Accessible AI : Enhancing Communication Accessibility

Digital content to Indian sign language

- *Team-Nooglers*

***

Abstract

The "Accessible AI" is an innovative web application designed to bridge the communication gap between hearing individuals and the deaf or hard-of-hearing community. This application leverages advanced speech recognition, natural language processing, and video processing technologies to convert spoken or written language into Indian Sign Language (ISL). The application supports multiple input options, including YouTube URLs, local video or audio files, and direct text input. The backend processes involve extracting audio, transcribing speech-to-text using Google's Speech-to-Text API, parsing and processing the text, and converting the processed text into ISL. The final output is a video that visually represents the input in sign language.

By utilizing tools such as Streamlit for the user interface, Google Cloud services for speech recognition and storage, and various Python libraries for audio and video manipulation, the application provides an efficient and user-friendly solution for translating spoken or written content into ISL.

Keywords *Indian Sign Language (ISL), Google's Speech-to-Text API.*

INTRODUCTION

In upcoming digital communication, our application covers the development of a system that converts digital content into Indian Sign Language (ISL).Our application addresses the communication barriers faced by the deaf and hard-of-hearing community, providing a tool that can bridge the gap between spoken and sign languages.

Our application aims to leverage the power of modern technologies such as natural language processing (NLP), machine learning, and cloud-based services to create a seamless and efficient conversion system. The system is designed to process both YouTube videos and local audio/video files, as well as direct text input, and convert the spoken or written content into ISL videos. By utilizing Google's Speech-to-Text API, the system transcribes spoken language into text. This text is then processed using advanced NLP techniques to ensure accurate and contextually appropriate translation into ISL. The project also incorporates the use of a vast library of sign language videos and gifs to create coherent and comprehensible ISL outputs.

The development of this system not only showcases the potential of integrating various technologies to solve real-world problems but also emphasizes the importance of accessibility in technological advancements. This project is a step towards making digital content more accessible to the deaf and hard-of-hearing community, promoting inclusivity, and enhancing communication in a diverse society.

Methodology

1. Dataset Description

The effectiveness and accuracy of the Speech-to-ISL conversion system heavily rely on the quality and comprehensiveness of the datasets used.

- The "**NLP_videos.csv"** dataset which helps in the text-to-ISL conversion process. It contains a detailed list of words along with their corresponding video URLs, start times, and end times, which are used to generate the ISL video clips.
- The "**alphabet**" dataset is also an component of dataset in which it helps in text-to-ISL conversion process, particularly for spelling out words or handling characters that do not have direct ISL video representations. This dataset contains GIF files representing each letter of the alphabet in ISL

2. Speech-to-text conversion

- The speech-to-text conversion component is integral to the project, transforming audio input into transcribed text that can be processed further for conversion into Indian Sign Language (ISL). This process involves several key steps, including audio extraction, uploading to Google Cloud Storage, and using Google's Speech-to-Text API to generate transcriptions.
- The first step in the speech-to-text conversion process is audio extraction. Depending on the input type, which could be a YouTube URL, a local video file, or an audio file, the system extracts the audio component. For YouTube videos, the yt-dlp library is used to download the audio track. If the input is a local video file, FFmpeg is utilized to extract the audio. The extracted audio is saved in WAV format, which is necessary for the subsequent processing steps.

- Once the audio is prepared, it is uploaded to Google Cloud Storage. This step involves creating a bucket in the storage service and uploading the audio file to this bucket. By storing the audio file in the cloud, the system can leverage

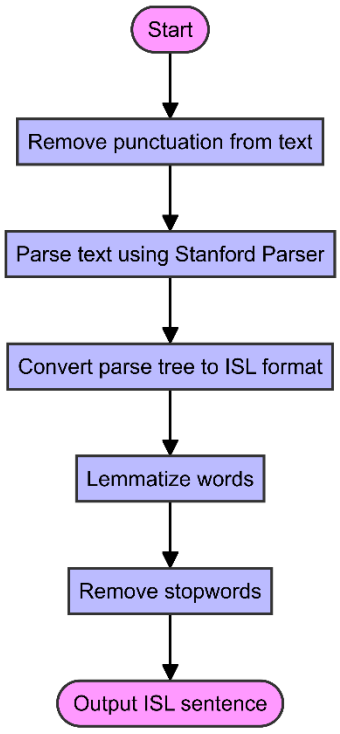
Google's powerful Speech-to-Text API, which requires the audio input to be accessible via a Google Cloud Storage URI

- The transcription process involves the API analyzing the audio and generating a text representation of the spoken words. This process can take some time, depending on the length and quality of the audio file. The API returns a structured response containing the transcription results, including alternatives for each segment of speech.

3. Natural Language Processing (NLP)

Text Parsing

- The first step in the NLP pipeline is text parsing. For this, we utilize the Stanford Parser, a robust tool capable of generating parse trees for sentences. These parse trees represent the grammatical structure of sentences, showing the relationships between words and phrases. The parser breaks down the transcribed text into its constituent parts, such as nouns, verbs, and adjectives, and organizes these parts into a hierarchical tree structure.

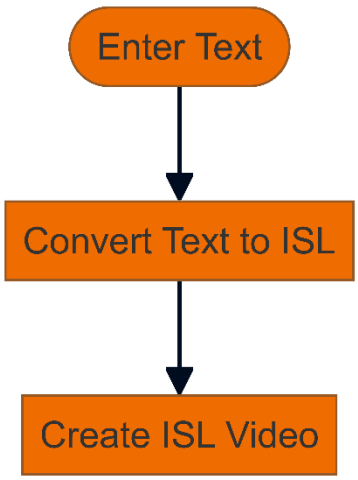


- Once we have the parse trees, the next step is tree manipulation. This involves rearranging the syntax of the parsed text to align with ISL grammar, which differs significantly from English grammar. ISL often follows a topic-comment structure, where the topic of the sentence is presented first, followed by the comment or action related to the topic. By reordering the elements of the parse tree, we can create a syntactic structure that accurately reflects ISL grammar.
- Following tree manipulation, the text undergoes lemmatization. This process reduces words to their base or root forms using the WordNet Lemmatizer. Lemmatization is essential for handling different inflected forms of words, ensuring that each word is represented in its simplest form. For example, verbs in different tenses and nouns in different cases are converted to their base forms, such as "running" to "run" and "geese" to "goose." This simplification helps in creating a more consistent and easily interpretable dataset for the ISL conversion process.

4. Text-to-ISL conversion

- The Text-to-ISL conversion component is a crucial part of the project, responsible for transforming the transcribed text into Indian Sign Language (ISL) video content. This process involves several steps, including text preprocessing, ISL grammar adaptation, and the generation of video sequences.

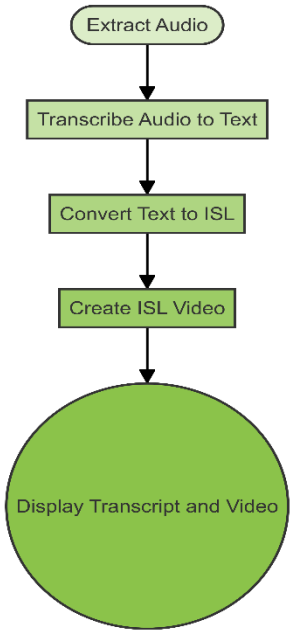
- The first step in the text-to-ISL conversion process is text preprocessing. This involves cleaning the transcribed text to ensure it is suitable for further processing. The text is tokenized, splitting it into individual words and punctuation marks. This is necessary because ISL, like other sign languages, has a different grammatical structure compared to spoken English. The tokenized text is then processed to remove any extraneous elements such as numbers, special characters, and non-essential punctuation.



- Next, the system adapts the cleaned text to match ISL grammar. ISL grammar differs significantly from English, often following a subject-object-verb (SOV) structure rather than the subject-verb-object (SVO) structure of English. With the text adapted to ISL grammar, the system proceeds to generate video sequences. This involves mapping each word in the adapted text to its corresponding sign in ISL

- The final step in the text-to-ISL conversion process is the synthesis of the video content. The system compiles the individual video clips, including non-manual signals and timing adjustments, into a single coherent video

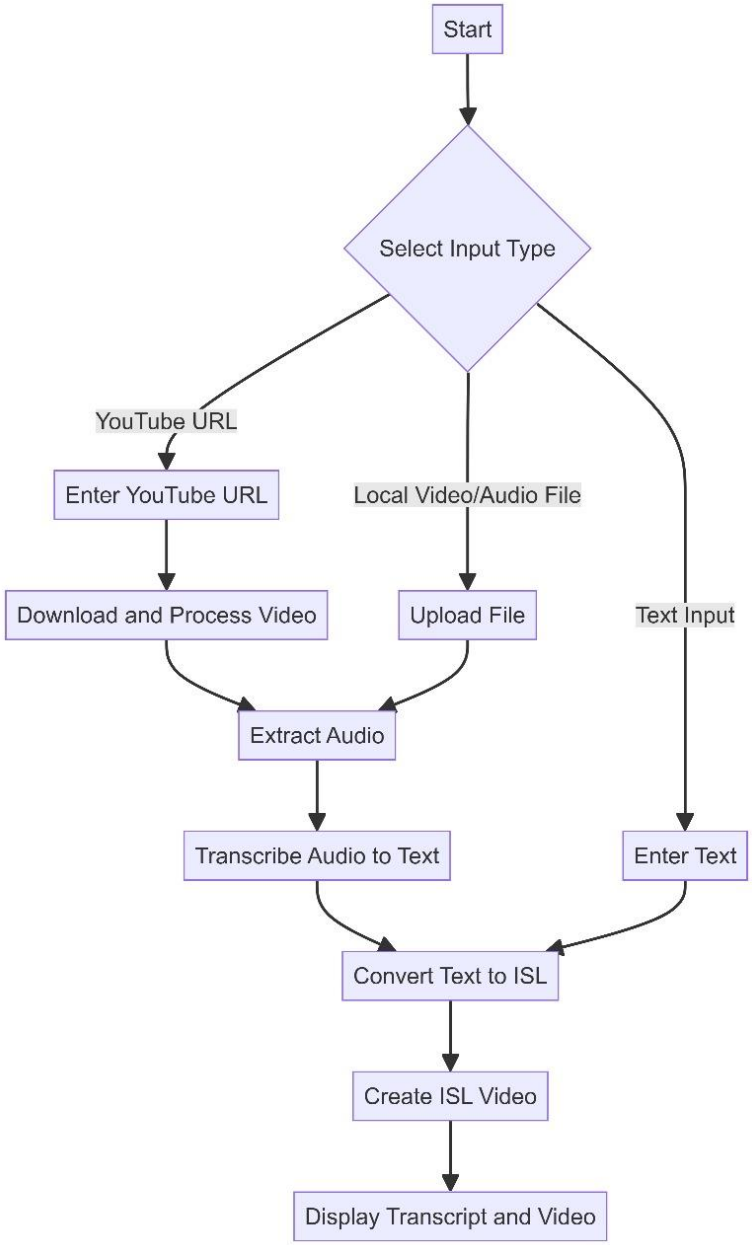
5. Video Processing



- The Video Processing component is responsible for synthesizing the final output video, which accurately represents the input speech in Indian Sign Language (ISL). This process involves multiple stages, including video generation, non-manual signal integration, timing adjustments, and final compilation.
- The initial stage of video processing is video generation, where the system translates the adapted text into a sequence of ISL signs. This is achieved using a pre-recorded library of ISL signs, which contains video clips of a signer performing each sign. The system maps

each word or phrase in the adapted text to its corresponding sign and retrieves the appropriate video clip from the library. For words or phrases without a direct sign, the system employs fingerspelling or contextual interpretation to ensure accurate representation. These individual clips are then arranged in sequence to form a video

How It works


Start
Select Input Type
YouTube URL
Local Video/Audio File
Text Input
Enter YouTube URL
Download and Process Video
Upload File
Extract Audio
Transcribe Audio to Text
Enter Text
Convert Text to ISL
Create ISL Video
Display Transcript and Video

Conclusion

In conclusion, the Speech-to-Indian Sign Language (ISL) conversion system represents a significant advancement in bridging the communication gap between the hearing and the deaf or hard-of-hearing communities. This Application covers technologies in speech recognition, natural language processing, and video synthesis to deliver an end-to-end solution capable of translating spoken English into fluent ISL.

The Application commenced with a comprehensive design phase, during which each component's functionality was well planned and integrated. The speech-to-text conversion module, powered by Google's ASR, demonstrated high accuracy in transcribing spoken words into text, even across various accents and speaking speeds. This transcription forms the foundation for subsequent processing stages, ensuring a reliable input for translation.

Natural Language Processing (NLP) techniques were employed to refine the transcribed text, adapting it to the grammatical and syntactic structures of ISL. Text parsing using the Stanford Parser, combined with lemmatization and stopword removal, ensured that the text was both concise and semantically appropriate for translation into ISL.

In summary, this project not only highlights the technical feasibility of automatic Speech-to-ISL translation but also underscores its impact on accessibility. The successful implementation of this system marks a step forward in making communication more accessible for all, fostering a more inclusive society.