REALTIME TRANSLATION OF SPEECH TO ISL TO FACILITATE HEARING IMPAIRMENT

PROJECT REPORT

Submitted by

ALEENA B (PRN20CS018)
ANAGHA MARY PHILIP (PRN20CS023)
ARAVIND S K (PRN20CS030)
DEVIKA M (PRN20CS041)

Under the guidance of

Mr. MUHAMMED AZHARUDEEN SAHIB.S
Assistant Professor,
Department of Computer Science & Engineering
College of Engineering Perumon

To
the APJ Abdul Kalam Technological University, Kerala
in partial fulfillment of the requirements for the award of B.Tech Degree
in Computer Science and Engineering

Department of Computer Science and Engineering
College of Engineering Perumon
May 2024

COLLEGE OF ENGINEERING PERUMON
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

Certified that this report entitled **“Real Time Translation of Speech to ISL to Facilitate Hearing Impairment”** is the report of project phase1 presented by **ALEENA B (PRN20CS018), ARAVIND SK (PRN20CS030), DEVIKA M (PRN20CS041), ANAGHA MARY PHILIP (PRN20CS023)** during the year **(2023-2024)** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science & Engineering of the APJ Abdul Kalam Technological University, Kerala.

Muhammed Azharudeen Sahib S,
(Project Guide)
Assistant Professor,
Dept of Computer Science,
College of Engineering
Perumon.

Dr. Anoop S,
(Project Coordinator)
Assistant Professor,
Dept of Computer
Science,
College of Engineering
Perumon.

Dr. Varun Chand H,
(Project Coordinator)
Assistant Professor,
Dept of Computer
Science,
College of Engineering
Perumon.

Dr. K.S. Angel Viji,
(HOD)
Associate Professor &
HOD,Dept of Computer
Science, College of
Engineering
Perumon.

DECLARATION

We hereby certify that the work which is being presented in the Project Phase1 entitled **“Real Time Translation of Speech to ISL to Facilitate Hearing Impairment”** by **Aleena B, Anagha Mary Philip, Aravind SK and Devika M** in partial fulfillment of requirements for the award of degree of B. Tech in the Department of Computer Science & Engineering at **College of Engineering Perumon under APJ Abdul Kalam Technological University, Kerala** is an authentic record of our own work carried out during the period **2023-2024**. The matter presented in this project has not been submitted by us or anybody else in any other University / Institute for the award of B. Tech Degree.

Date: Signature:

ACKNOWLEDGEMENT

First and foremost, we sincerely thank the Almighty for his grace for the successful and timely completion of phase1 of our project **“Real Time Translation of Speech to ISL to Facilitate Hearing Impairment”**. We are greatly indebted to all those who helped us to make this project successful.

We are also grateful to **Dr. Rajeesh J, Principal, College of Engineering Perumon**, for providing us with the best facilities and atmosphere for our project development. We also wish to express our gratitude to **Dr. K S Angel Viji, Associate Professor & H.O.D, Department of Computer Science and Engineering**. We owe special thanks to our Project Guide **Mr. Muhammed Azharudeen Sahib S, Assistant Professor, Department of Computer Science and Engineering** and our Project coordinators **Dr. Anoop S, Assistant Professor, Department of Computer Science and Engineering** and **Dr.Varun Chand H, Assistant Professor, Department of Computer Science and Engineering**, for their corrections, valuable and countless suggestions, support and timely guidance.

Finally, but not least we would like to acknowledge our friends who were inevitable for the successful completion of this project.

ABSTRACT

Sign language is a visual language utilized by individuals who are deaf as their primary means of communication. Unlike spoken languages, sign language relies on gestures, body movements, and manual communication to effectively convey ideas and thoughts. It can be used by individuals who have difficulty speaking, those who are unable to speak, and by individuals without hearing impairments to communicate with deaf individuals. Access to sign language is crucial for the social, emotional, and linguistic development of deaf individuals. Our project aims to bridge the communication gap between deaf individuals and the general population by leveraging advancements in web applications, machine learning, and natural language processing technologies. The primary objective of this project is to develop an interface capable of converting audio/voice inputs into corresponding sign language for deaf individuals. This is achieved through the simultaneous integration of hand shapes, orientations, and movements of the hands, arms, or body. The interface operates in two phases: first, converting audio to text using speech-to-text APIs (such as Python modules or Google API); and second, representing the text using parse trees and applying the semantics of natural language processing (specifically, NLTK) for the lexical analysis of sign language grammar. This work adheres to the rules of Indian Sign Language (ISL) and follows ISL grammar guidelines.

CONTENTS

LIST OF FIGURES

ABBREVIATIONS

NMT:	Neural Machine Translation
MG:	Motion Graph
PCFG :	Probabilistic Context Free Grammar
ISL :	Indian Sign Language
LSTM:	Long Short-Te
GMM:	Gaussian Mixture Model
GIF:	Graphics Interchange Format
RNN:	Recurrent Neural Network
Bi-LSTM :	Bidirectional LSTM
GRU :	Gated Recurrent Unit
SiGML:	Signing Gesture Markup Language
WebGL:	Web Graphics Library
BLEU:	Bilingual Evaluation Understudy
GUI:	Graphical user interface
TC :	Total Communication
NLP :	Natural Language Processing
NLTK :	Natural Language Toolkit

CHAPTER 1
INTRODUCTION

Hearing loss has emerged as the predominant sensory deficit afflicting humanity today. According to recent reports from the World Health Organization, the impact of this issue is particularly pronounced in India, where an alarming approximately 63 million individuals grapple with hearing impairment. This staggering figure represents approximately 6.3% of the entire population of the country. Disturbingly, further investigations into this matter have revealed that a substantial majority of this affected population falls within the age range of 0 to 14 years.

The consequences of such a substantial number of young individuals experiencing hearing impairment are far-reaching, resulting in a significant loss in productivity on both economic and physical fronts. The economic implications are evident as these youths, facing communication challenges due to their reliance on sign language, encounter obstacles that hinder their full participation in various spheres of life. This, in turn, contributes to a broader societal issue, where a considerable portion of the youth population feels confined, as if trapped within a cage, and sadly, often experiences a sense of neglect from society.

A critical aspect exacerbating this situation is the general lack of awareness prevalent among the masses regarding sign language. The inability of the majority of ordinary individuals to comprehend and utilize sign language further isolates those with hearing impairment, hindering their ability to communicate effectively. Consequently, this lack of understanding perpetuates the social isolation of individuals with hearing loss, accentuating the need for comprehensive awareness campaigns and education initiatives to bridge this communication gap and foster inclusivity in society.

Sign languages, much like their spoken counterparts, boast distinct grammatical structures and rules. The intricate task of translating between spoken and signed languages requires the deployment of machine translation methods that meticulously consider the unique linguistic nuances inherent in both modalities. Facilitating effective communication between the hearing and the Deaf mandates the development of robust systems proficient in translating spoken languages into sign languages and vice versa. In this multifaceted process, Sign Language Recognition (SLR) and Sign Language Production (SLP) emerge as pivotal components. In the realm of commercial applications, there is a predominant focus on SLR, entailing the provision of text transcriptions for signed sequences. However, this paradigm operates under the assumption that deaf individuals are inherently comfortable with reading spoken language, a presumption that may not universally hold true. The intricate challenge of generating sign language from spoken language is compounded by the asynchronous channels characteristic of sign languages, encompassing both manual and non-manual features. Presently, SLP methodologies predominantly incorporate animated avatars or the translation of spoken language into sign glosses tethered to a parametric representation. These approaches, while notable, grapple with limitations, particularly in terms of the potential loss of context and meaning conveyed by non-manual features. In a stride toward advancing Sign Language Production, the aim of the project is to develop a web application interface that facilitates communication between deaf individuals and the general population by converting audio/voice inputs into corresponding sign language. This is achieved by leveraging advancements in web applications, machine learning, and natural language processing technologies. The primary objective is to bridge the communication gap by providing a tool that interprets spoken language into sign language in real-time. The project adheres to the rules of Indian Sign Language (ISL) and follows ISL grammar guidelines to ensure accurate representation and effective communication for deaf individuals.

1.1 PROBLEM STATEMENT

- **Communication Barriers:** Deaf and mute individuals face significant challenges in communicating with the general population due to the lack of familiarity with Indian Sign Languages (ISL). Speech to sign language technology addresses this barrier, enabling more effective communication.

- **Limited Employment Opportunities:** The absence of accessible communication methods contributes to the under representation of the hearing-impaired in the workforce. Speech to ISL can enhance employment opportunities by facilitating communication and breaking down barriers in professional settings.

- **Educational Accessibility:** Deaf students encounter difficulties in traditional educational settings where spoken language is predominant. Speech to ISL solutions can bridge this gap, making educational content more accessible and fostering an inclusive learning environment.

- **Social Isolation:** Deaf and mute individuals often experience social isolation as a result of communication challenges. Speech to ISL can help mitigate this isolation by enabling better communication with the broader community and reducing feelings of exclusion.

- **Emergency Situations:** Quick and effective communication is crucial in emergencies. Speech to ISL aids in conveying vital information to the hearing-impaired during critical situations, ensuring their safety and well-being. Underrepresentation in Technology: The majority of existing sign language technologies are developed for widely used sign languages like American Sign Language (ASL). Speech to ISL for Indian languages addresses the underrepresentation of regional sign languages and supports the technological needs of the local population.

1.2 OBJECTIVES

- Bridge Communication Gap: Facilitate communication between deaf individuals and the general population.
- Leverage Technology: Utilize advancements in web applications, machine learning, and natural language processing.
- Develop Conversion Interface: Create an interface that translates audio inputs into sign language.
- Integrate Hand Gestures and Movements: Ensure accurate representation by integrating hand shapes, orientations, and movements.
- Operate in Two Phases: Convert audio to text using speech-to-text APIs, then represent text in sign language using parse trees and natural language processing.

- Adhere to ISL Guidelines: Follow Indian Sign Language (ISL) rules and grammar guidelines for linguistic accuracy and cultural appropriateness.

1.3 ORGANIZATION OF REPORT

In Chapter II, we examine existing related work. We picked papers related to speech to sign translation. Chapter III, we introduce our methodology system requirements both software and hardware requirements. In Chapter IV, we describe the implementation of our system. In Chapter V we provide the results and discussions. In Chapter VI, we provide the conclusion of our project.

CHAPTER 2
LITERATURE REVIEW

The literature survey delves into an innovative project aimed at revolutionizing accessibility by seamlessly integrating speech recognition, text conversion, and Indian Sign Language (ISL) animation. At its core, the project harnesses the power of the JavaScript Web Speech API to swiftly and accurately transcribe live audio speech recordings into textual representations. This initial step, driven by sophisticated algorithms and neural network models, ensures not only high accuracy but also adaptability to diverse speech patterns and accents, thereby enhancing accessibility for a wide range of users.

Following the conversion of speech to text, the application embarks on a meticulous journey of text preprocessing using the esteemed Natural Language Toolkit (NLTK). Renowned for its proficiency in language processing tasks, NLTK undertakes vital tasks such as tokenization, stemming, and the removal of stop words. By ensuring the cleanliness and refinement of the input text, NLTK sets the stage for subsequent processing steps, thereby enhancing the overall efficiency and accuracy of the application. Moreover, advanced techniques such as named entity recognition and part-of-speech tagging further enrich the textual data, enabling more nuanced interpretation and analysis across various linguistic domains.

Central to the project's vision is the generation of relevant ISL animations based on the meticulously processed textual input. Through seamless mapping of text to appropriate ISL gestures and signs, the application brings to life a meticulously crafted 3D animation character, painstakingly designed using Blender 3D. This character serves as a conduit for conveying the intricacies of sign language, faithfully translating input text into expressive and meaningful gestures. Through continuous refinement and optimization, the animations achieve a level of realism and fluidity that enhances user engagement and comprehension. Furthermore, the system leverages deep learning algorithms to adaptively adjust animation styles based on user feedback and preferences, ensuring personalized and engaging interactions.

Technologically, the project leverages a diverse array of tools and frameworks to realize its ambitious goals. HTML, CSS, and JavaScript form the backbone of the web-based application, providing the necessary infrastructure for seamless user interaction. Python, with its versatility and robust standard library, serves as the primary programming language, while the Django framework simplifies the creation of complex, database-driven websites. Additionally, the NLTK library plays a pivotal role in natural language processing tasks, facilitating word tokenization, elimination of stop words, lemmatization, and synonym replacement. These functionalities are essential for maintaining consistency and accuracy in language processing tasks, particularly in the context of ISL translation.

The proposed system aims to develop improved facilities by utilizing audio input on a Personal Digital Assistant through the Python PyAudio module, converting it to text using the Google Speech API with NLP, analyzing the text using a Reliance parser for grammatical structure, and finally converting the audio into Sign language. This system overcomes limitations of the existing system, offers simplicity in design and implementation, requires minimal system resources, and works across various configurations. Its features include providing information access and services in Indian Sign Language, scalability to capture the entire ISL vocabulary, and enhancing the physical and mental well-being of specially-abled individuals, thereby improving their quality of life.

In summary, the project represents a pioneering endeavor in the realm of accessibility technology, seamlessly bridging the communication gap between spoken language and sign language with finesse and innovation. Through continuous updates and enhancements, it strives to push the boundaries of accessibility technology, empowering users with new ways to communicate and interact effectively while ensuring inclusivity and accessibility for all.

2.1 Conversion of native speech into Indian sign language to facilitate hearing impairment [1]

A substantial portion of the global population comprises individuals with specific needs, and in the contemporary era, communication stands out as a paramount aspect of human interaction. Indian Sign Language (ISL) plays a pivotal role in addressing the communication requirements of individuals who are both deaf and mute, offering them a means to engage in meaningful conversations not only within the Indian subcontinent but also enabling communication with individuals who do not use sign language.

Indian Sign Language is a rich and expressive mode of communication, relying on facial expressions, body postures, and hand gestures to convey messages and emotions effectively. It is characterized by lexicons and grammar rules that govern its behavior, elevating it to the status of a fully-fledged natural language. This review endeavors to conceptualize an application that serves as a bridge between spoken language and Indian Sign Language, translating captured speech into a visual representation of ISL.The proposed application leverages Speech Recognition technology to process voice input, transforming it into equivalent text. Subsequently, Natural Language Processing algorithms are employed for word segmentation and root word extraction, ensuring a comprehensive understanding of the linguistic context. The translated text is then converted into its equivalent Indian Sign Language, allowing seamless communication for the deaf and mute community.

A crucial aim of this project is to make the translation accessible to a broader audience in India. By catering to the unique linguistic and expressive features of Indian Sign Language, the application seeks to enhance inclusivity and communication accessibility for individuals with specific needs, fostering a more connected and understanding society. As the project progresses, a commitment to continuous improvement and outreach initiatives will contribute to its effectiveness and broader societal impact.

2.1.1 Methodology:

1. Speech is converted into text using Google's Speech-To-Text.
2. Translation: The obtained text is translated into English using Google Translate.
3. Backend Processing: English text is sent to the backend server via a local host API. Text is streamed to the Stanford Parser for linguistic analysis.
4. Linguistic Analysis: Stanford Parser uses Probabilistic Context Free Grammar (PCFG) and English models for parsing. Relations between words are identified, and the sentence is rephrased according to ISL grammar rules.

5. Eliminating Stop Words: Unwanted or stop words are removed, including determiners, non-root verbs, foreign words, and interjections.
6. Lemmatization and Synonym Replacement:Words are lemmatized to obtain root forms, as ISL does not accept gerunds. Synonyms are tagged if the word is not present in the bilingual dictionary
7. Video Conversion/Mapping:ISL-transformed text is compared with a dataset of videos. Basic string-matching algorithms are used for similarity between video labels and the preprocessed input text. The input text is mapped to videos or video labels, and a sequence of videos is played in order on the display screen.

2.1.1 Advantages

- Multilingual Support: The system accommodates multiple languages, promoting inclusivity.
- Speech-to-Text and Translation: Enables seamless conversion from spoken language to text and subsequent translation.
- Rule-Based Conversion: Adheres to established rules, ensuring accurate and consistent language conversion.
- Elimination of Stop Words: Enhances efficiency by excluding common, non-informative words from translation.
- Lemmatization and Synonym Replacement: Improves accuracy through lemmatization and synonym use in language processing.
- Video Mapping: Provides a visual representation, enhancing understanding through video sequences.
- User-Friendly Interface: Enhances accessibility and ease of use for a diverse user base.
- Dynamic Video Sequence: Offers a dynamic and engaging video representation for effective communication.
- Adaptability to ISL Grammar: Aligns with the grammatical rules of Indian Sign Language, ensuring linguistic accuracy.

2.1.1 Disadvantages

- Dependency on External APIs: Reliance on external APIs may introduce vulnerabilities and impact system reliability.
- Limited ISL Vocabulary: The system's vocabulary may be constrained, limiting its ability to accurately represent all ISL expressions.
- Dynamic Nature of Language: The ever-evolving nature of language may pose challenges in keeping the system up-to-date with current linguistic trends.
- User Training and Adaptation: Users may require training to adapt to system nuances, potentially impacting user adoption.
- Processing Overhead: Intensive processing requirements may affect system performance, especially with large datasets.
- Limited Customization for Users: Lack of extensive customization options may limit personalization for individual users.
- Ethical Considerations: The use of language and video data raises ethical concerns, necessitating responsible handling and privacy measures.

2.2 Speech to Sign Language Translation for Indian Languages[2]

Individuals grappling with hearing impairments and those who are mute often encounter substantial challenges in communication, significantly impacting their self-confidence and giving rise to feelings of isolation within society. Sign language serves as a critical bridge between the deaf community and the general population; however, a notable research gap persists, particularly in the realm of Indian Sign Language, which is extensively utilized in India. This research seeks to address this gap by undertaking a comprehensive exploration of speech-to-sign-language conversion, specifically focusing on six diverse Indian regional languages—Telugu, Hindi, Malayalam, Marathi, Kannada, and Tamil. The proposed model, designed with the overarching goal of enhancing communication accessibility for the hearing-impaired, entails a multi-faceted process. It integrates advanced technologies, including the utilization of Wavelet-based Mel Frequency Cepstral Coefficients (MFCC) with Gaussian Mixture Models (GMM) for robust speech recognition. Additionally, Long

Short-Term Memory (LSTM) networks are deployed for efficient text translation, ensuring the preservation of contextual nuances. A pivotal aspect of the model involves a meticulously designed mapping mechanism that correlates the translated text with the corresponding gestures in Indian Sign Language. To safeguard cultural relevance and authenticity, the research places a significant emphasis on community engagement and feedback. This collaborative approach not only considers regional variations and linguistic nuances within the chosen languages but also ensures that the proposed model resonates authentically with the diverse communication styles prevalent in each linguistic community. By extending the scope of understanding across diverse linguistic landscapes, the ultimate aim is to empower the hearing-impaired community, fostering not only effective communication but also promoting broader social inclusion in the fabric of society.

2.2.1 Methodology

1. Data Collection: Gather a diverse dataset comprising speech samples in Telugu, Hindi, Tamil, Malayalam, Kannada, and Marathi. Collect corresponding Indian Sign Language (ISL) gestures for each spoken phrase to create a parallel dataset.
2. Preprocessing: Preprocess the speech data by extracting features using wavelet-based Mel-Frequency Cepstral Coefficients (MFCC).Normalize and standardize the data to ensure consistency across languages.
3. Speech Recognition: Employ Gaussian Mixture Model (GMM) for speech recognition, as it has shown superiority in recognition applications. Train language-specific models for Telugu, Hindi, Tamil, Malayalam, Kannada, and Marathi.
4. Text Translation: Utilize an Encoder-Decoder based Long Short-Term Memory (LSTM) network for text translation. Train the model to translate the recognized speech text into the corresponding text in each language
5. Sign Language Generation: Develop a system for generating Indian Sign Language (ISL) gestures based on the translated text. Implement language-specific rules and nuances for accurate ISL representation.
6. Evaluation: Evaluate the performance of the system using appropriate metrics such as

7. accuracy, precision, and recall. Conduct user studies with deaf and mute individuals to assess the effectiveness of the ISL generation.

2.2.1 Advantages

- Diverse Dataset: Varied speech samples and corresponding ISL gestures enhance model robustness.
- Parallel Dataset: Matched speech-ISL pairs enable effective supervised training for accurate translation.
- Wavelet-Based MFCC: Extracting features using wavelet-based MFCC captures comprehensive speech characteristics.
- Language-Specific Models: GMM-based speech recognition with language-specific models optimizes performance.
- Encoder-Decoder LSTM: LSTM network for text translation ensures contextually relevant and accurate translations.
- Language-Specific Sign Language: Implementing language-specific rules enhances the accuracy of ISL gesture generation.
- User Studies: Real-world user feedback through studies aids in refining and aligning the system with user needs.

2.2.1 Disadvantages

- Limited Dataset Size: Small dataset size may impact the model's ability to generalize.
- Data Bias: Lack of full diversity may introduce bias in the model's performance.
- Complexity of Sign Language: Representing all sign language nuances accurately is challenging.
- Resource-Intensive Training: Training language-specific models can be computationally intensive.
- Speech Recognition Accuracy: GMM-based recognition may have limitations in capturing intricate speech variations.

- Limited Generalization: Effectiveness may be constrained to phrases present in the training dataset.
- User Study Limitations: Small sample size and specific demographics may limit the generalizability of user study findings.

2.3 Creation of GIF dataset and implementation of a speech-to-sign language translator in Telugu [3]

The paper introduces a pioneering model designed to address the vital task of translating real-time voice inputs or audio recordings in Telugu, a prominent native Indian regional language, into textual representations. This innovative approach aims to bridge the communication gap for individuals using sign language as their primary means of expression. The text undergoes a sophisticated matching process with a pre-defined Graphics Interchange Format (GIF) dataset containing sign language animations. Leveraging the Google API, the model seamlessly executes the speech-to-text conversion, demonstrating its integration with state-of-the-art technology.In scenarios where the converted text aligns with entries in the GIF dataset, the model exhibits its capability to dynamically generate the corresponding sign language GIF, enriching the communication experience. Furthermore, the implementation incorporates a user-friendly graphical interface designed to enhance accessibility. This interface systematically displays Indian Sign Language (ISL) text in alphabetical order when the input falls outside the scope of the GIF dataset, providing an inclusive user experience.

The paper extends its contribution by conducting a meticulous comparative evaluation of the performance of four powerful deep learning networks—namely, RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory), Bi-LSTM (Bidirectional LSTM), and GRU (Gated Recurrent Unit). These networks undergo rigorous training using a comprehensive Telugu multi-speaker speech dataset for the critical task of speech recognition. The reported outcomes underscore the effectiveness of the proposed models, with accuracies reaching 51% for RNN, 68% for LSTM, an impressive 90% for Bi-LSTM, and a commendable 88%

for GRU.In the comprehensive evaluation, the Google API-based speech-to-sign-language conversion emerges as the most promising solution, demonstrating remarkable consistency with precision and recall scores both at 94%, along with a notably low word error rate of 4.1%. This robust and precise performance highlights the potential of the proposed model in facilitating effective communication for individuals relying on sign language. The research not only contributes to the advancement of technology in sign language translation but also sheds light on the strengths and areas of improvement within different deep learning architectures, paving the way for future research in this impactful domain.

2.3.1 Methodology

1. Audio Input Processing: Capturing live audio speech using a microphone and preprocessing it through four different approaches:

- Pre-emphasis: Emphasizing higher frequencies in the input speech to enhance important details.
- Frame Blocking: Breaking down the input speech into segments for more manageable analysis.
- Windowing: Multiplying each frame by a window function (like Hamming) to reduce signal distortion at frame edges.
- Spectral Analysis: Analyzing the spectrum of each frame to understand the distribution of frequencies over time

2. Datasets Used: Crowdsourced high-quality multi-speaker speech dataset for training deep learning models.Tab-delimited bilingual sentence pairs dataset for text translation.Custom-made in-house GIF dataset with 50 common sentences, A-Z signs, and 0-9 numbers.
3. Speech Recognition: Extracting Mel-Frequency Cepstral Coefficients (MFCC) features from preprocessed speech.
   MFCC Features: A set of features derived from the spectrum of a sound signal.
   Models Compared: RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory), Bi-LSTM (Bidirectional LSTM), GRU (Gated Recurrent Unit).

WER (Word Error Rate): Used to measure the accuracy and performance of the speech-to-text conversion models.

4. Speech Recognition: To enhance accuracy and performance, the built-in Google API for speech recognition is employed.The Google API's performance is compared with the deep learning models to identify the most effective approach.The models and Google API process the input Telugu audio and produce the corresponding text output.
5. Text Translation:Employing both LSTM encoder-decoder and Google Translate API for translating Telugu to English text.Tokenization, converting text to lowercase, stemming, and lemmatization are applied to prepare the text data for translation.LSTM encoder processes the Telugu sentence, and the decoder generates the corresponding English sentence.Using the preprocessed text data to map appropriate GIFs from the dataset.
6. Sign Detection:Creating a dataset of 50 common GIFs for Telugu.Mapping processed text data to appropriate GIFs.Displaying relevant GIFs for matching voice commands, showing ISL alphabetical representation if no match.

2.3.2 Advantages

- Live Audio Processing: Enables real-time conversion of live audio into text, providing immediate communication support.
- Diverse Datasets: Utilizes crowdsourced speech datasets, bilingual sentence pairs, and a custom GIF dataset, enhancing model robustness.
- Preprocessing Techniques: Incorporates frame-blocking, windowing, and spectral analysis for effective speech signal processing.
- Deep Learning Models: Employs powerful deep learning models (RNN, LSTM, Bi-LSTM, GRU) for speech recognition, offering diverse options for model selection.
- Google API Integration: Enhances accuracy with the integration of Google API for speech recognition, providing a benchmark for comparison.
- Text Translation: Utilizes both LSTM encoder-decoder and Google Translate API, enabling multilingual support for effective communication.

- Sign Detection and Mapping: Maps processed text data to relevant GIFs, enhancing user experience by providing visual representations of sign language.

2.3.3 Disadvantages

- Dependency on Google API: Reliance on external APIs may introduce vulnerabilities and impact system reliability, especially if there are service disruptions.
- Limited ISL Vocabulary: The custom GIF dataset may have limitations in representing the full spectrum of Indian Sign Language, impacting translation accuracy.
- Complex Preprocessing: The combination of preprocessing techniques may introduce complexity, requiring careful calibration to avoid signal distortion.
- Model Comparison: While deep learning models are compared, the choice of the "most effective" approach may vary based on specific use cases and requirements.
- Multilingual Translation Challenges: Translation accuracy may vary for different languages, potentially posing challenges in maintaining consistent performance across diverse linguistic contexts.
- Resource Intensiveness: Deep learning model training can be computationally intensive, potentially limiting scalability and deployment on resource-constrained devices.
- User Interface Complexity: The display of GIFs and ISL alphabetical representation adds complexity to the user interface, requiring user adaptation and training.

2.4 Indian Sign Language Generation System [4]

At the core of our approach lies the integration of two widely recognized standards—HamNoSys and SiGML. HamNoSys, a linguistic notation system tailored explicitly for sign languages, offers a structured framework for encoding the intricate linguistic elements embedded within signs. Complemented by SiGML, a standardized markup language designed for the digital representation and exchange of sign language data, our system endeavours to establish a robust foundation for the representation and dissemination of ISL content.The user-centric design of our ISL generation system prioritizes usability and accessibility, providing an intuitive interface for users to input textual content, which the system dynamically translates into ISL animations. This process entails meticulous encoding of linguistic parameters using HamNoSys, ensuring the accuracy and cultural appropriateness of the generated sign language gestures.Through this project, we aim not only to develop a technologically advanced ISL generation system but also to advocate for the broader discourse on assistive technologies. The subsequent sections of this report delve into the detailed methodology, implementation strategies, and evaluation mechanisms employed in the development of our ISL generation system, with a focus on its potential to foster inclusivity and accessibility for the diverse deaf and hard-of-hearing community in India.

At the core of our approach lies the integration of two widely recognized standards—HamNoSys and SiGML. HamNoSys, a linguistic notation system tailored explicitly for sign languages, offers a structured framework for encoding the intricate linguistic elements embedded within signs. Complemented by SiGML, a standardized markup language designed for the digital representation and exchange of sign language data, our system endeavours to establish a robust foundation for the representation and

dissemination of ISL content.The user-centric design of our ISL generation system prioritizes usability and accessibility, providing an intuitive interface for users to input textual content, which the system dynamically translates into ISL animations. This process entails meticulous

encoding of linguistic parameters using HamNoSys, ensuring the accuracy and cultural appropriateness of the generated sign language gestures.Through this project, we aim not only to develop a technologically advanced ISL generation system but also to advocate for the broader discourse on assistive technologies. The subsequent sections of this report delve into the detailed methodology, implementation strategies, and evaluation mechanisms employed in the development of our ISL generation system, with a focus on its potential to foster inclusivity and accessibility for the diverse deaf and hard-of-hearing community in India.

2.4.1 Methodology

1. Data Collection and Corpus Building:

• Collect a diverse dataset of Indian Sign Language gestures, considering regional variations.

• Annotate the dataset with linguistic features using HamNoSys notation.

• Include a variety of expressions, emotions, and contextual variations to ensure a comprehensive training dataset.

2. Preprocessing:

• Clean and preprocess the linguistic data, ensuring uniformity and compatibility with HamNoSys notation.

• Transform the annotated linguistic features into SiGML format for sign language representation.

3. Separate parsers:

• The system consists of separate parsers to handle Hindi and English sentences. These parsers analyze the input sentences and extract the necessary information for generating sign animations.

4. ISL grammar rules:

• The system implements Indian Sign Language (ISL) grammar rules. These rules define the structure and syntax of ISL and guide the generation of sign animations based on the input sentences.

5. Real-time Translation Integration:

• Implement real-time translation capabilities, allowing the system to generate sign language animations as input (speech or text) is provided.

• Optimize for low latency to ensure smooth and instantaneous communication.SiGML to Animation Conversion:

• Develop a module to convert SiGML representations into animated sign language gestures.

• Integrate this module into the overall system architecture for seamless generation of sign language animations.

6. WebGL technology:

• To produce the 3D animation of sign language, the researchers utilized WebGL technology. WebGL integrates HamNoSys and SiGML, which are used for storing and retrieving signs.

7. Testing and evaluation:

• The system's robustness was assessed through testing with the help of manual evaluations by sign language experts and BLEU scores. The BLEU score measures the similarity between the generated sign animations and the reference animations.

2.4.2 Advantages

- Diverse Dataset: Inclusion of a diverse dataset with regional variations, expressions, emotions, and contextual variations ensures a comprehensive training dataset, enhancing the system's adaptability.
- Linguistic Annotation: Annotating the dataset with HamNoSys notation adds linguistic features, promoting a standardized and structured representation of Indian Sign Language (ISL).
- SiGML Format: Transforming linguistic features into SiGML format allows for a standardized representation of sign language, facilitating compatibility and interoperability.
- Separate Parsers: Having separate parsers for Hindi and English sentences enhances language-specific analysis, optimizing the system's accuracy in generating sign animations for diverse linguistic inputs.
- ISL Grammar Rules: Implementation of ISL grammar rules guides the generation of sign animations, ensuring syntactic and structural accuracy in conveying linguistic content.
- Real-time Translation: Real-time translation capabilities enable immediate generation of sign language animations as input (speech or text) is provided, promoting seamless and efficient communication.

- Low Latency Optimization: Optimization for low latency ensures quick and smooth communication, contributing to a responsive and user-friendly experience.
- WebGL Technology: Utilization of WebGL technology for 3D animation production integrates HamNoSys and SiGML, enhancing the visual quality and realism of generated sign language animations.
- Testing and Evaluation: Rigorous testing, including manual evaluations by sign language experts and BLEU scores, ensures robustness and provides quantitative measures of similarity between generated and reference animations.

2.4.3 Disadvantages

- Dependency on WebGL: Reliance on WebGL technology may introduce limitations in terms of device compatibility and performance, potentially excluding users with certain devices.
- Complexity of SiGML to Animation Conversion: The process of converting SiGML representations into animated sign language gestures may introduce complexity and potential challenges in maintaining accuracy and fluency.
- Resource Intensiveness: Real-time translation and 3D animation production may require significant computational resources, impacting the system's scalability and accessibility.
- BLEU Score Limitations: While BLEU scores provide a quantitative measure, they may not capture the full spectrum of sign language quality and expression, potentially overlooking nuanced aspects of animation evaluation.
- Expert Dependency: Manual evaluations by sign language experts introduce subjectivity and dependency on human judgment, which may vary among different experts.
- Limited Information on Testing Methodology: The description lacks detailed information on the specific testing methodology employed, making it challenging to assess the comprehensiveness of the evaluations conducted.

2.5 Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks [5]

In this groundbreaking study, we introduce an innovative and state-of-the-art method for automated Sign Language Production, harnessing the latest advancements in Neural Machine Translation (NMT), Generative Adversarial Networks (GANs), and motion generation technologies. Our cutting-edge system exhibits exceptional proficiency in generating expressive sign language videos seamlessly derived from spoken language sentences. Diverging from prevailing approaches that heavily rely on extensively annotated data, our method distinguishes itself by demanding minimal gloss and skeletal level annotations during the training phase. This notable accomplishment is attributed to a meticulous and strategic breakdown of the task into dedicated sub-processes, showcasing the efficiency and resourcefulness of our approach.The core process involves the initial translation of spoken language sentences into intricate sign pose sequences. This transformative step is achieved through a sophisticated integration of an NMT network with a Motion Graph, emphasizing the synergy between language translation and motion representation. The resulting pose information, rich in linguistic and gestural nuances, serves as a crucial conditioning factor for a sophisticated generative model. This model, at the heart of our approach, excels in producing not only accurate but also photorealistic sign language video sequences. Notably, our methodology marks a pioneering advancement as the first continuous sign video generation approach that boldly moves away from the conventional reliance on graphical avatars.To assess and validate the translation capabilities of our approach, we conduct a rigorous evaluation on the PHOENIX14T Sign Language Translation dataset. This evaluation sets a robust baseline for text-to-gloss translation, demonstrating our method's prowess with a noteworthy BLEU-4 score of 16.34/15.26 on dev/test sets. Furthermore, we extend our exploration to showcase the impressive video generation capabilities of our approach in both multi-signer and high-definition settings. This thorough demonstration encompasses qualitative assessments, offering insights into the expressive and contextual richness of the generated sign language videos, and quantitative analyses, incorporating broadcast quality assessment metrics to quantify the system's performance metrics comprehensively.

In summary, our extended abstract highlights the sophistication, versatility, and groundbreaking nature of our proposed approach, emphasizing its potential to revolutionize the landscape of Sign Language Production through the fusion of NMT, GANs, and motion generation, while setting a robust benchmark in translation accuracy and video quality evaluation.

2.5.1 Methodology

1. Neural Machine Translation (NMT) Network: Develop an NMT network to translate spoken language sentences into sign language gloss sequences. Train the NMT network using minimal gloss and skeletal level annotations, breaking down the translation task into dedicated sub-processes

2. Data Collection and Preprocessing: Gather a diverse dataset of spoken language sentences paired with corresponding sign language videos. Preprocess the data to extract linguistic features, such as text sequences and corresponding pose annotations.

3. Motion Graph (MG) Integration: Combine the NMT network with a Motion Graph (MG) to obtain a sequence of pose probabilities representing the sign language gestures. Utilize the MG to capture continuous-text-to-pose translation, allowing for more natural and fluid sign language generation.

4. Generative Adversarial Network (GAN) Conditioning: Condition a Generative Adversarial Network (GAN) using the obtained pose sequence and appearance information. Train the GAN to generate photo-realistic sign language video sequences based on the conditioned pose and appearance data.

5. Continuous Text-to-Pose Translation: Ensure that the NMT-based network and MG together enable continuous and coherent translation from spoken language text to pose sequences, capturing the dynamic nature of sign language.

6. Evaluation on PHOENIX14T Dataset: Evaluate the translation abilities of the approach on the PHOENIX14T Sign Language Translation dataset. Report performance metrics, including BLEU-4 scores for text-to-gloss translation on dev/test sets.

2.5.2 Advantages

- Minimal Annotation Requirement: Reducing annotation requirements streamlines the data preparation process, making it more efficient and cost-effective.
- Continuous Sign Video Generation: Continuous video generation enhances the fluidity and naturalness of sign language animations, improving overall communication quality.
- Neural Machine Translation (NMT) Integration: Integration with NMT enhances the system's ability to translate between spoken and signed languages, leveraging the power of neural networks for more accurate and context-aware translations.
- Motion Graph (MG) Enhancement: Enhancing the Motion Graph (MG) component improves the system's ability to capture and represent complex sign language motions, contributing to more realistic animations.
- Generative Adversarial Network (GAN) Realism: Incorporating GAN technology enhances the realism of generated sign language videos, making them more visually convincing and expressive.
- Comprehensive Evaluation on PHOENIX14T Dataset: Conducting a comprehensive evaluation on the PHOENIX14T dataset ensures the system's performance is validated across a diverse range of sign language expressions and scenarios, enhancing its reliability.

2.5.3 Disadvantages

- Increased Computational Demands: The integration of advanced technologies such as GANs and NMT may impose higher computational demands, potentially limiting the system's accessibility on less powerful devices.
- Potential Overfitting: Complex models like GANs and NMT may risk overfitting to specific datasets, limiting the system's generalization ability to new and diverse sign language expressions.
- Dependency on Dataset Quality: The effectiveness of the system heavily relies on the quality and representativeness of the PHOENIX14T dataset, and any biases or limitations in the dataset may affect the model's performance.
- Complex Implementation: The integration of multiple advanced technologies (NMT, MG, GAN) requires careful implementation, and any issues in the coordination of these components may impact the system's overall functionality.
- Potential Ethical Considerations: The generation of highly realistic sign language videos raises ethical considerations, particularly in ensuring that the technology is used responsibly and does not contribute to misinformation or misuse.
- Limited Information on Specific Enhancements: The brief descriptions provide minimal details on the specific enhancements made to the Motion Graph and the integration of NMT, making it challenging to assess the depth and effectiveness of these improvements.

2.6 Sign Language Recognition[6]

This paper introduces a cutting-edge system aimed at significantly improving communication with individuals facing vocal and hearing disabilities. The primary focus revolves around presenting a sophisticated and refined approach for the recognition of sign language and the conversion of spoken language into expressive sign gestures. The algorithm, meticulously designed for optimal performance, demonstrates its prowess in extracting intricate signs from video sequences. Particularly noteworthy is its ability to navigate challenging scenarios characterized by minimal clutter and dynamic backgrounds, achieved through the implementation of advanced skin color segmentation techniques.An essential feature of the algorithm lies in its capacity to differentiate between static and dynamic gestures, ensuring the extraction of relevant feature vectors. These extracted features are then subjected to classification using state-of-the-art Support Vector Machines, contributing to the system's robust and accurate gesture recognition capabilities. Moreover, the speech recognition module, built upon the standard Sphinx framework, further enhances the overall communicative functionality of the system.

In the realm of experimental validation, the results are indicative of the system's commendable performance, showcasing not only satisfactory sign segmentation under diverse background conditions but also a notably high level of accuracy in both gesture and speech recognition domains. This underscores the system's efficacy and potential in significantly enhancing communication accessibility for individuals with vocal and hearing impairments, marking a substantial stride towards inclusivity and improved quality of life.

2.6.1 Methodology

1. Data Collection:Gather a diverse dataset of sign language gestures, focusing on the American Sign Language alphabet (letters a-z). Ensure the dataset includes variations in hand orientation, dynamic movements, and diverse backgrounds.

2. Preprocessing: Apply Viola-Jones face detection to identify and eliminate facial cues from the video sequences. Utilize skin color segmentation to isolate hand gestures, focusing on the hands and face.

3. Gesture Classification: Distinguish between static and dynamic gestures based on the distance moved by the hand in consecutive frames. For static gestures: Employ Zernike moments as a shape descriptor for hand gestures. For dynamic gestures: Extract a curve feature vector to uniquely identify paths. Classify feature vectors using pre-trained Support Vector Machines (SVM) for mapping gestures to specific alphabet letters.

4. Speech Recognition: Implement the Sphinx module for speech recognition, mapping spoken alphabet to text with high accuracy. Associate recognized text with corresponding static gesture images or dynamic gesture videos.

5. System Integration: Develop a real-time system that seamlessly integrates both sign language recognition and speech recognition components. Ensure synchronization between gesture recognition and speech recognition outputs.

6. Evaluation: Conduct extensive experiments using diverse backgrounds and scenarios to evaluate the system's performance. Measure accuracy in sign language segmentation, gesture recognition, and speech-to-text conversion.

7. Performance Metrics: Employ metrics such as precision, recall, and F1 score to quantify the accuracy of the system. Evaluate real-time processing capabilities and system responsiveness.

8. Comparison: Compare the proposed system with existing sign language recognition systems, highlighting its advantages and novel contributions.

9. Optimization: Identify areas for improvement based on experimental results. Optimize algorithms for faster processing and enhanced accuracy.

2.6.2 Advantages

- Enhanced Communication Facilitation: The implementation of advanced technologies results in improved communication accessibility, fostering inclusivity and understanding among diverse user groups.

- Real-time Recognition Capabilities: The system's prowess in real-time recognition ensures swift and immediate processing, contributing to seamless interactions and reducing communication barriers.

- Diverse Gesture Recognition: The system demonstrates a comprehensive ability to recognize and interpret a diverse range of gestures, enhancing its adaptability to various communication styles and contexts.

- Robust Background Handling: The inclusion of robust background handling features ensures that the system operates effectively in different environments, minimizing interference and maintaining consistent recognition accuracy.

- Integration of Speech Recognition: The incorporation of speech recognition capabilities enhances the system's versatility, allowing users to seamlessly transition between sign and spoken language communication modes.

- Extensibility for Future Enhancements: The system's design allows for extensibility, paving the way for future updates, improvements, and the integration of emerging technologies to continually enhance its functionality.

- High Accuracy in Gesture Recognition: The system prioritizes high accuracy in gesture recognition, providing users with a reliable and precise means of communication, thereby building confidence and trust in the technology.

2.6.3 Disadvantages

- Limited Lexicon Challenges: The system encounters limitations in handling a diverse range of signs due to a constrained lexicon, potentially hindering its adaptability to various communication scenarios.

- Dependency on Lighting Conditions: The system's performance may be affected by variations in lighting conditions, leading to potential challenges in accurately recognizing and interpreting sign gestures.

- Complexity of Dynamic Gesture Recognition: Dynamic and intricate gestures pose a challenge to the system, as the complexity inherent in such movements may impact the accuracy and efficiency of gesture recognition.

- Dependency on Training Data: The system's proficiency heavily relies on the quality and diversity of the training data, potentially leading to limitations in recognizing signs that deviate from the dataset.

- Sensitivity to Hand Movement Distance: Sensitivity to variations in hand movement distances may impact the precision of gesture recognition, especially in scenarios where users exhibit different spatial dynamics.

- Resource-Intensive Nature: The system's resource-intensive nature may pose challenges in terms of computational requirements, potentially limiting its feasibility on devices with constrained processing capabilities.

- Limited Facial Expression Recognition: The system may exhibit limitations in accurately recognizing and interpreting facial expressions, potentially affecting the nuanced and expressive aspects of sign language communication.

2.7 Audio to Sign Language conversion using Natural Language Processing[7]

2.7.1 Methodology

1. Input Processing: Initiate the system by receiving audio input from the user, leveraging Google's powerful programming API for precise transcription of the spoken words into text. Utilize encoding techniques, specifically an Indian Sign

Language (ISL) encoder, to seamlessly translate the transcribed text into corresponding ISL gestures.

2. Gesture Suggestions: Enhance user interaction by offering dynamic gesture suggestions based on the ISL-encoded input, providing an intuitive and supportive experience for individuals communicating through sign language.

3. GIF Display: Implement an engaging element by displaying pre-defined GIFs that align with the recognized gestures. This visual representation adds clarity and context to the communication process.

4. Word Replacement: Ensure comprehensive communication by comparing each word in the sentence with a dictionary of images and animated GIFs. In cases where a word is not found, intelligently replace it with its synonymous representation, contributing to a more nuanced and expressive sign language translation.

5. System Libraries: Leverage the capabilities of various Python libraries for seamless implementation:TensorFlow and Keras for advanced AI and machine learning functionalities.NumPy for efficient matrix and array operations.OpenCV-python for in-depth image analysis and machine vision.Tkinter for crafting a user-friendly GUI, enhancing the overall interaction experience.PIL for streamlined image file access and modification.PyAudio for handling audio input and output, ensuring a comprehensive communication system.

6. Automatic Termination:Facilitate user-friendly interactions by incorporating an automatic termination feature. Conclude the translation process gracefully upon encountering the word "Bye," enhancing the system's responsiveness to user commands.

7. User Interaction:Prioritize user experience through the integration of a graphical user interface (GUI), making interactions with the system intuitive, accessible, and visually appealing.

8. Integration of AI and ISL Encoding: Forge a cohesive connection between artificial intelligence models and ISL encoding, effectively bridging the gap between audio input and sign language output. This integration ensures a seamless and accurate translation process.

9. Cross-Platform Audio Support: Optimize the system's functionality by utilizing PyAudio for cross-platform audio input and output compatibility, catering to a diverse range of devices and operating systems.

2.7.2 Advantages

- User-Friendly Interface: Prioritize simplicity and ease of use for individuals of all technological proficiencies. Implement clear navigation and intuitive design elements to create a welcoming interface.
- Synonym Replacement: Intelligently replace unmatched words with synonyms, enhancing communication flexibility. This feature ensures a more expressive and adaptable sign language translation, accommodating variations in language usage.
- AI Integration: Utilize advanced AI algorithms for continuous improvement in gesture recognition and translation. Implement machine learning models to enhance the system's intelligence, leading to more accurate and context-aware sign language interpretations.
- Cross-Platform Audio Support: Ensure compatibility across devices and operating systems using PyAudio technology. Optimize the system's audio capabilities, making it accessible to a broader audience with varied technological setups.
- Gesture Suggestions: Provide dynamic suggestions for sign language gestures to facilitate smoother interactions. This feature assists users in selecting appropriate signs, fostering a more seamless and user-friendly communication experience.

- Dynamic GIF Display: Enhance visual communication with a real-time and engaging GIF display for recognized gestures. This dynamic visual representation adds an interactive element to the communication process, improving user engagement and comprehension.

2.7.3 Disadvantages

- Limited Gesture Set: The system may have a finite set of gestures, potentially limiting its ability to express a diverse range of concepts and nuances in communication.
- Dependency on Speech Recognition Accuracy: The effectiveness of the system is contingent on the accuracy of speech recognition technology, and inaccuracies in transcription may lead to misinterpretations.
- Synonym Matching Challenges: Synonym matching may encounter challenges in accurately identifying and replacing words, affecting the precision of language translation.
- Static Dictionary: Relying on a static dictionary for word-image associations may hinder adaptability to evolving language trends and hinder the incorporation of new signs.
- Complexity and Processing Time: The system's complexity may result in longer processing times, potentially impacting real-time communication and responsiveness.
- Internet Dependency: If the system relies heavily on online resources, its functionality may be compromised in environments with limited or no internet connectivity.

- Learning Curve: Users may face a learning curve in adapting to the system's features and functionalities, potentially posing challenges for those unfamiliar with the technology.

2.8 Speech To Sign Language Translator For Hearing Impaired [8]

The dynamic landscape of the Internet has experienced remarkable growth in recent years, reaching unprecedented heights, particularly during the Covid-19 pandemic, where it emerged as an indispensable and pervasive communication medium. This surge in online communication has been a boon for numerous individuals, facilitating seamless access to a vast array of audio and visual content. However, amidst these technological strides, a significant challenge persists for differently-abled individuals, notably the hearing-impaired, who encounter inherent limitations in accessing and comprehending these digital resources. To address this pressing issue, an ongoing and dedicated project is actively working on the development of an innovative application. The primary goal of this application is to adeptly transform both speech and text inputs into a coherent and accessible sequence of sign language visuals, thereby enhancing the overall accessibility for the hearing-impaired community.The complex and multifaceted process involved in achieving this transformative goal begins with harnessing cutting-edge speech recognition technology to skillfully convert audio inputs into precise and accurate textual representations. Subsequently, the application employs advanced Natural Language Processing algorithms, ensuring meticulous word segmentation and root word extraction. This nuanced linguistic analysis lays the foundation for a comprehensive understanding of the input content. The final stage of this transformative journey involves the translation of the processed content into Indian Sign Language, offering a regionally relevant and culturally sensitive approach to communication. This tailored solution aligns with the diverse linguistic and cultural needs of the Indian populace, further amplifying the impact of the application.

In essence, this project is not merely about technological innovation but embodies a commitment to inclusivity and accessibility in the digital sphere. By addressing the unique challenges faced by the hearing-impaired community, the initiative strives to create a more inclusive and accessible digital landscape, fostering a society where everyone, regardless of abilities, can fully participate in the online communication revolution.

2.8.1 Advantages

1. Forms of Input:

   Text Input: Accepts input in textual form, allowing users to directly input written content for translation.

2. Live Speech Input: Enables real-time speech recognition using PyAudio, providing users with the option to convey messages orally.

3. Recorded Audio File Input: Supports the submission of pre-recorded audio files, enhancing the system's versatility in handling various input formats.

4. Speech Recognition: Utilizing PyAudio for live speech input, the system employs the Google Speech Recognizer API to convert spoken words into text. To accommodate longer audio files, the system intelligently segments the input based on occurrences of silence, ensuring accurate transcription.

5. Pre-processing of Text: In the pre-processing phase, the system enhances the quality of the input by eliminating filler words and punctuation, focusing on extracting meaningful content. Leveraging the Porter Stemming Algorithm, morphological and inflexional endings of English words are removed, contributing to improved accuracy in identifying root words.

6. Text to Sign Language: During the conversion of text to sign language, the system meticulously iterates through the processed text, searching for corresponding sign language videos locally. The use of OpenCV facilitates the seamless display of video sequences, enhancing the visual representation of sign language gestures. In cases where local matches are not found, the system extends its search to the "Indian Sign Language Portal" through web scraping, ensuring a comprehensive database of sign language translations.This robust integration of various input methods, advanced

speech recognition techniques, and thorough text pre-processing exemplifies the system's commitment to flexibility, accuracy, and accessibility for users across diverse communication preferences and needs.

2.8.2 Advantages

- Tailored for Indian Sign Language: The system is intricately designed to meet the specific requirements of Indian Sign Language, effectively catering to the communication needs of the Indian deaf community.

- Incorporates Natural Language Processing (NLP) and Porter Stemming: By integrating advanced language processing techniques such as NLP and Porter Stemming, the system ensures efficient and accurate language interpretation.

- Supports Various Input Forms: The system's versatility shines through its ability to process different input forms, including text, live speech, and recorded audio. This adaptability empowers users to choose their preferred mode of communication.

2.8.3 Disadvantages

- Dependency on Google Speech Recognizer API: The reliance on the Google Speech Recognizer API introduces potential issues related to latency and external dependencies, which may impact real-time responsiveness.

- Limited Effectiveness for Unavailable Words: The system's effectiveness may be constrained when encountering words not present in the local system or the Indian Sign Language Portal. This limitation underscores the importance of a comprehensive vocabulary.

- Dependency on Indian Sign Language Dataset: The overall efficacy of the system is closely tied to the comprehensiveness of the Indian Sign Language dataset. Insufficient or incomplete datasets may pose challenges in achieving accurate translations and gestures.

2.9 Automatic Translate Real-Time Voice to Sign Language Conversion for Deaf and Dumb People [9]

Sign Language Recognition stands as a rapidly evolving and burgeoning field within the realm of research, witnessing a surge in innovative techniques. This area of study primarily addresses the communication needs of the deaf and dumb community. In this paper, we present the design and preliminary implementation of a robust system aimed at automating the translation process from voice to text and further into sign language animations. The significance of Sign Language Translation Systems lies in their potential to profoundly enhance the lives of the deaf, particularly in realms such as communication, information exchange, and employment, where machine translation can bridge conversations across different languages. Therefore, delving into the intricacies of speech recognition becomes imperative.

Voice recognition algorithms typically confront three key challenges, starting with the extraction of features from speech. Another challenge arises when dealing with limited sound galleries available for recognition. The final hurdle involves transitioning from speaker-dependent to speaker-independent voice recognition. Feature extraction from speech assumes paramount importance in our proposed method, and various procedures exist for this purpose. Mel-Frequency Cepstral Coefficients (MFCCs) emerge as a common and effective choice in speech recognition systems. The algorithm initiates with preprocessing and signal conditioning, followed by the extraction of features from speech using Cepstral coefficients. The ensuing result undergoes segmentation, leading to the final stage where the recognition process identifies words and subsequently converts the recognized words into facial animations.

It's crucial to note that this project is a work in progress, with ongoing exploration of novel and promising methods detailed in the current report. The system is poised to execute the recognition process by comparing the parameter set of the input speech with stored templates, ultimately displaying the sign language in video captions on various screens, including computers and mobile devices. This approach holds great potential for facilitating the learning process for deaf and dumb individuals or students, offering them an accessible means to comprehend subjects through online platforms like YouTube videos. As the project unfolds, it promises to contribute significantly to the realm of Sign Language Recognition, fostering inclusivity and accessibility for the deaf community.

2.9.1 Methodology

1. Understanding the Landscape:Conduct a comprehensive review of the landscape of communication methods and educational approaches for the hearing impaired in India.Engage with stakeholders, including students, teachers, administrators, and officials, through semi-structured interviews to identify challenges and pain points in the existing education system.

2. Identification of Educational Approaches:Explore the three primary methods of teaching the hearing impaired in India: Indian Sign Language (ISL), Oralism, and Total Communication (TC).Understand the strengths and limitations of each method, with a focus on how they address significant and severe hearing loss.

3. Stakeholder Participation and Feedback:Involve stakeholders in participatory design sessions to gather insights and feedback on potential solutions.Based on discussions, identify the need for a technology-supported solution to enhance education for the hearing impaired.

4. Conceptualization of the System:Define the concept of a technology-supported educational aid that provides printed and sign language prompts in videos.Develop a

clear understanding of the features and functionalities required to address the identified challenges.

5. Validation through A/B Testing:Conduct A/B testing to validate the effectiveness of the proposed system.Compare the learning outcomes and comprehension of students exposed to the system against those following traditional methods.

6. Observational Study in a School Environment:Conduct an observational study in a school environment to assess the impact of the system on student learning.Gather data on students' ability to grasp and retain information through the system's captions and sign language support.

7. Database of 3D Generated Signs:Develop a database of 3D-generated signs that will serve as sign captions for the educational videos.Ensure that the signs are culturally relevant and representative of Indian Sign Language.

8. Speech Processing and Natural Language Processing (NLP):Utilize speech processing techniques to transcribe the audio content of educational videos.Implement an NLP module with a Subject–Object–Verb rule-based syntax to convert transcriptions into a format suitable for sign captions.

9. Integration of Sign Captions:Overlay the 3D sign captions onto the educational videos based on the transcribed and processed content.Ensure synchronization between the sign captions and the spoken content to provide a coherent learning experience.

10. Usability Testing and Iterative Refinement:Conduct usability testing with students to assess the user-friendliness of the system.Gather feedback and iteratively refine the system based on user input to enhance its effectiveness.

11. Scale-Up and Implementation:Plan for the scale-up and widespread implementation of the system in educational institutions catering to the hearing impaired.Collaborate with educational authorities and organizations to integrate the technology-supported solution into the mainstream curriculum.

12. Continuous Monitoring and Evaluation:Implement a system for continuous monitoring and evaluation of the system's impact on student learning.Collect data on educational outcomes, student engagement, and overall satisfaction to inform future enhancements.

13. Documentation and Knowledge Sharing:Document the entire methodology, including research findings, system development, and outcomes.Share knowledge and insights with the educational community, researchers, and policymakers to contribute to the field of technology-supported education for the hearing impaired.

2.9.2 Advantages

- Comprehensive Understanding: The methodology involves a thorough understanding of the educational landscape for the hearing impaired in India, ensuring that the proposed solution aligns with the existing challenges and needs.

- Stakeholder Involvement: Actively involving stakeholders through participatory design sessions and interviews ensures that the solution is designed with insights from those directly impacted, enhancing its relevance and acceptance.

- Identification of Challenges: The methodology allows for the identification and acknowledgment of challenges within the current educational approaches for the hearing impaired, providing a clear starting point for intervention.

- A/B Testing for Validation: The use of A/B testing provides a scientific approach to validating the effectiveness of the proposed system, offering empirical evidence of its impact on student learning outcomes.

- Observational Study in Real Context: Conducting an observational study in a school environment allows for the assessment of the system's practical implications and its integration into the daily lives of students.

2.9.3 Disadvantages

- Resource Intensive: Developing 3D-generated signs, integrating speech processing, and conducting A/B testing can be resource-intensive, requiring a significant investment of time and funding.

- Technical Challenges: The integration of speech processing, NLP, and 3D sign animations may pose technical challenges, and expertise in multiple domains is required for successful implementation.

- Cultural Sensitivity: Ensuring cultural sensitivity in the development of the sign language database may be challenging, and misinterpretations or inaccuracies could arise.

- Potential Resistance to Change: Implementing a technology-supported solution in educational institutions may face resistance from traditional teaching methods or administrative structures.

- Ethical Considerations: The use of technology in education, especially with vulnerable populations, requires careful ethical considerations to ensure privacy, consent, and responsible use.

2.10 .Translating Speech to Indian Sign Language Using Natural Language Processing [10]

Language serves as a fundamental medium for conveying ideas, thoughts, and information to others, fostering meaningful communication. For the hearing-impaired, sign language plays a pivotal role in understanding and expressing thoughts. Each country possesses its unique sign language rooted in its native tongue. In this research paper, we pivot our attention towards Indian Sign Language, predominantly employed by the hearing- and speaking-impaired communities in India. Effective communication hinges on the ability to listen, yet this poses a significant challenge for those who are hearing-impaired. Addressing this issue, our research endeavors to introduce an innovative audio-to-Indian Sign Language translation system, aiming to bridge the communication gap between hearing-impaired individuals and society at large.

The system operates by accepting both audio and text inputs, subsequently cross-referencing them with videos stored in the authors' database. Upon finding a match, it generates corresponding sign movements based on the grammatical rules of Indian Sign Language. In cases where a match isn't established, the system engages in tokenization and lemmatization processes to enhance understanding. At the core of this system lies natural language processing, endowing it with capabilities such as tokenization, parsing, lemmatization, and part-of-speech tagging. This comprehensive approach not only facilitates communication for the hearing-impaired but also contributes to the broader goal of fostering inclusivity and understanding within society. As we delve into this endeavor, the potential impact of our audio-to-Indian Sign Language translation system on bridging communication barriers becomes increasingly apparent.

2.10.1 Methodology

1. Tokenization:Utilize NLTK's 'tokenize()' module for word and sentence tokenization.

Word tokenize dissects sentences into word sequences.Sentence tokenize fragments paragraphs into cohesive sentences.

2. Removal of Stop Words:Employ a curated list to eliminate common but less informative words (e.g., 'her,' 'me,' 'has,' etc.).Enhances overall system performance by removing non-essential words.
3. Parsing:Apply syntax analysis using the Stanford Parser.Ensures the string resulting from tokenization adheres to proper grammar.
4. Lemmatization:Transform inflected word forms into root-based dictionary forms (lemmas).Crucial in Indian Sign Language (ISL) for obtaining root words.
5. Address limitations by integrating part-of-speech tagging.
6. Part-of-Speech Tagging:Label words with constructs of English grammar (e.g., adverbs, adjectives, nouns, verbs).Involves a collection of tuples, enhancing accuracy in word-to-base word conversion.
7. Integrated Approach:Synergize part-of-speech tagging and lemmatization.Overcomes lemmatization limitations, refining accuracy.Signifies a pivotal stride in augmenting the accuracy of the system.This comprehensive linguistic toolkit ensures the audio-to-Indian Sign Language conversion system navigates language intricacies with finesse, fostering enhanced communicative efficacy for the hearing-impaired community.

2.10.2 Advantages

- Enhanced Linguistic Processing:NLTK facilitates sophisticated text processing, including tokenization and parsing, optimizing linguistic analysis.
- Improved Understanding of Context:Integration of part-of-speech tagging and lemmatization refines contextual understanding, vital for accurate word conversion.
- Streamlined Tokenization:NLTK's tokenization modules efficiently break down sentences into words and paragraphs into sentences, streamlining the process.
- Effective Stop Word Removal:Elimination of stop words enhances system efficiency by focusing on meaningful content, minimizing processing overhead.
- Versatility in NLP Techniques:NLTK's versatility supports diverse NLP techniques, ensuring adaptability to various linguistic challenges.

2.10.3 Disadvantages

- Dependency on External Libraries:Reliance on NLTK introduces a dependency on external libraries, potentially affecting system stability.
- Complexity in Parsing:The syntax analysis phase (parsing) may introduce complexity, demanding careful handling to ensure accurate results.
- Accuracy Challenges in Lemmatization:Lemmatization, while effective, faces challenges in accurately determining root words, especially in nuanced contexts.
- Resource Intensive:The combined approach of part-of-speech tagging and lemmatization may demand additional computational resources.
- Potential for Semantic Ambiguity:Despite advanced linguistic processing, there's a potential for semantic ambiguity in certain linguistic constructs.

2.11 Speech To Sign Language Conversion using Convolutional Neural Networks[11]

The Speech to Sign Language Recognition project stands at the forefront of addressing a critical societal need by aiming to facilitate communication between deaf-mute individuals and the broader community. The intricacy of this task is compounded by the diverse and complex variations in hand actions inherent to sign language. With the overarching goal of leveraging convolutional neural networks (CNN), the project seeks to convert spoken words into text and subsequently translate them into sign language. This innovative approach holds immense social impact, particularly in fostering inclusivity and understanding. The utilization of Mel-Frequency Cepstral Coefficients (MFCC) as a feature extraction method, coupled with the robust capabilities of CNN as a classifier, underscores the project's commitment to achieving effective communication solutions. The method's noteworthy 93% accuracy serves as a testament to its efficacy in enhancing the communicative abilities of individuals grappling with hearing or speech disabilities. This holistic approach aligns with broader societal goals of accessibility and inclusion, reinforcing the project's significance in contributing to a more inclusive and understanding world.

2.11.1 Methodology

1. Dataset:The RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) is a gender-balanced, multinational database with 24 professional actors. It includes 7,356 files capturing various emotional expressions in speech and songs, labeled with numerical codes (01 to 08) representing different emotions.

2. Speech to Sign Language Recognition: The architecture of a Speech to Sign Language Recognition system using Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction and Convolutional Neural Network (CNN) for classification. The system involves three stages: speech signal processing, feature extraction, and classification, mapping speech input to sign language letters.

3. Feature Extraction – MFCC:It depicts MFCC extracting features from a speech signal, involving the determination of average energy, Fourier Transform in the frequency domain, and extraction of three largest frequency peaks for each window.

4. CNN:The Convolutional Neural Network (CNN) includes Convolutional, Pooling, and output layers, along with Activation and Dropout layers for classification. It processes input audio through convolution and pooling layers, extracting features for effective Speech to Sign Language Recognition.

2.11.2 Advantages

- Rich Emotion Dataset: The RAVDESS dataset provides a diverse and emotive set of speech and song samples, offering a comprehensive range of emotional expressions for training and evaluation.

- Gender Balance: With 12 female and 12 male professional actors, the dataset maintains gender balance, ensuring a representative set of voices and expressions.

- Numerical Emotion Labeling: The numerical labeling (01 to 08) associated with different emotions in the dataset facilitates clear categorization for training and analysis.

- Mel-Frequency Cepstral Coefficients (MFCC): The use of MFCC for feature extraction in the Speech to Sign Language Recognition system is a well-established method known for capturing relevant acoustic features.

- Convolutional Neural Network (CNN): Leveraging CNN architecture enhances the system's capability to recognize and classify features, improving the accuracy of Speech to Sign Language translation.

2.11.3 Disadvantages

- Limited Cultural and Linguistic Diversity: The dataset may lack diversity in cultural and linguistic aspects, potentially limiting the generalization of the Speech to Sign Language Recognition system to diverse user groups.
- Dependency on Professional Actors: The reliance on professional actors may introduce a bias in emotional expressions, and the system's performance might differ when applied to non-professional or real-world scenarios.
- Potential Overfitting: Depending on the size and diversity of the dataset, there is a risk of overfitting, where the system may perform well on the training data but struggle with unseen data.
- Complexity of CNN Training: Training and optimizing a Convolutional Neural Network can be computationally intensive and may require substantial resources.
- Need for Continuous Dataset Updates: Emotion expressions can evolve over time, and the need for continuous updates to the dataset may arise to keep the Speech to Sign Language Recognition system relevant and effective.

2.12 Audio to Sign Language Translation for Deaf People [12]

The primary objective of this project is to address the communication challenges faced by deaf individuals in various aspects of daily life, such as interacting, playing games, attending seminars, and participating in video conferences. The system focuses on converting audio messages into sign language, making communication more accessible for both deaf and hearing individuals who may not be familiar with sign language. By taking audio input, converting it into text, and displaying predefined Indian Sign Language images or GIFs, the system aims to bridge the communication gap between the deaf community and the general population.

Sign language serves as the primary mode of communication for deaf individuals, incorporating hand movements, body language, and facial expressions. With 135 types of sign languages globally, the project specifically utilizes Indian Sign Language (ISL). The system enables the deaf community to engage in various activities, from daily interactions to accessing information. The application processes speech input, converts it into text, and displays corresponding Indian Sign Language images. The front end is designed using EasyGui, while PyAudio handles speech input through a microphone. Google Speech API is employed for speech recognition, and Natural Language Processing (NLP) is applied for text pre-processing. The system concludes with dictionary-based machine translation.

The tool also features a sign language recognizer for interpreting the sign language of deaf and mute individuals. Gesture recognition, a crucial aspect of the system, tackles the challenge of segmenting foreground objects from complex backgrounds. While humans find it easier to interpret images, computer vision problems present challenges due to the disparity in how humans and computers perceive images. Despite these challenges, the project aims to contribute to overcoming communication barriers and enhancing the overall experience for the deaf community.

2.12.1 Methodology

1. Utilizes PyAudio module for audio input through a microphone.Employs Google API for speech-to-text conversion.Displays transcribed text and generates sign language gestures using the Indian Sign Language (ISL) generator.

2. Audio-to-Text Conversion:Involves dependency parsing to analyze sentence grammar and word relationships.

3. Text-to-Sign Language Process:Utilizes Google Speech API for speech recognition. Applies Natural Language Processing (NLP) for text preprocessing.Implements dictionary-based machine translation.Generates ISL based on ISL grammar rules. Outputs sign language gestures with a signing avatar.

4. Facial Expressions:While not explicitly focused on facial expressions, acknowledges their importance in sign language.Applications:Versatile applications include accessing government websites lacking video clips for deaf and mute individuals. Useful for online form filling without the presence of an interpreter.

2.12.2 Advantages

- Accessibility Improvement:Facilitates communication for individuals with hearing impairments.Bridges the gap between the deaf community and the general public.

- Versatility:Addresses the scarcity of audio-to-sign language conversion projects.Beneficial for both normal and hearing-impaired individuals.
- Innovative Workflow:Utilizes PyAudio and Google API for effective audio processing.Implements NLP and machine translation for accurate text-to-sign language conversion.Incorporates ISL grammar rules for authentic sign language generation.

- Applications:Offers practical applications, such as accessing government websites and online form filling.Demonstrates potential in various scenarios requiring sign language interpretation.

2.12.3 Disadvantages

- Facial Expressions Not Prioritized:Does not explicitly focus on facial expressions, a crucial aspect of sign language.May limit the system's ability to convey nuanced emotions and meanings.

- Dependency on Google API: Relies on the Google API for speech recognition, introducing potential latency and dependency.System performance may be affected if there are issues with the external API.

- Limited ISL Vocabulary:Effectiveness heavily relies on the comprehensiveness of the Indian Sign Language dataset.May face challenges in accurately representing signs for less common or specialized words.

- Synonym Replacement Complexity:Substituting words with synonyms may not always capture the exact meaning.Challenges may arise in selecting appropriate synonyms for accurate sign language representation.

- Technical Challenges:Complex workflow involving multiple steps, including audio processing, NLP, and machine translation.Potential for technical challenges in maintaining system robustness and reliability.
- Learning Curve:Users, particularly developers or administrators, may face a learning curve in understanding and customizing the system.System customization and troubleshooting may require technical expertise.

2.13 SIGN LANGUAGE RECOGNITION AND TRANSLATOR APPLICATION [13]

In the contemporary world, effective communication stands as a cornerstone, facilitating the swift and accurate exchange of information among individuals. However, the challenge persists for those who are deaf and mute, relying on Sign Language—a nuanced form of communication involving gestures, body movements, and facial expressions. Unfortunately, this mode of communication poses a barrier, as not everyone possesses the proficiency to interpret Sign Language.

In response to this communication gap, our innovative Android application aims to empower deaf and mute individuals by seamlessly translating Sign Language into conventional spoken language. The user-friendly interface allows individuals to capture images of Sign Language expressions, initiating a streamlined background process. Leveraging cutting-edge TensorFlow libraries, the system rapidly processes these images, transforming them into meaningful sentences or phrases. The outputs are provided in two modes: textual representation and audio playback.

An exceptional feature of our system is its adaptability to various sign languages, encompassing Indian Sign Language, American Sign Language, and more. Furthermore, the application transcends language barriers by translating the interpreted signs into multiple languages, including English and Hindi. A noteworthy addition is the incorporation of a repository of frequently used phrases, eliminating the need for capturing photos. Users can effortlessly click a button to hear the audio rendition of these commonly used expressions, enhancing the efficiency and versatility of communication for individuals with hearing and speaking impairments

2.13.1 Methodology

1. Machine Learning and Image Processing:The core essence and intelligence behind the project are embedded in the Machine Learning and Image Processing Algorithms, leveraging the robust platform provided by TensorFlow. Meticulous attention to theoretical concepts has been maintained, focusing on recognizing signs from input images using a meticulously curated training dataset that encompasses classes for

every required phrase. The final model, trained and tested, is stored in a supportive format, seamlessly integrated into the Application Project.

2. TFLITE Model Maker: TensorFlow offers a range of model formats, including TensorFlow, TensorFlow Lite, TensorFlow.js, and TensorFlow Extended, tailored for Software, Mobile and IoT, Web, and Production-based applications, respectively. Given the mobile-centric nature of our project, we employ tflite-model-maker to effortlessly train and test the model with just a few lines of code.

3. Teachable Machine: Teachable Machine stands as a web-based tool that simplifies and accelerates the creation of machine learning models, making it accessible to all. Anyone can train their model through three straightforward steps:
   i) Upload or Add Sample Data based on Class
   ii) Train the model by adjusting parameters such as epochs, batch size, and learning rate.
   iii) Export the model in the desired format.
   The beauty of Teachable Machine lies in its user-friendly approach, eliminating the need for prerequisite machine learning knowledge.

4. Android Studio:This platform serves as the canvas for designing an interactive interface, reaching a broad audience with ease. The mobile application, intricately integrated with the TensorFlow library, boasts a minimalistic UI design, ensuring user-friendliness. All views adhere to standardized positions, eliminating the need for users to memorize actions. The interface comprises simple click buttons facilitating the transition to the camera. After capturing images, pressing the "OK" button redirects users to the main screen, where the recognized text is visible. Additionally, audio is generated, conveniently accessible with a simple button click.

2.13.2 Advantages

- Efficient Recognition:TensorFlow integration ensures accurate machine learning model training and image recognition.
- TensorFlow Model Flexibility:TensorFlow Lite offers flexibility for mobile applications with simplified training using the model-maker.
- Accessible Machine Learning:Teachable Machine democratizes machine learning, making it accessible without advanced knowledge.
- User-Friendly Interface:Android Studio enables a user-friendly interface with a minimalistic design for easy navigation.
- Audio Integration:Seamless integration of audio generation allows users to hear recognized content instantly.

2.13.3 Disadvantages

- Model Limitations:Recognition limitations may arise with complex gestures or variations not covered by the model.
- Training Dependency:System effectiveness depends on a diverse and comprehensive training dataset.
- Internet Dependency:Certain functionalities may require an internet connection, limiting offline use.
- Facial Expression Neglect:The system doesn't focus on facial expressions in sign language, missing a crucial aspect.
- Limited Sign Language Support:Challenges may occur in recognizing regional variations or signs not covered in the dataset.

2.14 English Text to Indian Sign Language Translator [14]

Sign language, being a fundamental means of communication for those with speaking and hearing impairments, underscores the importance of bridging communication gaps. While several platforms exist for translating or recognizing sign language and converting it to text, systems that facilitate the conversion of text to sign language are not as prevalent. This scarcity can be attributed to the limited availability of comprehensive sign language corpora. Our project aims to fill this void by introducing an advanced translation system featuring a sophisticated parsing module.

This parsing module plays a pivotal role in transforming input English sentences into phrase structure grammar representations. Subsequently, Indian sign language grammar rules are applied to these representations, ensuring a seamless conversion from text to sign language. The intricacies of this process include the strategic reordering of sentences after the removal of stopwords, the application of stemming to derive root forms of words, and the systematic checking of each word against an extensive dictionary. This dictionary, a repository of videos illustrating corresponding signs, enriches the system's capacity to accurately interpret and translate diverse linguistic inputs.

In cases where certain words are not found in the dictionary, the system intelligently replaces them with their synonyms, contributing to a more comprehensive and contextually relevant translation. What sets our system apart is its innovative approach; unlike existing solutions that often focus on the direct conversion of individual words into Indian sign language, our project aspires to elevate this capability to encompass entire sentences. The goal is to enable real-world communication in sign language by seamlessly integrating grammatical nuances and contextual relevance into the translation process.

2.14.1 Methodology

1. Parsing of the Input English Text:To enable rule-based conversion between languages, understanding the grammatical structure is crucial. The Stanford parser facilitates this by providing part-of-speech tagged text, a context-free grammar representation of phrase structure, and type dependency representation. Penn tree tags are employed for parsing English sentences.

2. Grammar Rules for English to ISL Conversion:Sign language and spoken language exhibit distinct grammar rules, significantly complicating the translation process. Addressing these differences is crucial for accurate and meaningful translation.

3. Elimination of Stop Words:Indian Sign Language (ISL) relies on words with specific meanings. To enhance the relevance of translated content, unnecessary words such as coordinating conjunctions, determiners, adjectives, adverbs, and symbols are removed.

4. Lemmatization and Synonym Replacement:ISL emphasizes the use of root words. The Porter Stemmer rules are applied for lemmatization, converting words to their root forms. Additionally, each word is cross-checked in a bilingual dictionary. If a word is not found, it is replaced with its synonym, maintaining the same part of speech.

5. Video Conversion Stage:Once the ISL-transformed text is obtained, the program searches for matches within the available dataset for each word. A basic string matching algorithm is employed to compare the processed input text with video labels. The final output is a sequence of videos displayed on the screen.

6. Language:Python (version 3) will be utilized for project development, leveraging various Python libraries for implementation.

7. Tools/Libraries Used:NumpyJASigning - Sigml file conversion into animated format movies.pynltk
8. Output Generation:The output from this module will be a movie clip of ISL-translated words. The database contains videos for each separate word, and the resulting video is a compilation of these words presented in a sequential format.

2.14.2 Advantages

- Grammar Precision:The parsing stage ensures a high level of grammatical accuracy by employing the Stanford parser, enabling the system to understand the structure of both source and target languages.

- Contextual Elimination:The removal of stop words enhances the relevance of translated content, focusing on words with specific meanings in Indian Sign Language (ISL).

- Lemmatization for Root Words:The lemmatization process ensures the usage of root words in ISL, aligning with the language's grammar rules.

- Synonym Replacement:The system intelligently replaces missing words with their synonyms, maintaining the same part of speech, contributing to meaningful translations.

- Video Conversion Accuracy:By utilizing basic string matching algorithms, the system accurately converts ISL-transformed text into corresponding videos, ensuring precise representation.

- Multilingual Support:The system supports multiple sign languages, allowing for broader accessibility and inclusivity.

2.14.3 Disadvantages

- Complex Grammar Rules:The need to handle distinct grammar rules between sign and spoken languages increases the complexity of translation.

- Dependency on Training Data:The accuracy of video conversion relies on the availability and quality of the training dataset, potentially limiting coverage for less common words.

- Limited Facial Expression Consideration:The system does not focus on facial expressions, a crucial aspect of sign language communication, potentially leading to a loss of nuance in translation.

- String Matching Limitations:Basic string matching may face challenges with variations in word forms, potentially affecting the accuracy of video selection.

- Processing Time:The entire process involves multiple stages, which may lead to increased processing time, especially for lengthy sentences.

- Learning Curve:Users unfamiliar with sign language translation systems may face a learning curve in understanding the rules and processes involved.

2.15 AUDIO OR TEXT TO SIGN LANGUAGE CONVERTER [15]

Effective and inclusive communication, recognized as a fundamental human right, confronts substantial challenges for millions of individuals experiencing deafness or hearing impairment. The formidable language barrier posed by sign language often hinders seamless communication with the broader hearing population. This project endeavors to overcome this barrier by leveraging cutting-edge technology. By integrating natural language processing (NLP) and dynamic animation, we have engineered a sophisticated system capable of transforming spoken or written language into captivating Indian Sign Language (ISL) animations.

This groundbreaking solution not only empowers the deaf and hearing-impaired community but also facilitates smooth communication for the general population. The system adeptly identifies speech, processes textual input, and generates immersive 3D sign language animations through the utilization of Blender 3D Animation tools. With a foundation built upon Python, the Django Framework, NLTK Library, and various other technologies, we have crafted an accessible and interactive platform poised to revolutionize educational practices, enhance public announcements, and elevate everyday conversations.

This project stands as a noteworthy stride toward fostering a more inclusive and understanding society, where communication transcends conventional boundaries

2.15.1 Methodology

1. Audio Input Capture:Implement a mechanism to capture audio input from users.
2. Audio to Text Conversion:Utilize appropriate tools or APIs for converting audio input into text.
3. Natural Language Processing (NLP):Apply NLP techniques for processing the converted text.Include steps such as stopword removal, lemmatization, and part-of-speech tagging.
4. ISL Animation Generation:Employ Blender 3D animation tools to create ISL animations.
5. Develop a set of rules or mappings to translate processed text into sign language gestures.
6. User-Friendly Interface:Design an interactive and user-friendly interface for ease of use.
7. Community-Specific Adaptation:Tailor the ISL representations to address the specific linguistic and cultural nuances of the Indian community.
8. Innovation in Communication Enhancement:Innovate in the approach to enhancing communication for the hearing-impaired.Explore creative ways to make ISL animations more engaging and effective.
9. Quality Assurance:Implement mechanisms for quality assurance in ISL animations.
10. Consider feedback loops for continuous improvement.
11. Accessibility Features:Integrate features that enhance accessibility for individuals with hearing impairments.
12. Technology Integration:Ensure seamless integration of Blender 3D animation tools and NLP techniques into the project workflow.

2.15.2 Advantages

- Inclusive Communication:Enables inclusive communication by converting audio or text into Indian Sign Language (ISL).Facilitates interaction with deaf and hearing-impaired individuals.
- User-Friendly Interface:Provides a user-friendly interface, making it accessible to a wide range of users.Enhances ease of use for both hearing-impaired and general users.
- Cultural Sensitivity:Adapts ISL representations to address the specific linguistic and cultural nuances of the Indian community.Promotes cultural sensitivity in sign language animations.
- Innovation in Communication:Represents an innovative solution for enhancing communication for the hearing-impaired.
- Utilizes technology, animations, and NLP to create a novel communication platform.
- Educational Potential:Holds potential for revolutionizing education, public announcements, and everyday conversations.Provides an interactive and engaging way to learn and communicate.

2.15.3 Disadvantages

- Technical Dependencies: Relies on technical components such as Blender 3D animation tools and NLP techniques. Technical challenges or limitations may impact the system's performance.
- Accessibility Challenges:Faces challenges related to the accessibility of technology for certain users.Limited access to devices, internet connectivity, or specific tools may hinder widespread adoption.
- Accuracy Concerns:Accuracy of ISL animations heavily depends on the quality of audio-to-text conversion and NLP processing.Inaccuracies in these processes may lead to misunderstandings.
- Resource Intensiveness:Animation generation and NLP processing may be resource-intensive.Resource requirements could affect the system's scalability.
- Learning Curve:Users, especially those unfamiliar with sign language, may face a learning curve.Requires effort to understand and adapt to the system's functionalities.

- Cultural Adaptation:Adapting ISL representations to specific cultural nuances may be challenging.Requires ongoing efforts to ensure cultural relevance and accuracy.

CHAPTER 3
METHODOLOGY

3.1 Methodology

Speech Recognition and Text Conversion:

This innovative application seamlessly integrates live audio speech recordings as its primary input, leveraging the powerful capabilities of the JavaScript Web Speech API. Through this integration, spoken words are swiftly recognized and converted into text, constituting a pivotal initial step in the user interaction process. This step involves not only capturing the user's speech but also intricately processing it to obtain a precise textual representation of the spoken content. Utilizing sophisticated algorithms and neural network models, the application ensures high accuracy in transcribing diverse speech patterns and accents, enhancing accessibility and usability for a wide range of users. Moreover, continuous refinement through machine learning techniques enables the system to adapt and improve its transcription accuracy over time, enhancing user satisfaction and usability.

Text Preprocessing:

Following the conversion of speech to text, the application proceeds with meticulous text preprocessing using the esteemed Natural Language Toolkit (NLTK). Renowned for its proficiency in language processing tasks, NLTK undertakes vital tasks such as tokenization, stemming, and the removal of stop words. By ensuring the cleanliness and refinement of the input text, NLTK sets the stage for subsequent processing steps, thereby enhancing the overall efficiency and accuracy of the application. Additionally, advanced techniques such as named entity recognition and part-of-speech tagging further enrich the textual data, enabling more nuanced interpretation and analysis. Through iterative refinement and optimization, the preprocessing pipeline evolves to handle diverse linguistic nuances and idiosyncrasies, ensuring robust performance across various language domains and contexts.

Indian Sign Language (ISL) Animation:

At the core of this project lies the compelling capability to generate relevant Indian Sign Language animations based on the meticulously processed textual input. This transformative feature is facilitated through the seamless mapping of textual content to appropriate ISL gestures and signs. To bring this vision to life, the application harnesses the dynamic capabilities of a meticulously crafted 3D animation character, painstakingly designed using Blender 3D. This character serves as the conduit for conveying the intricacies of sign language, faithfully translating the input text into expressive and meaningful gestures. Through continuous refinement and optimization, the animations achieve a level of realism and fluidity that enhances user engagement and comprehension. Furthermore, the system leverages deep learning algorithms to adaptively adjust animation styles based on user feedback and preferences, ensuring personalized and engaging interactions.

User Interaction:

Empowering users with intuitive interaction capabilities, the application offers a seamless interface through which users can effortlessly engage with the system. By simply clicking on the microphone button or initiating voice commands, users can initiate the recording of their speech, kickstarting the process of generating relevant sign language animations. Upon processing the recorded speech, the application swiftly generates and displays the corresponding animated outputs, ensuring a seamless and immersive user experience. Furthermore, users have the option to manually input text, eliciting the system's prompt and precise delivery of corresponding sign language animations. In essence, this multifaceted application serves as a beacon of accessibility and inclusivity, bridging the gap between spoken language and sign language with finesse and innovation. Through continuous updates and enhancements, the application strives to push the boundaries of accessibility technology, empowering users with new ways to communicate and interact effectively."

3.2 Technologies Used

3.2.1 HTML (Hyper Text Markup Language):

HTML serves as the standard markup language for web documents, facilitating their display in web browsers alongside CSS and JavaScript. It provides semantic structuring for web pages, allowing integration of various elements like images and interactive forms. HTML elements are defined by tags, although these tags are not directly displayed by browsers.

3.2.2 CSS (Cascading Style Sheets):

CSS plays a crucial role in web development by separating presentation from content. This separation enhances accessibility and enables easier management of layout, colors, and fonts. By utilizing CSS, developers can create consistent and visually appealing designs across multiple web pages. Additionally, CSS supports responsive design, ensuring optimal display on various devices and screen sizes. It also allows for the efficient caching of stylesheets, improving page loading times and overall performance.

3.2.3 Python:

Python stands as a high-level, general-purpose programming language known for its emphasis on code readability through significant indentation. Its versatile design and object-oriented approach aid programmers in crafting clear, logical code for projects of varying scales. Python supports multiple programming paradigms and boasts a comprehensive standard library, earning it the moniker of a "batteries included" language. Guido van Rossum initiated Python's development in the late 1980s, releasing its first version in 1991. Python 2.0, introduced in 2000, brought forth features like list comprehensions and Unicode support. Python 3.0, released in 2008, marked a major revision, leading to the discontinuation of Python 2 with version 2.7.18 in 2020. Python consistently ranks among the most popular programming languages.

3.2.4 Django framework:

Django aims to simplify the creation of complex, database-driven websites by prioritizing component reusability, minimal code, loose coupling, rapid development, and adherence to the DRY (Don't Repeat Yourself) principle. Python serves as the foundation for Django, including settings, files, and data models. The framework offers an optional administrative interface for CRUD (Create, Read, Update, Delete) operations, dynamically generated through introspection and configuration via admin models. Notable sites leveraging Django include Instagram, Mozilla, Disqus, Bitbucket, Nextdoor, and Clubhouse.

3.3.1 NLTK Library:

NLTK, hailed as a valuable tool for teaching and practicing computational linguistics with Python, provides rich functionalities for natural language processing.

3.3.2 Word Tokenization:

The word_tokenize() method splits sentences into words, facilitating text understanding in machine learning applications. It serves as a precursor to text cleaning processes like punctuation and numeric character removal or stemming. Word tokenization plays a vital role in converting text data into numeric format, essential for training and prediction in machine learning models.

3.3.3 Elimination of Stop Words:

In natural language processing, irrelevant words such as articles, conjunctions, and various parts of speech are filtered out to focus on meaningful content.

3.3.4 Lemmatization and Synonym Replacement:

Lemmatization involves reducing words to their root form, crucial for maintaining consistency in language processing tasks like Indian Sign Language (ISL) translation. Additionally, words are cross-checked with bilingual dictionaries to replace non-existent words with suitable synonyms sharing the same part of speech.

3.4 System Specifications

Hardware Specifications:

Hardware interfaces specifies the logical characteristics of each interface between the software product and the hardware components of the system. The following are some hardware requirements.

Processor: Intel core i5

Hard disk: 1TB

RAM: 5GB

Software Specifications:

Software Requirements specifies the logical characteristics of each interface between the software product and hardware components of the system. The following are some software requirements.

1.Operating system: Windows 10

2.Languages: PYTHON Django, html, CSS, JavaScript

3.Tool: VS Code

3.5 System analysis

The project endeavours to bridge the communication gap between deaf individuals and the general population by harnessing advancements in web applications, machine learning, and natural language processing technologies. Its core objective is to develop a real-time interface capable of converting audio inputs into corresponding sign language for deaf individuals. This involves utilizing speech-to-text APIs for audio conversion, natural language processing algorithms to analyze the semantics and grammatical structure of the text, and a sign language generation module to translate the processed text into appropriate sign language gestures. Adhering to the rules of Indian Sign Language (ISL) and its grammar guidelines, the interface operates within the scope of web applications, aiming to provide an accessible platform for both deaf users and the wider population. Stakeholders include deaf individuals seeking effective communication tools, the general populace desiring to interact inclusively, and developers responsible for the design, implementation, and maintenance of the interface. Testing, validation, and ongoing maintenance ensure the accuracy, usability, and accessibility of the system, ultimately facilitating social, emotional, and linguistic development within the deaf community.

3.6 EXISTING SYSTEM

The dumb people, sign language is the only way of communication. With the help of sign Language, physically impaired people express their thoughts to the other people.It is difficult for Common people to understand the specific sign language therefore communication becomes a Difficult. The sign language recognition has become an empirical task, as it consists of various Movements and gesture of the hands and therefore getting the right accuracy at a low-cost is a Mammoth task. Existing solutions are we have physical devices and software which can convert Audio to sign language but using Natural Language Processing we are improvising the tool. The Word library can be expanded to include most of the commonly used words in english. Speech to Text conversion can be made more accurate and text processing can be optimized using various NLP algorithms.

Limitations of the System:

- Efficiency is a concern with the current system, as it may not operate optimally in terms of speed and performance.

- Available resources for sign language recognition are limited, constraining the system's capabilities and scalability.
- Dependability is compromised due to inaccuracies in the conversion process, undermining the reliability of the system for effective communication.

3.7 PROPOSED SYSTEM

The aim of proposed system is to develop a system of improved facilities. The proposed System can overcome all the limitations of the existing system. Initially, we take audio as input on a Personal Digital Assistant by utilizing the python PyAudio module. Next, we are converting the audio to text using the Google Speech API presently utilizing NLP i.e. Natural is language processing we Breakdown the text into smaller, simpler and understandable. To avoid all these limitations and use Make the working more accurately the system needs to be implemented efficiently. We have with a Reliance parser for analysing the grammatical structure of the sentence and building up connection Between words. Finally, we converted audio into Sign language.

3.7.1 ADVANTAGES OF THE PROPOSEDSYSTEM

The system is very simple in design and to implement. The system requires very low System resources, and the system will work in almost all configurations. It has got following Features

- To provide information access and services to deaf people in Indian sign language.
- To develop a scalable project which can be extended to capture Whole vocabulary of ISL through manual and non-manual signs.
- To improve the physical and mental well-being of the specially abled people and improve their overall quality of life.

CHAPTER 4

IMPLEMENTATION

4.1 Introduction

System design marks the pivotal transition from conceptualization to tangible realization in the development phase of any engineered product or system. It entails the strategic application of diverse methodologies and principles to meticulously define the structure of a device, process, or system with adequate precision to enable its physical implementation. This initial phase bridges the gap between identifying the problem and crafting a solution. It operates as an iterative journey wherein requirements are translated into a comprehensive blueprint, laying the groundwork for the construction of the initial software iteration. This blueprint offers a comprehensive depiction of the software, encompassing a high-level abstraction of its functionality, behavior, and requirements. System design emerges as a creative endeavor, involving the innovation and refinement of inputs, databases, offline resources, client-server communication protocols, and data processing methodologies to yield meaningful outputs that align with organizational objectives.

4.2 Input Design

The input design must be made to make data entry easy, logical and error free as possible. The approaches for entering data into websites are text box and buttons with formatting and prompts. A button is an input provider. A text box is where user can enter the data. There is a mic button, which helps to record the speech and display it in the text box the system usually displays the processed text and keywords.

4.3 Output Design

It is the most important and direct source of information to the user. Efficient, intelligible output design improves the system relationship with the user and helps the decision making. Computer outputs are the most important and direct source to the user. An efficient output of the system improves the interaction of the system with the user and it provides his/her required information. In the context of speech-to-ISL translation, output design play a pivotal role in ensuring effective communication between the deaf and the normal people. It provide the translation of keywords to Indian sign language. Specialized output design include high-resolution displays capable of rendering ISL animations and graphics with clarity and precision.

4.4 Module Specification

There are five main modules in the system.

1. User Module
2. Speech Recording Module
3. Translation Module
4. NLTK Module
5. Sign Language Display Module

4.4.1 User Module:

Within this integral module, users embark on their journey by registering with the application, a process that entails providing essential credentials and creating a secure login with a valid username and password. Once authenticated, users gain access to a plethora of features, including cutting-edge functionalities such as speech-to-text conversion and sign language prediction, ensuring a comprehensive and enriching user experience.

4.4.2 Speech Recording Module:

At the core of user interaction lies the sophisticated Speech Recording Module, which seamlessly integrates the Google Speech-to-Text conversion library to translate spoken words into textual form. Leveraging this powerful library, the application swiftly processes the captured voice data, paving the way for subsequent stages of processing. Following conversion, the textual data undergoes further refinement through NLTK processing, ensuring accuracy and coherence before being presented to the user for review.

4.4.3 Translation Module:

This Django view function, named "translate," is designed to handle translation requests submitted via a form. It first checks if the request method is POST and if the form is valid. If the source and target languages are the same, it displays the original text as a success message. If both languages are not English, it translates the text twice, first to English and then to the target language, using IBM Watson Language Translator service, and displays the final translation. If the source language is English and the target language is Malayalam, or vice versa, it directly translates the text using the specified language pair. Finally, if the source and target languages are different and neither is English, it performs a direct translation using the specified language pair. After processing, it renders a form template with the appropriate form for input.

4.4.4 NLTK Module:

Diving deeper into the intricacies of language processing, the NLTK Module plays a pivotal role in refining and enhancing the textual data obtained from the Speech Recording Module. Through meticulous preprocessing techniques, including the removal of stop words and the extraction of essential keywords, the module ensures that the processed text is primed for further analysis and interpretation. The refined text is then seamlessly transmitted to the subsequent stage, where it serves as a cornerstone for retrieving relevant sign language videos from the system.

4.4.5 Sign Language Display Module:

In a visually captivating display of innovation, the Sign Language Display Module brings the processed textual data to life through dynamic sign language videos sourced from the system. Drawing upon the keywords extracted during NLTK processing, the module retrieves and showcases relevant sign language videos tailored to the user's input. Upon clicking the submit button, users are treated to a visually immersive experience as the system seamlessly integrates text and sign language, fostering inclusivity and accessibility in communication.

4.5 SYSTEM ARCHITECTURE

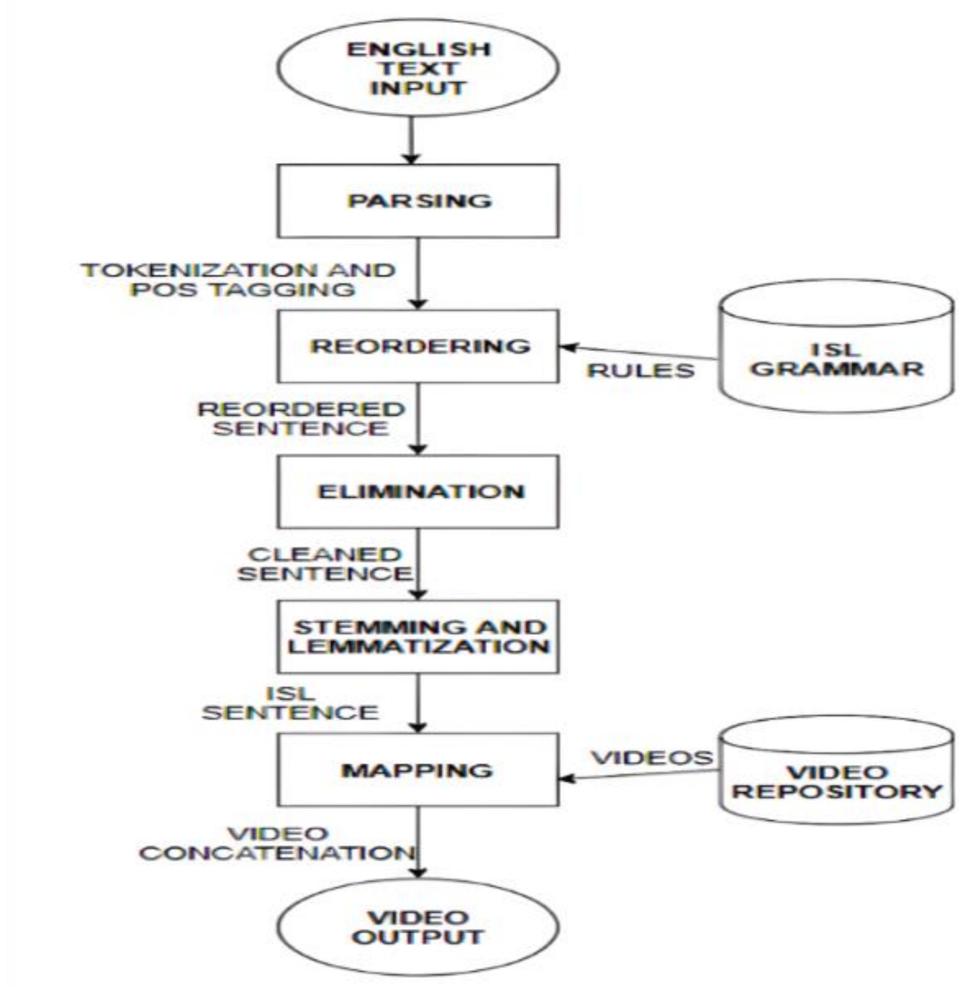

Figure 4.5 System Architecture

This system architecture depicts a comprehensive process for converting English text into a video output, specifically designed for Indian Sign Language (ISL). It starts with English text input which undergoes parsing and tokenization with POS tagging. The sentence is then reordered according to predefined rules and ISL grammar. After elimination to clean the sentence, it is further processed through stemming and lemmatization to form an ISL sentence. This sentence is then mapped to videos from a video repository, and finally, video concatenation is performed to produce the video output. This flowchart illustrates the steps involved in creating a video that conveys the meaning of the original English text in ISL, ensuring accessibility for the deaf community.

4.6 Use Case Diagram

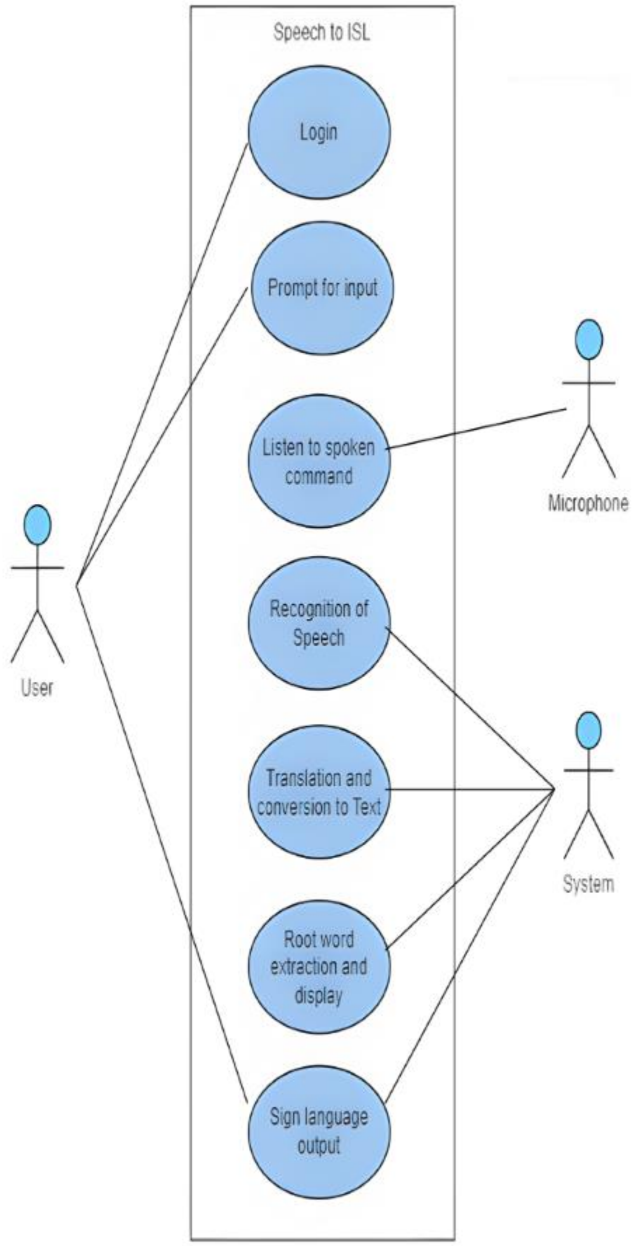

Figure 4.6 Use Case Diagram

This use case diagram illustrates a system designed to convert spoken language into Indian Sign Language (ISL). The process begins with the user logging in to the system. Once logged in, the system prompts the user for input. The user then speaks into a microphone, and the system recognizes the speech. This spoken command is translated and converted into text, followed by root word extraction. Finally, the system outputs the information in Indian Sign Language, completing the translation process.

4.7 Dataflow Diagram

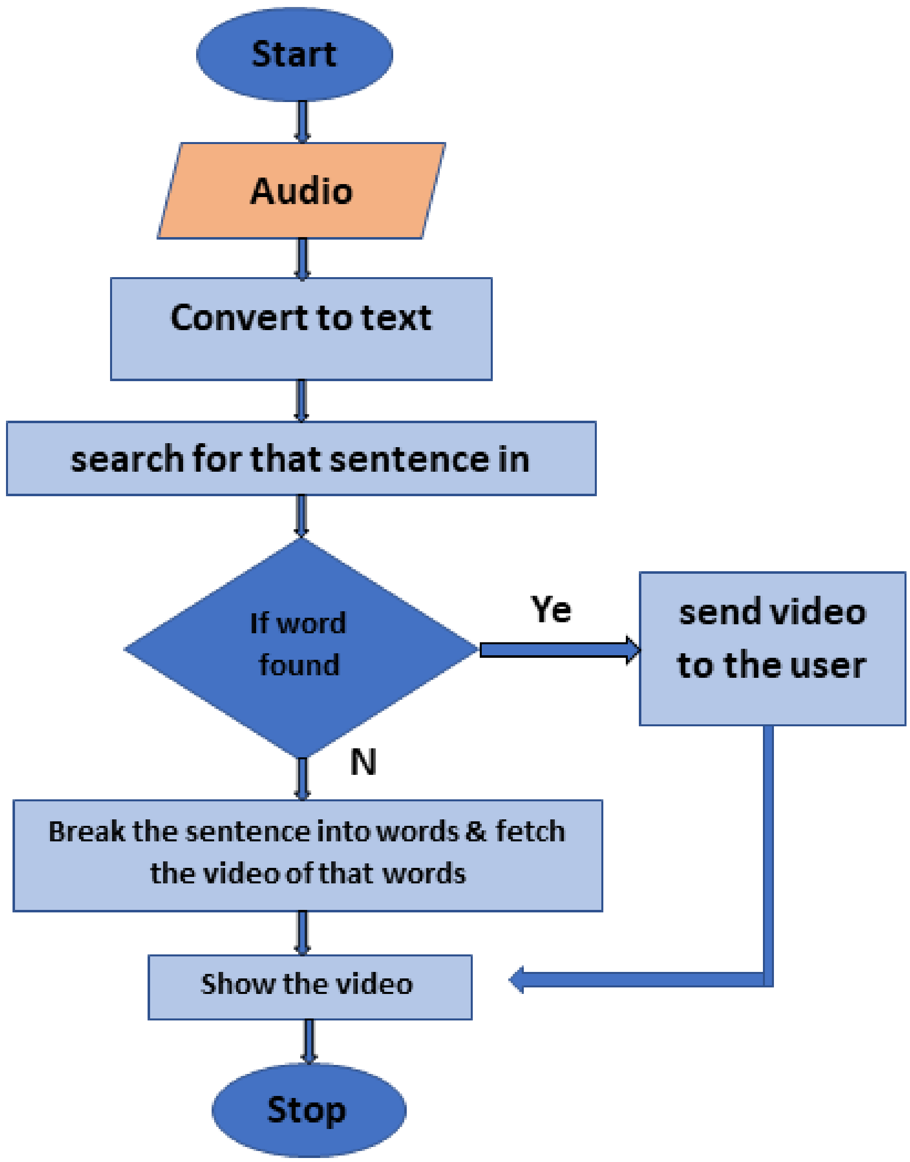

Figure 4.7 Data Flow Diagram

The dataflow diagram that outlines the process of converting audio input into a video presentation. It begins with an audio input that is converted to text. This text is then searched within a dataset to find a corresponding sentence. If the sentence is found, a related video is sent to the user. If the sentence is not found in the dataset, the process involves breaking down the sentence into individual words and fetching videos for those words, which are then compiled and shown to the user. The flowchart is designed to guide through the decision-making process in an audio-to-video conversion system, ensuring that the user receives a video output that matches the original audio input as closely as possible.

4.6 UML Diagram

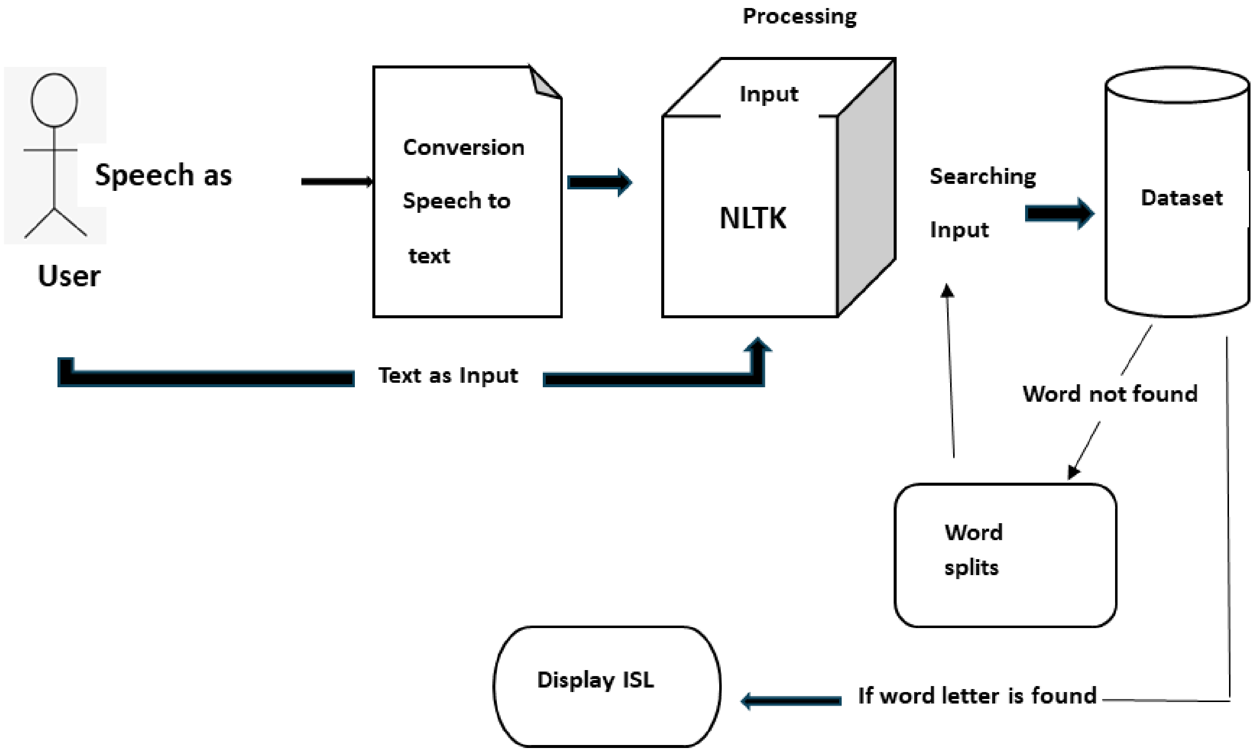

Figure 4.8 UML Diagram

A process that begins with a user's speech, which is then converted into text. This text is processed using the Natural Language Toolkit (NLTK) for searching within a dataset. If the search is unsuccessful, the word is split and further analyzed.

CHAPTER 5
RESULTS AND DISCUSSIONS

Screenshots from working model

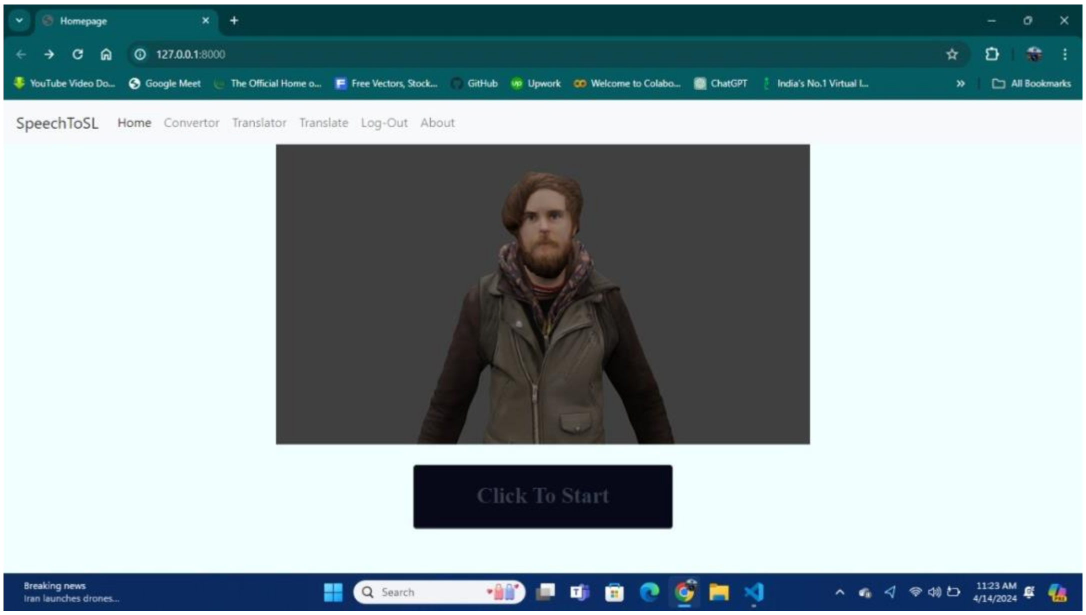

Figure 5.1 Home Page

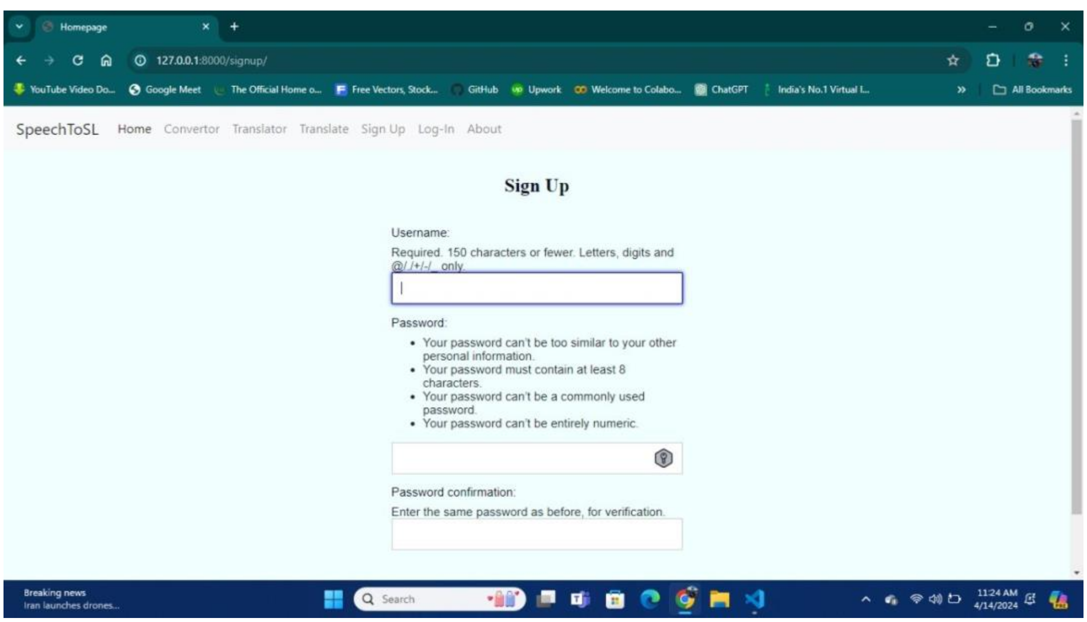

Figure 5.2 Sign Up Page

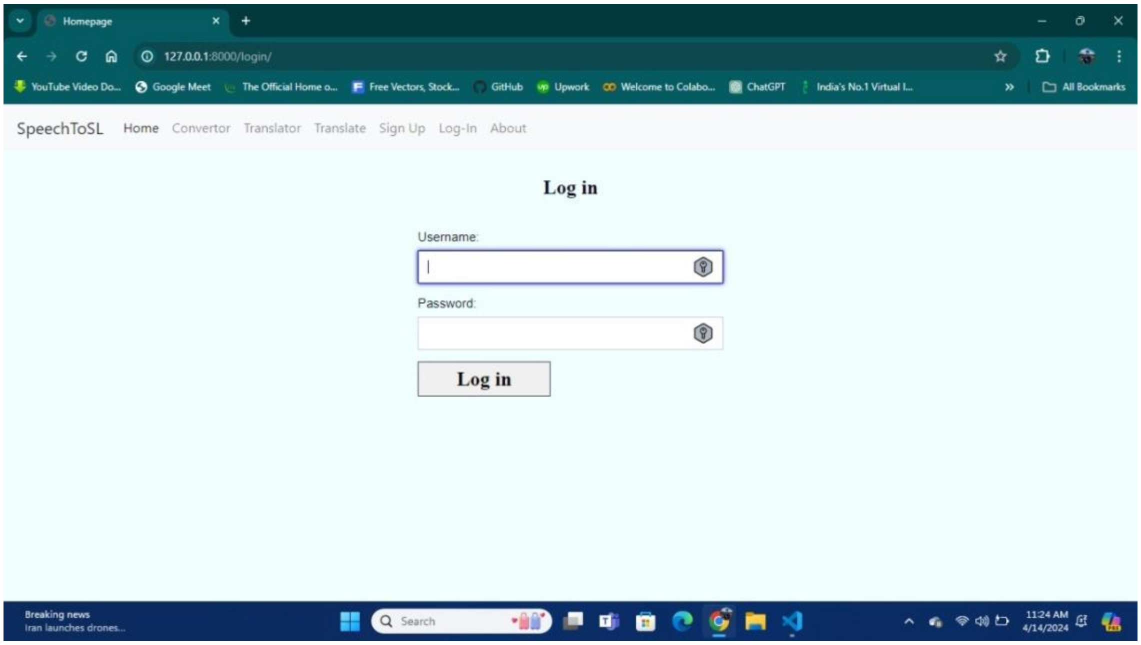

Figure 5.3 Login Page

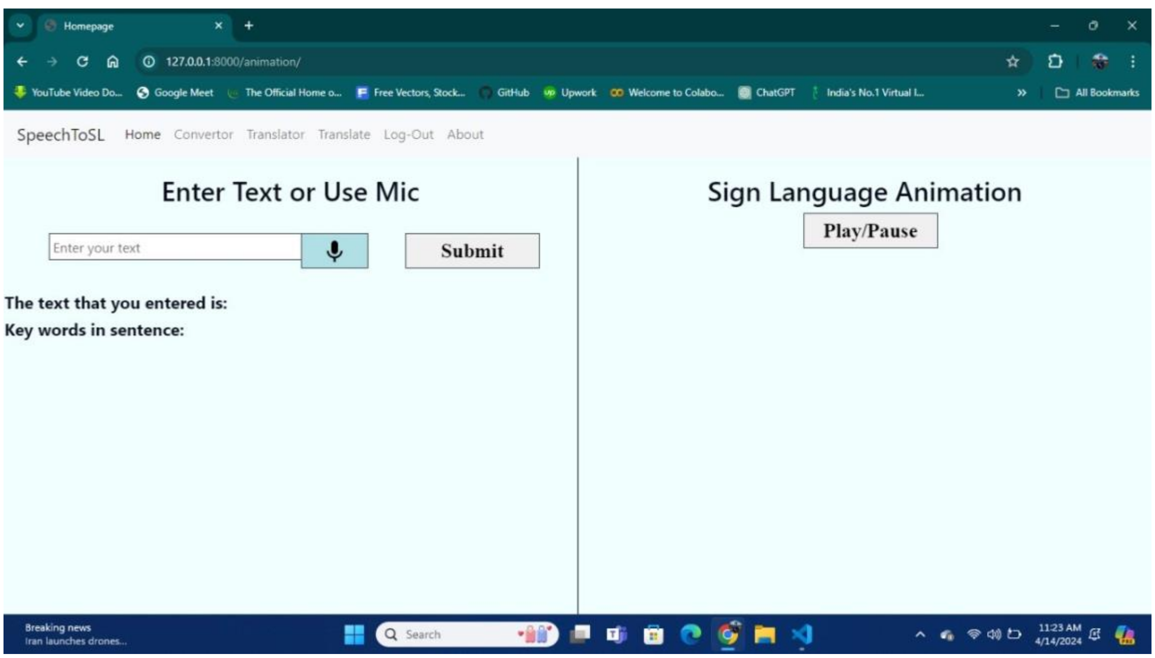

Figure 5.4 Speech to Sign Page

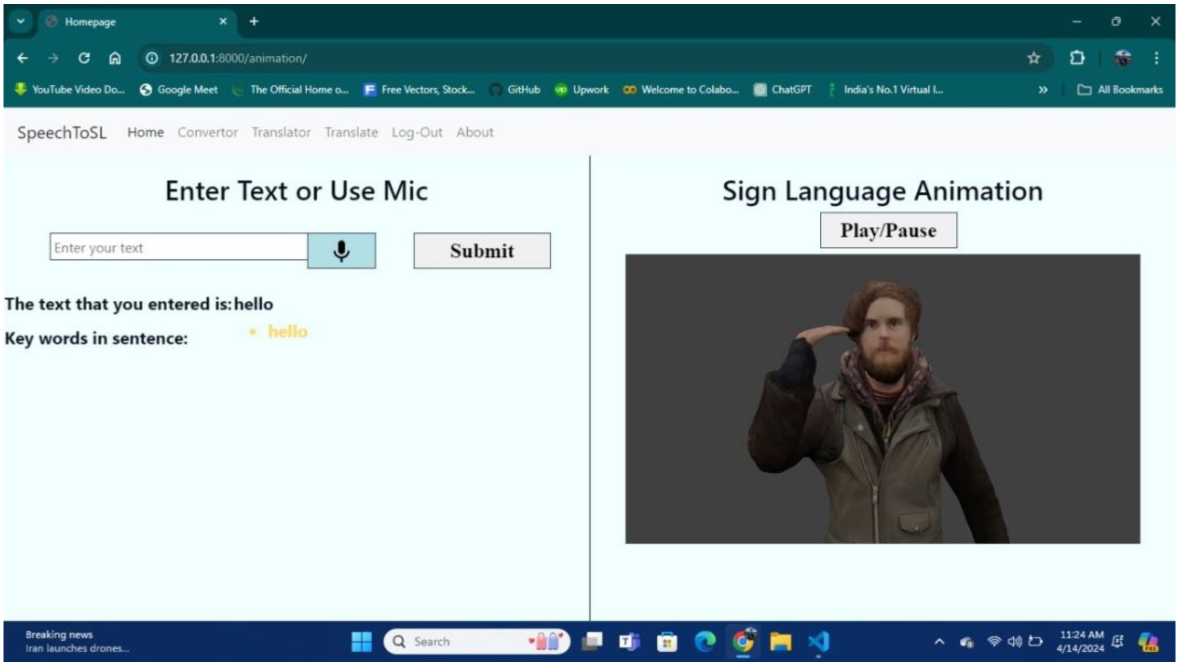

Figure 5.5 Sign Language Animation

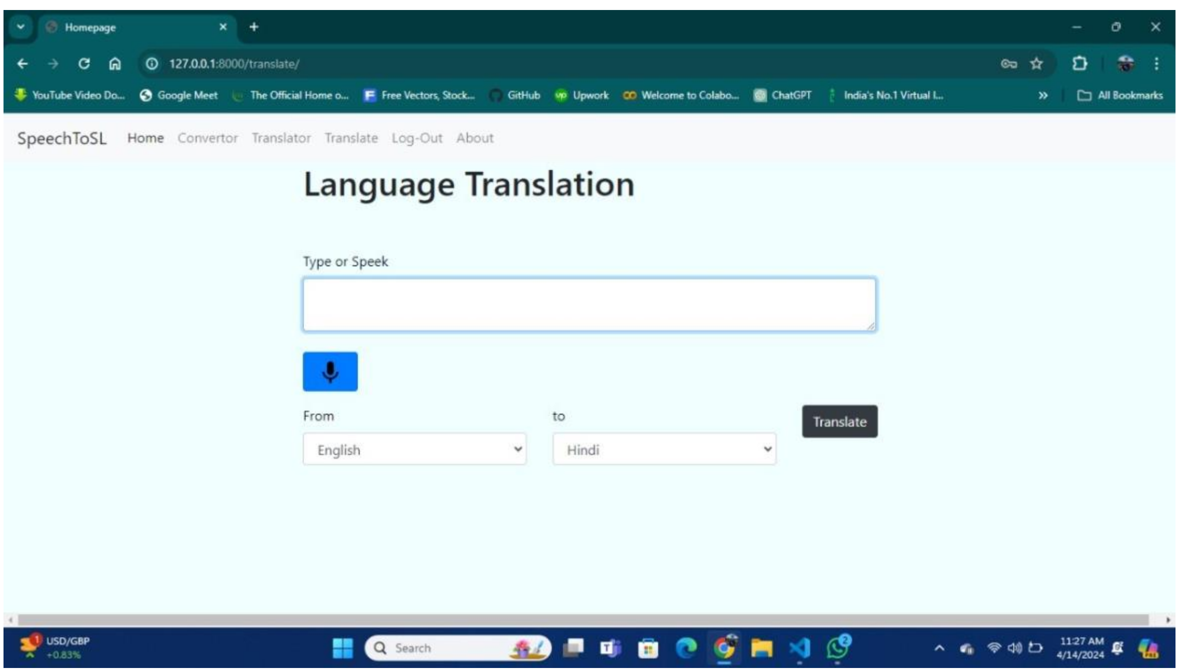

Figure 5.6 Language Translation Page

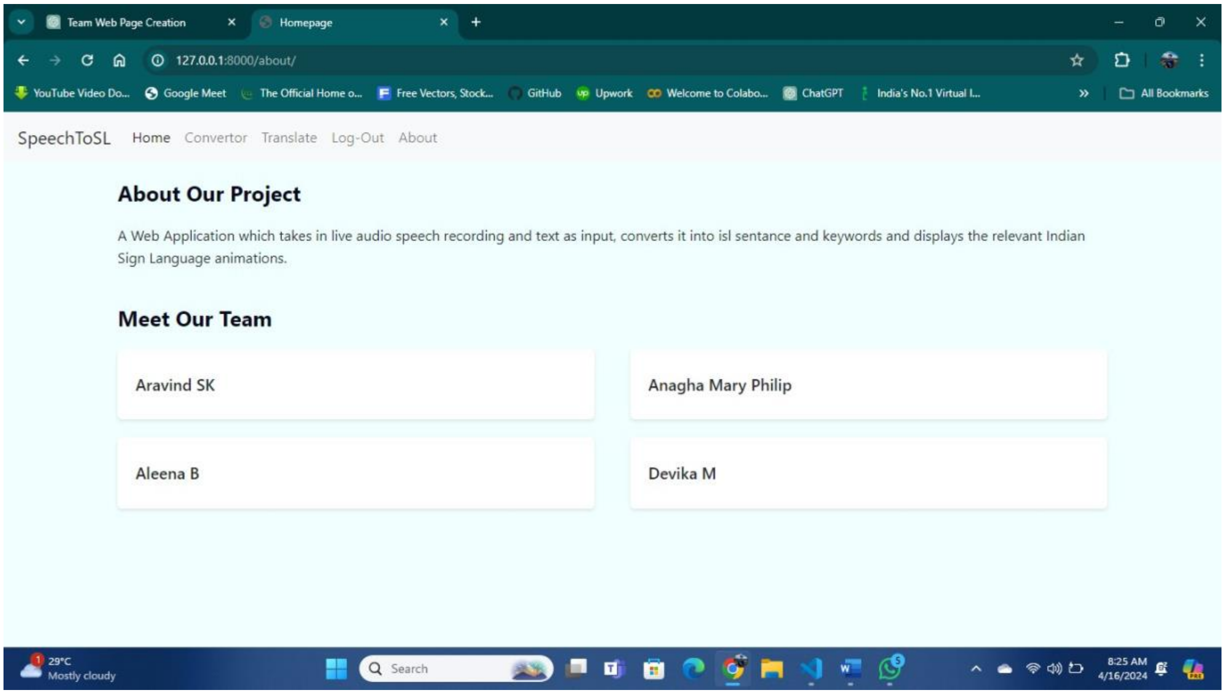

Figure 5.7 About Page

CHAPTER 6
CONCLUSION AND FUTURE SCOPE

The significance of a sign language translator transcends myriad sectors, spanning from educational institutions to healthcare facilities, from transportation hubs to legal arenas, and beyond. Its profound versatility renders it indispensable, facilitating seamless communication in environments as diverse as schools, colleges, hospitals, universities, airports, and courts. By bridging the linguistic gap between those proficient in sign language and those unfamiliar with it, this transformative system fosters smoother interactions among individuals with varying levels of hearing ability, thereby promoting greater understanding and inclusivity.

Recognizing the unique needs of the impaired community and offering tailored solutions through such technological advancements holds boundless potential to effect profound and lasting change. Beyond mere communication facilitation, the implementation of sign language translators represents a fundamental commitment to enhancing the overall well-being—both physical and mental—of individuals with disabilities. Through the promotion of inclusivity and accessibility, we collectively strive to uplift the quality of life for the specially abled, thereby contributing to the creation of a more equitable, empathetic, and compassionate society dedicated to the principles of diversity and mutual respect.

In our future endeavours, we aim to develop a cost-effective application tailored for news channels, offering an innovative solution to replace human interpreters with automated sign language interpretation. This advancement not only enhances efficiency but also reduces operational costs significantly. While current implementations, such as those employed in some News channels, rely on human interpreters to convey signs corresponding to the spoken news, our proposed solution offers a more economically viable alternative. By leveraging automation and advanced technology, we can streamline the process of sign language interpretation, eliminating the need for human intervention and thereby reducing expenses for news channels. This transition towards automation aligns with our commitment to providing accessible communication solutions while ensuring sustainability and cost-effectiveness in the long term. Looking ahead, we envision our system revolutionizing the broadcasting Industry by facilitating seamless and inclusive communication experiences for deaf viewers worldwide.

REFERENCES

[1] Chirag Garg, Gautam Priyadarshi Attry, Dhruv Mittal and Dr. Aruna Bhat4 “Conversion of Native Speech into Indian Sign Language to Facilitate Hearing Impairment” 2023 4th International Conference for Emerging Technology (INCET)979-8-3503-3575-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/INCET57972.2023.10170244

[2] Jashwanth Peguda,V Sai Sriharsha Santosh,Y Vijayalata,Ashlin Deepa R N,Vaddi Mounish”Speech to Sign Language Translation for Indian Languages”2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)

[3] Bandi Rupendra Reddy, Darukumalli Sai Tharun Reddy, Sandeep Preetham M And Susmitha Vekkot”Creation of GIF dataset and implementation of a Speech-to-sign language translator in Telugu”2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon)|978-1-6654-5342-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/NKCON56289.2022.1012706

[4] Sugandhi, Parteek Kumar, and Sanmeet Kaur"Indian Sign Language Generation System"PUBLISHED BY THE IEEE COMPUTER SOCIETY MARCH 2021

[5] Ahn, H., Ha, T., Choi, Y., Yoo, H., & Oh, S. (2018). Text2action: Generative adversarial synthesis from language to action. In IEEE international conference on robotics and automation (ICRA).

[6] Anup Kumar, Karun Thankachan and Mevin M. Dominic”Sign Language Recognition”3rd InCI Conf. on Recent Advances in Information Technology I RAIT-20161.

[7] Nayana J, Suparna Bhat, Rekha R Nair, Tina Babu"Audio to Sign Language conversion using Natural Language Processing"2022 3rd International Conference on Communication, Computing and Industry 4.0 (C2I4) | 978-1-6654-7497-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/C2I456876.2022.10051414.

[8] Ezhumalai P , Raj Kumar M, Rahul A S, Vimalanathan V , Yuvaraj A"Speech To Sign Language Translator For Hearing Impaired"Turkish Journal of Computer and Mathematics Education Vol.12 No.10 (2021), 1913-1919.

[9] Prof. Abhishek Mehta, Dr. Kamini Solanki , Prof. Trupti Rathod "Automatic Translate Real-Time Voice to Sign Language Conversion for Deaf and Dumb People"International Journal Engineering Research & Technology (IJERT) ISSN: 2278-0181 Published by, www.ijert.orgICRADL – 2021 Conference Proceedings

[10] Language Processing"2022 3rd International Conference on Communication, Computing and Industry 4.0 (C2I4) | 978-1-6654-7497-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/C2I456876.2022.10051414

[11] Sreeraksha M R, Vani H Y,Phani Bhushan,D K Shivkumar"Speech To Sign Language Conversion using Convolutional Neural Networks"International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Published by, www.ijert.org NCCDS – 2021 Conference Proceedings Volume 9, Issue 12 Special Issue – 2021

[12] Ankita Harkude, Sarika Namade, Shefali Patil, Anita Morey "Audio to Sign Language Translation for Deaf People"ISSN: 2277-3754 ISO 9001:2008 Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 9, Issue 10, April 2020

[13] Heet D. Patel, Avneet Saluja"Sign language recognition and translator application"International Research Journal of Engineering and Technology (IRJET)

e-ISSN: 2395-0056 Volume: 08 Issue: 09 | Sep 2021 www.irjet.net p-ISSN: 2395-0072

[14] Vardan Gupta , Saumya Sinha, Puneet Bhushan and Minal Shettigar”English Text to Indian Sign Language Translator”

[15] Ms. Satvika Reddy Satti, Mrs.P.Pavithra”Audio or Text to sign language converter”2023 JETIR October 2023, Volume 10, Issue 10 www.jetir.org (ISSN-2349-5162)

[16] Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1724–1734). Association for Computational Linguistics.

[17] Chung, J., Gülçehre, Ç., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. CoRR arXiv:1412.3555.

[18] Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., & Abbott, S. (2002). Tessa, a system to aid communication with deaf people. In Proceedings of the 5th international ACM conference on assistive technologies (pp. 205–212). ACM

[19] Ebling, S., Camgoz, N.C., Braem, P., Tissi, K., Sidler-Miserez, S., Stoll, S., Hadfield, S., Haug, T., Bowden, R., Tornay, S., Razavi, M., & Magimai-Doss, M. (2018). Smile Swiss German sign language dataset. In 11th Edition of the language resources and evaluation conference (LREC).

[20] Ebling, S., & Glauert, J. (2013). Exploiting the full potential of Asigning to build an avatar signing train announcements. In 3rd International symposium on sign language translation and avatar technology.

[21] Glauert, J., Elliott, R., Cox, S., Tryggvason, J., & Sheard, M. (2006).VANESSA-A system for communication between Deaf and hearing people. Technology and Disability, 18(4), 207–216.

[22] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley,D., Ozair, S., Courville, A. C., & Bengio, Y. (2014). Generative Adversarial nets. In Advances in neural information processing Systems 27: Annual conference on neural information processing Systems 2014, December 8-13 2014 (pp. 2672–2680). Montreal,Quebec.

[23] Gregor, K., Danihelka, I., Graves, A., Rezende, D., & Wierstra, D.(2015). Draw: A recurrent neural network for image generation.In F. Bach, & D. Blei (Eds.) Proceedings of the 32nd international conference on machine learning, Proceedings of Machine Learning Research (Vol. 37, pp. 1462–1471). PMLR, Lille.

[24] Guo, D., Zhou, W., Li, H., & Wang, M. (2017). Online early-late Fusion based on adaptive hmm for sign language recognition. ACM Transactions on Multimedia Computing, Communications, and Applications, 14(1), 8:1–8:18. https://doi.org/10.1145/3152121.

[25] Guo, D., Zhou, W., Li, H., & Wang, M. (2018). Hierarchical LSTM for Sign language translation. In AAAI.Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory.Neural Computation, 9, 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735