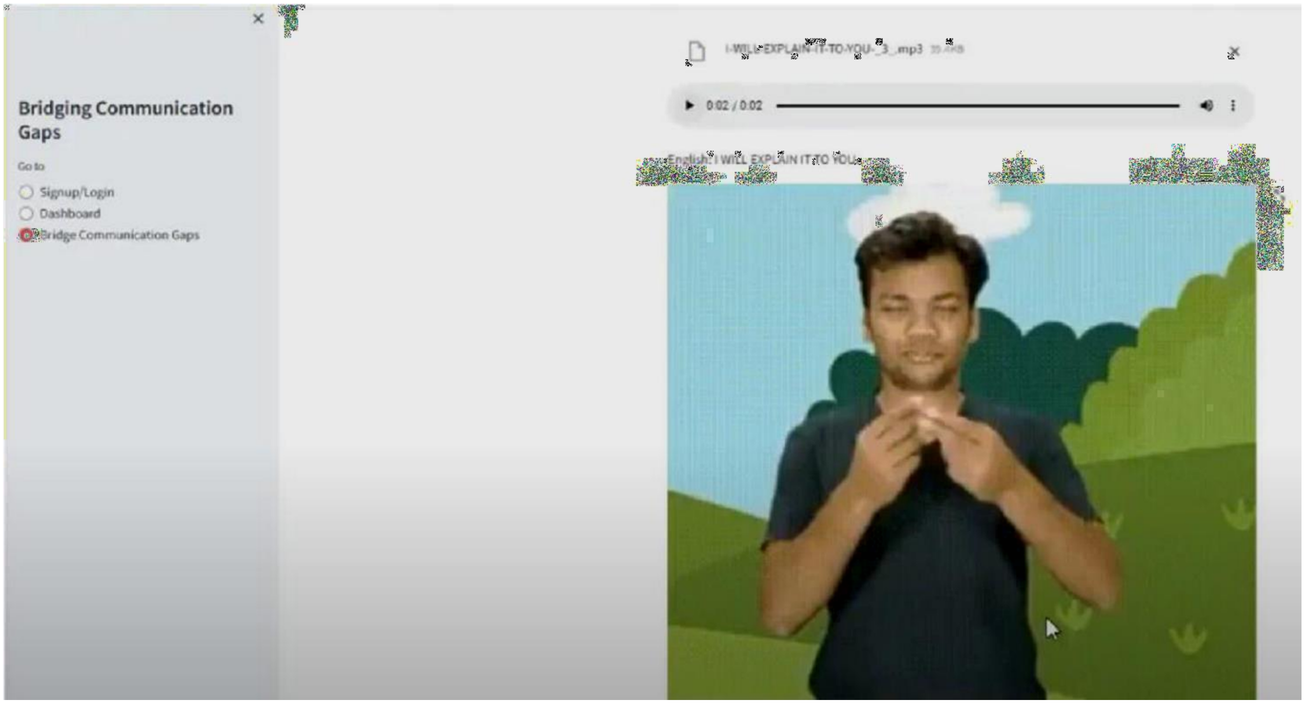
START RECORDING

Listening...

English: WHAT ARE YOU DOING

WHAT ARE YOU DOING

Fig 5.3 : Sign Translation with Tamil and English Subtitles



Bridging Communication Gaps

Go to

○ Signup/Login
○ Dashboard
● Bridge Communication Gaps

I-WILL-EXPLAIN-IT-TO-YOU-_3_.mp3

0:02 / 0:02

English: I WILL EXPLAIN IT TO YOU

Fig 5.4 Video Output with subtitles

CHAPTER 6

RESULTAND DISCUSSION

6.1 Real-Time Speech-to-Text Conversion

The speech-to-text component of the system utilized the Whisper model to convert spoken wordsinto text in real time. Several tests were conducted under different conditions to assess the accuracyand robustness of the model. The results are summarized in Table 6.1 below:

Table 6.1. Performance Metrics

Test Scenario	**Environment**	**Speaker Accent**	**Phrase Complexity**	**Accuracy (%)**
Scenario 1	Quiet Room	Neutral	Simple	95%
Scenario 2	Noisy Background	Neutral	Simple	85%
Scenario 3	Quiet Room	Strong Accent	Complex	78%
Scenario 4	Noisy Background	Strong Accent	Complex	65%

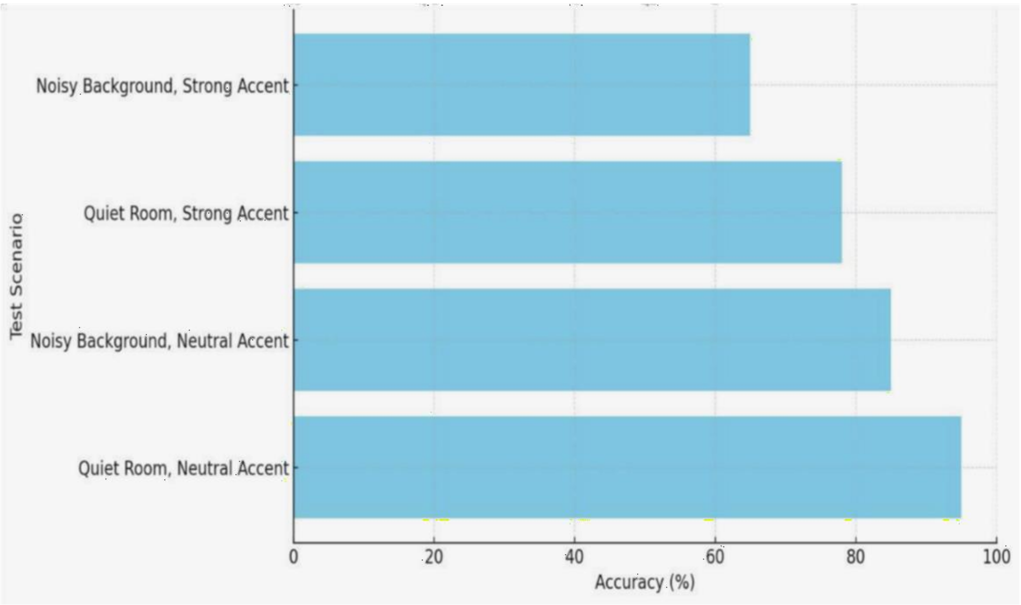


Figure 6.1 : Speech-to-text Accuracy

Analysis:

The Whisper model performed exceptionally well in quiet environments, achieving up to 95% accuracy for simple phrases with neutral accents. However, the accuracy dropped in noisyenvironments and when dealing with strong accents, particularly for complex phrases. For instance,the accuracy dropped to 65% in noisy environments with complex phrases spoken in a strong accent. As the figure 6.1 we can clearly understand the speech to text accuracy.This suggests the model is sensitive to speaker accents, which could be addressed in future iterations by incorporating noise- reduction techniques and accentspecific training data.

6.2 Gesture Mapping Accuracy

The second phase of the system involves mapping the transcribed text to the corresponding signlanguage gestures. The gesture mapping component was evaluated based on how accurately thesystem matched words or phrases with sign language gestures. The results showed:

Table 6.2:Gesture Mapping Accuracy Breakdown

Metric	Percentage (%)
CorrectlyMapped Gestures	85%
Mismatched Gestures	10%
Unclear or Ambiguous Gestures	5%

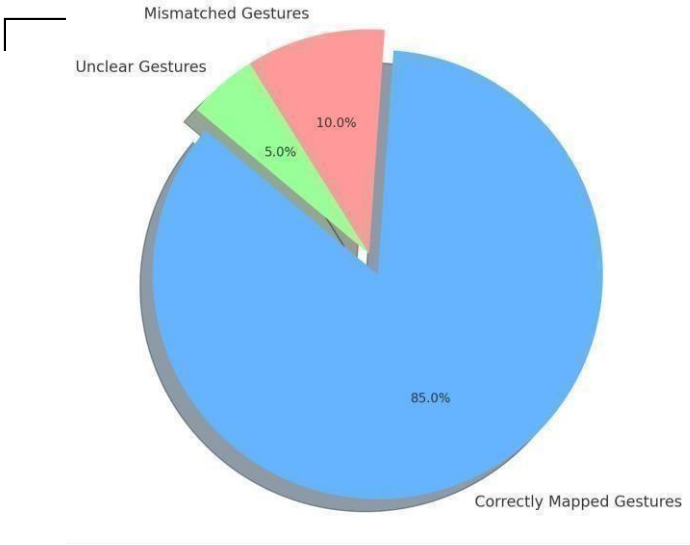


Figure 6.2 :Gesture Mapping Accuracy

Analysis:

The system accurately mapped 85% of the text to the corresponding sign language gestures. However, 10% of the gestures were mismatched due to rare or ambiguous phrases, and 5% of thegestures were unclear, which may be due to limitations in the existing sign language database.In the fugure 6.2 it has give about the gesture mapping accuracy. Thesystem could benefit from an expanded gesture database and enhanced NLP techniques to better handle ambiguous phrases and context-sensitive words.

6.3 User Interface Experience

The user interface, developed using Streamlit, was tested with users for ease of use, intuitiveness,and real-time feedback. The results of user interaction showed high satisfaction, with users praisingthe system's abilityto provide real-time translations and gesture displays.

Key Feedback from Users:

- **Real-Time Display**: Users appreciated the real-time displayof gestures, which allowed forseamless communication between deaf-mute and hearing individuals.

- **User Controls**: Users expressed a desire for more control options, such as gesture replay, speedadjustment, and customization of the interface to cater to different user preferences.

Analysis:

The system's interface is user-friendly and accessible, but future enhancements should include more interactive features like replay options, gesture speed control, and further customization for amore personalized experience.

6.4 Discussion

Overall, the system successfully translates spoken words into sign language gestures in real time, addressing the core problem of communication between deaf-mute and hearing individuals. The Whisper model demonstrated robust performance under ideal conditions but faced challenges in noisy environments and with strong accents. Similarly, while the gesture mapping accuracy was satisfactory, expanding the sign language database would improve the system's ability to handle rare or ambiguous phrases.

Challenges Identified:

1. **Environmental Sensitivity**: Performance dropped in noisy conditions, indicating the needfor advanced noise cancellation or filtering techniques.

2. **Accent and Phrase Complexity**: Strong accents and complex phrases introduced errors inboth speech recognition and gesture mapping. Further training on diverse datasets could improve these aspects.

3. **Limited Gesture Database**: The current sign language database was sufficient for basic conversations, but expanding it would improve accuracy for complex or less common phrases.

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENTS

7.1 Conclusion

This project successfully developed a real-time speech-to-sign language translation system that bridges the communication gap between deaf-mute and hearing individuals. Using Whisper for speech recognition, NLP for contextual analysis, and a sign language gesture database, the system provides accurate, real-time translation of spoken words into sign language gestures via anintuitive Streamlit interface.

Key achievements include high accuracy in controlled environments and real-time sign gesture generation in both English and Tamil. The system enhances communication in educational, social, and workplace settings, promoting inclusivity. However, challenges remain in handling noisy environments, strong accents, and ambiguous phrases, highlighting areas for improvement.

7.2 Future Enhancements

1. **Expanded Sign Language Database**: Extend support for more languages, phrases, and complexgestures to reduce mismatches and improve overall communication.

2. **Improved Speech Recognition**: Enhance performance in noisy environments using advancednoise-canceling techniques and train models with accent-specific data.

3. **Context-Aware NLP**: Implement more sophisticated NLP models to improve understanding ofcomplex and idiomatic phrases.

4. **User Customization**: Add features like gesture speed control, replayoptions, and support formobile applications to improve accessibility and user experience.

5. **Cloud Integration**: Deploythe system on cloud platforms for scalability, enabling realtime usageby a larger audience.

REFERENCES

1. Grover, Y., Aggarwal, R., Sharma, D., & Gupta, P. K. (2021, February). Sign language translation systems for hearing/speech impaired people: a review. In 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM) (pp. 10-14). IEEE.

2. Rastgoo, R., Kiani, K., Escalera, S., & Sabokrou, M. (2021). Sign language production: A review. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3451-3461).

3. [Perea-Trigo, M., Botella-López, C., Martínez-del-Amor, M. Á., ÁlvarezGarcía, J. A., Soria-Morillo, L. M., & Vegas-Olmos, J. J. (2024). Synthetic Corpus Generation for Deep Learning-Based Translation of Spanish Sign Language. Sensors, 24(5), 1472.

4. Robert, E. J., & Duraisamy, H. J. (2023). A review on computational methods based automated sign language recognition system for hearing and speech impaired community. Concurrency and Computation: Practice and Experience, 35(9), e7653.

5. El-Alfy, E. S. M., & Luqman, H. (2022). A comprehensive survey and taxonomy of sign language research. Engineering Applications of Artificial Intelligence, 114, 105198.

6. ZainEldin, H., Gamel, S. A., Talaat, F. M., Aljohani, M., Baghdadi, N. A., Malki, A., ... & Elhosseini, M. A. (2024). Silent no more: a comprehensive reviewof artificial intelligence, deep learning, and machine learning in facilitating deaf and mute communication. Artificial Intelligence Review, 57(7), 188.

7. Probierz, B., Kozak, J., Piasecki, A., & Podlaszewska, A. (2023, September). Sign language interpreting-relationships between research in different areasoverview. In 2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS) (pp. 213-223). IEEE.

8. Oak, S., Shroff, T., Kulkarni, A., Jadhav, R., & Donkar, V. (2022, January). RETRACTED CHAPTER: Literature Review on Sign Language Generation. In the International Conference on Data Management, Analytics & Innovation (pp. 373-385). Singapore: Springer Nature Singapore.

9. Farzana, W., Sarker, F., Chau, T., & Mamun, K. A. (2021). Technological evolvement in AAC modalities to Foster communications of verbally challenged ASD children: A systematic review. IEEE Access.

10. Horton, L., & Singleton, J. (2022). Acquisition of turn-taking in sign language conversations: An overview of language modality and turn structure. Frontiers in Psychology, 13, 935342.

11. Chen, Y., Huang, Y., & Wu, H. (2021). A survey on sign language recognition and translation systems. Journal of Ambient Intelligence and Humanized Computing, 12(1), 1-18.

12. Zeng, Y., Chen, C., & Huang, Y. (2022). Enhancing sign language translation with multimodal learning. Sensors, 22(4), 1295.

13. Koller, O., & Ney, H. (2019). Real-time sign language recognition using hybrid deep learning models. IEEE Transactions on Neural Networks and Learning Systems, 31(5), 1398-1408.

14. Kim, S., Park, J., & Lee, S. (2020). An end-to-end system for continuous sign language recognition using attention-based models. Artificial Intelligence Review, 54(1), 67-85.

15. Alshahrani, M., & Al-Qurashi, A. (2023). Developing an intelligent sign language recognition system using convolutional neural networks. Journal of King Saud University - Computer and Information Sciences, 35(2), 295-305.

16. Buehler, C., Geller, L., & O'Reilly, M. (2021). Evaluating real-time sign language translation technologies: A systematic review. International Journal of HumanComputer Interaction, 37(3), 260-279.

17. Khamis, M., & Mowafi, M. (2022). Speech-to-sign language translation: Current technologies and future directions. Journal of Communication and Computer, 19(3), 23-29.

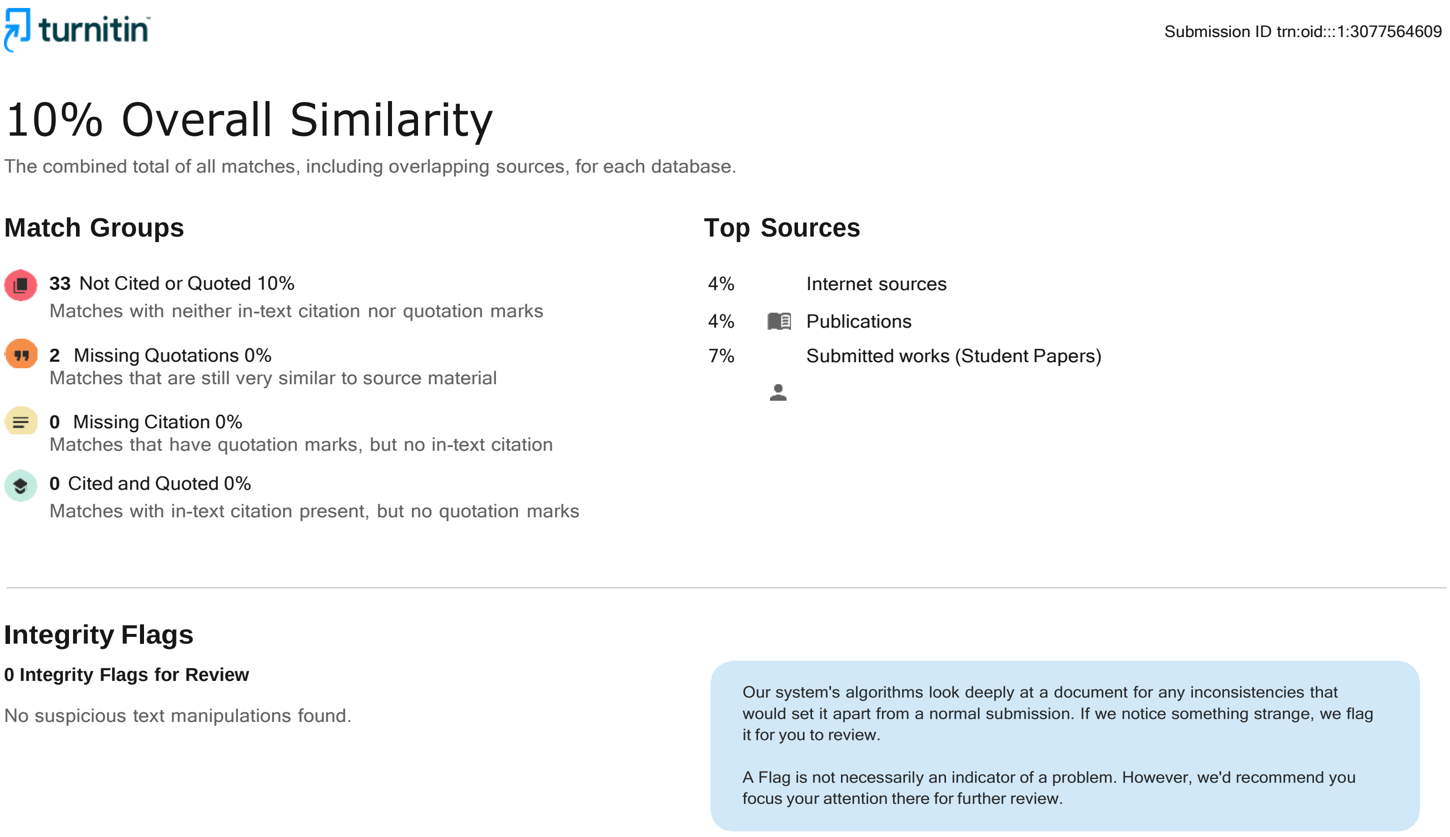
18. Liu, Y., & Zheng, H. (2021). Deep learning for sign language translation: A review. IEEE Access, 9, 76450-76465.

19. Alshahrani, M., & Al-Fagih, A. (2021). Automatic sign language recognition: A systematic review. Journal of Real-Time Image Processing, 18(4), 1091-1110.

20. Elakkiya, R., & Ezhil, A. (2022). A novel approach for sign language translation using deep learning models. Computational Intelligence and Neuroscience, 2022, 1-10.

APPENDIX 1

PLAGIARISM REPORT



turnitin

Submission ID trn:oid:::1:3077564609

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

33 Not Cited or Quoted 10%
Matches with neither in-text citation nor quotation marks

2 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

4%	Internet sources
4%	Publications
7%	Submitted works (Student Papers)

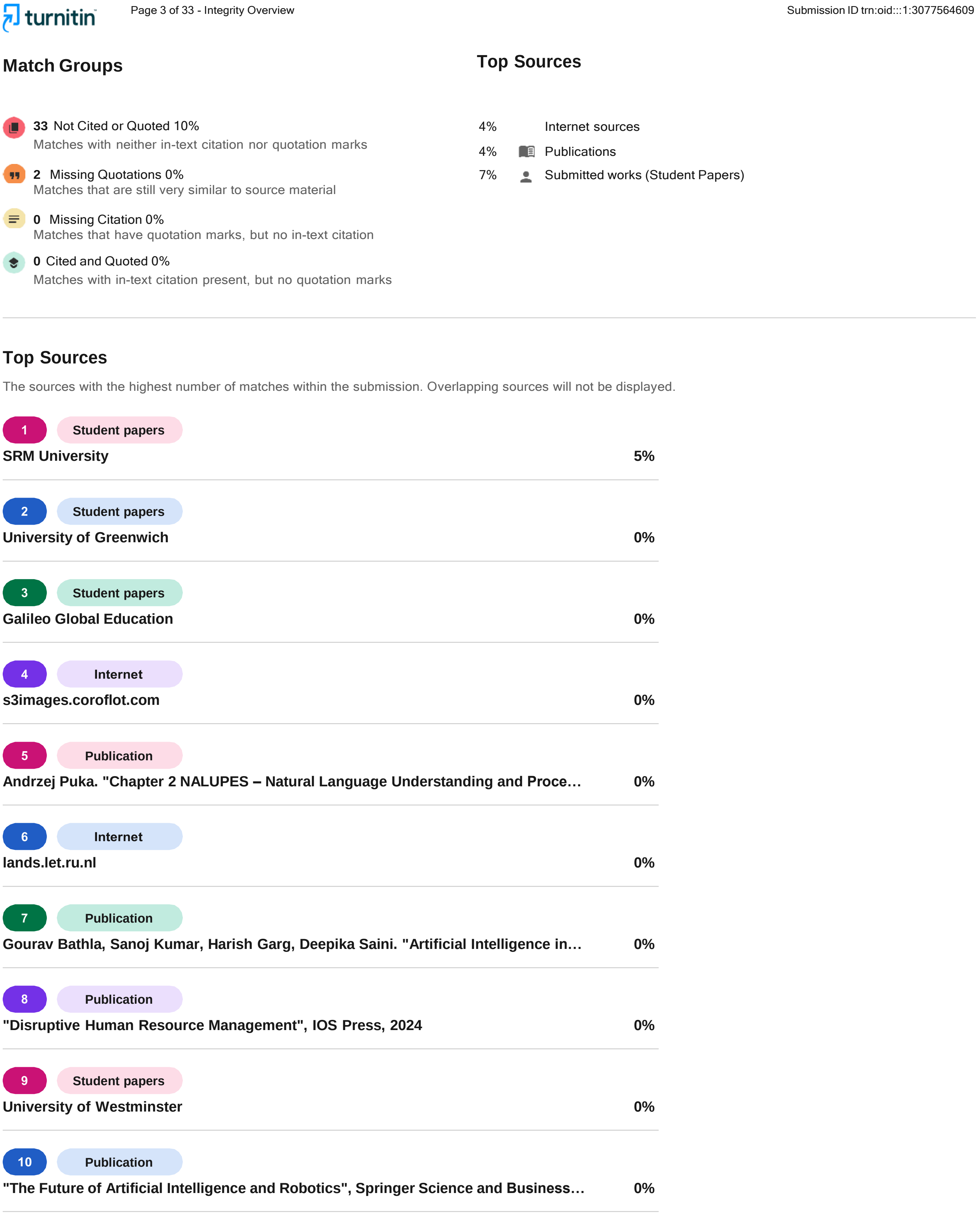
Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.



turnitin Page 3 of 33 - Integrity Overview Submission ID trn:oid:::1:3077564609

Match Groups

33 Not Cited or Quoted 10%
Matches with neither in-text citation nor quotation marks

2 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

4%	Internet sources
4%	Publications
7%	Submitted works (Student Papers)

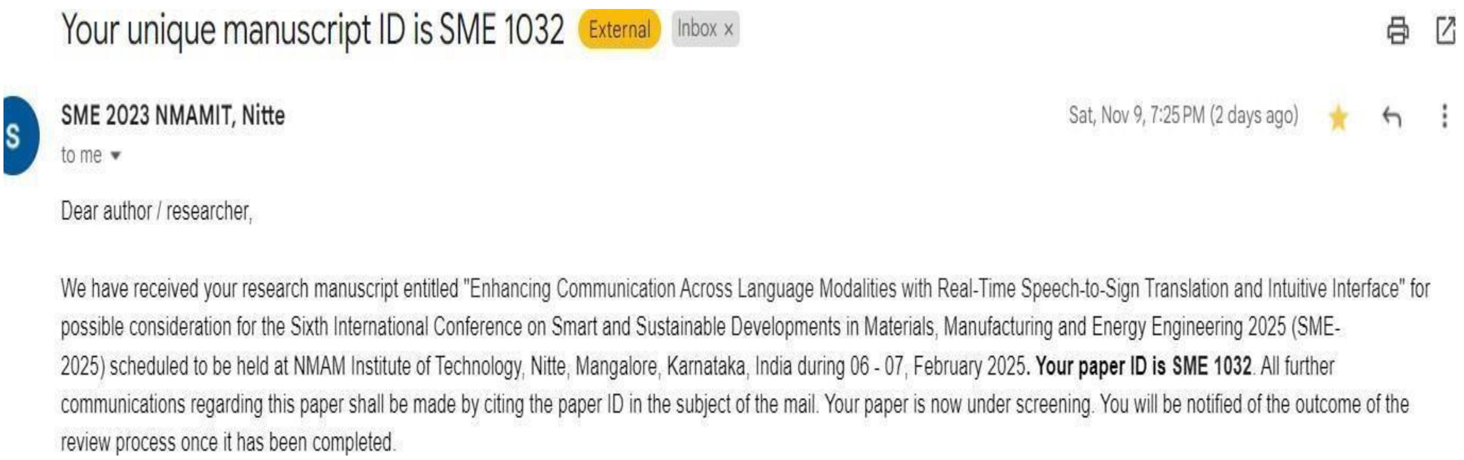
Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	SRM University	5%
2	Student papers	University of Greenwich	0%
3	Student papers	Galileo Global Education	0%
4	Internet	s3images.coroflot.com	0%
5	Publication	Andrzej Puka. "Chapter 2 NALUPES – Natural Language Understanding and Proce…	0%
6	Internet	lands.let.ru.nl	0%
7	Publication	Gourav Bathla, Sanoj Kumar, Harish Garg, Deepika Saini. "Artificial Intelligence in…	0%
8	Publication	"Disruptive Human Resource Management", IOS Press, 2024	0%
9	Student papers	University of Westminster	0%
10	Publication	"The Future of Artificial Intelligence and Robotics", Springer Science and Business…	0%

APPENDIX B

CONFERENCE PRESENTATION



Your unique manuscript ID is SME 1032

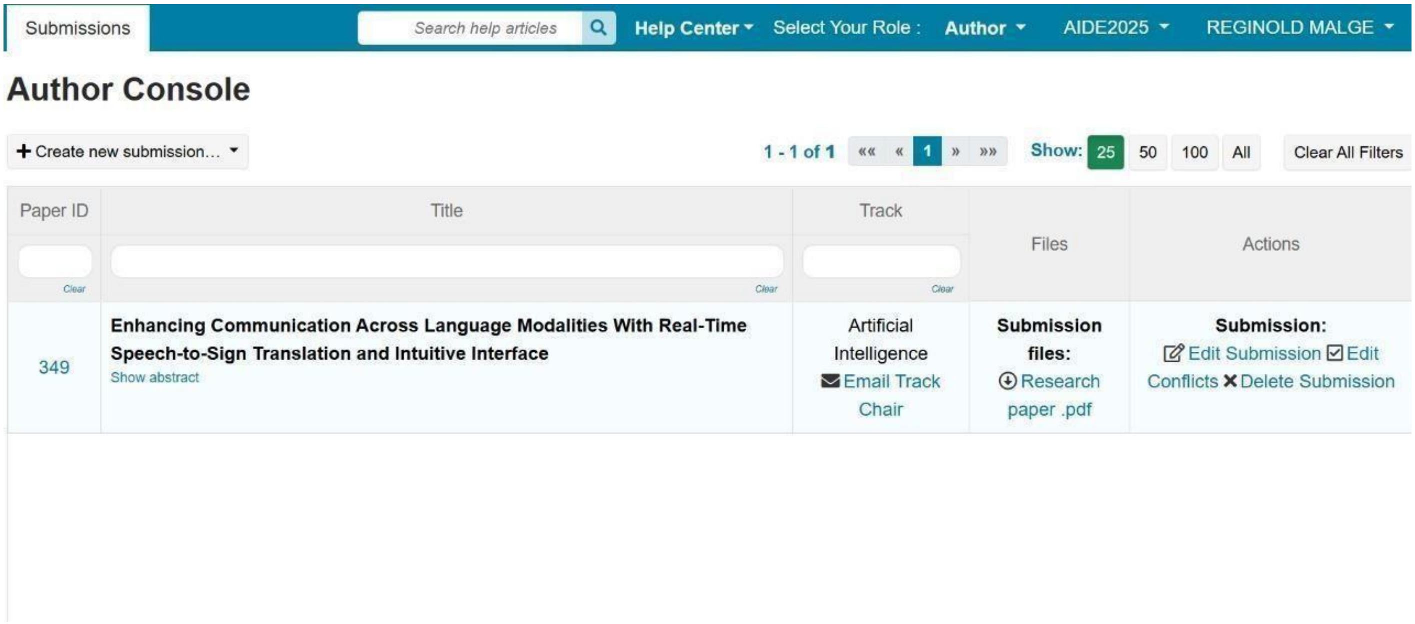
SME 2023 NMAMIT, Nitte

to me

Sat, Nov 9, 7:25 PM (2 days ago)

Dear author / researcher,

We have received your research manuscript entitled "Enhancing Communication Across Language Modalities with Real-Time Speech-to-Sign Translation and Intuitive Interface" for possible consideration for the Sixth International Conference on Smart and Sustainable Developments in Materials, Manufacturing and Energy Engineering 2025 (SME-2025) scheduled to be held at NMAM Institute of Technology, Nitte, Mangalore, Karnataka, India during 06 - 07, February 2025. **Your paper ID is SME 1032**. All further communications regarding this paper shall be made by citing the paper ID in the subject of the mail. Your paper is now under screening. You will be notified of the outcome of the review process once it has been completed.



Author Console

Paper ID	Title	Track	Files	Actions
349	**Enhancing Communication Across Language Modalities With Real-Time Speech-to-Sign Translation and Intuitive Interface** Show abstract	Artificial Intelligence ✉ Email Track Chair	**Submission files:** Research paper .pdf	**Submission:** Edit Submission ☑ Edit Conflicts ✕ Delete Submission