AI POWERED SPEECH RECOGNITION AND TEXT

TRANSLATION TO OTHER LANGUAGE

Dr. Kavitha Soppari

Assoc. Professor and Head

Dept. of CSE (AI & ML)

ACE Engineering College

Hyderabad, India

 Mrs. Jangam Bhargavi

Assistant Professor

Project Coordinator and Project Guide

Dept. of CSE (AI & ML)

ACE Engineering College Hyderabad, India

Ankarla Bhargavraj

Student, CSE-AI&ML Dept

ACE Engineering College

Hyderabad, India

Valaboju Bala Brahmam

Student, CSE-AI&ML Dept

ACE Engineering College

Hyderabad, India

Katta VarunTeja

Student, CSE-AI&ML Dept

ACE Engineering College

Hyderabad, India

**ABSTRACT**

This is an AI-based speech-to-text translation system designed for real-time tracking and translation across different languages. The model leverages advanced speech recognition and neural machine translation technologies to process spoken input and produce accurate text output in another language. The system has two main components: a robust Automatic Speech Recognition (ASR) module that transcribes spoken input, and an advanced neural translation model that translates the recognized text into a grammatically correct and properly localized variant of the target language. The primary goal of this system is to create a smooth speech-to-text translation interface, addressing language barriers in real time. This system has applications in various fields, including multilingual communication, education, and broadcasting, where real-time translation fosters greater understanding and collaboration. To ensure robust performance and easy deployment, special measures will be taken to address challenges such as variations in accent, background noise, and colloquial expressions.

**Keywords:** Speech Recognition, Microphone Input, Text-to-Speech (TTS), Ambient Noise Adjustment, Translation Source Language, Translation Target Language

**INTRODUCTION**

Communication across languages remains a significant challenge in our globally connected world, despite language being one of humanity's most valuable assets. Bridging linguistic gaps is essential for promoting worldwide dialogue, teamwork, and mutual comprehension. To tackle these issues, progress in artificial intelligence (AI) and machine learning (ML) has paved the way for advanced systems capable of surmounting these obstacles. Among these innovations, AI-powered speech-to-text translation platforms have emerged as game-changing solutions, able to process spoken words in real-time and deliver precise translations into other languages.

This system presents an AI-driven speech-to-text translation platform designed for instantaneous multilingual interaction. Harnessing state-of-the-art developments in Automatic Speech Recognition (ASR) and neural machine translation, the system is engineered to detect spoken input, transcribe it accurately, and convert the resulting text into a grammatically correct and contextually appropriate output in the desired language. This dual functionality of real-time transcription and translation addresses the need for effortless communication across diverse linguistic environments.

**Need for Real-Time Speech Translation**

The swift expansion of global commerce, learning, and media demands tools that facilitate seamless and accurate communication across language barriers. For example, international corporations require solutions to overcome linguistic challenges during gatherings, symposiums, and discussions. Educational facilities can likewise employ instantaneous translation technologies to foster inclusive learning spaces for linguistically diverse student populations. The broadcasting and media sectors also greatly benefit from systems enabling immediate multilingual interpretations, ensuring worldwide access to news, happenings, and entertainment content. Conventional translation methods often depend on written text as input, which introduces communication lags. While these techniques are suitable in certain contexts, they fall short in real-time verbal exchanges. AI-powered speech-to-text translation systems address this shortcoming by directly converting spoken language and delivering near-instantaneous translations.

**Evolution of Speech-to-Text and Translation Technologies**

Advancements in artificial intelligence, machine learning, and natural language processing (NLP) have propelled the evolution of speech-to-text and translation technologies. Initial systems employed rule-based algorithms, which faced challenges in handling the intricacies and diversity of human communication. Contemporary systems, leveraging deep learning models, have attained impressive levels of precision and adaptability, enabling practical applications that were previously confined to the realm of science fiction.

**Key Features of the AI-Based System**

The proposed AI-based speech-to-text translation system distinguishes itself through its accuracy, speed, and adaptability, seamlessly integrating Automatic Speech Recognition (ASR) and Neural Machine Translation (NMT) for an enhanced user experience; the ASR module, acting as the system's core, converts spoken input to text with high precision, trained on diverse datasets to handle accents, environmental noise, and provide real-time processing, while the NMT model builds upon the ASR output to translate text into the target language, preserving meaning and context by capturing nuances, ensuring grammatical accuracy, and adapting to regional linguistic and cultural preferences.

**ARCHITECTURE**



The architecture presented in the flowchart demonstrates a layered approach for a real-time Hindi-to-Telugu speech translation system. It begins with the Input Layer, where the user's spoken Hindi is captured by a microphone. This audio input is then processed by the Speech Recognition module within the Processing Layer, which converts it into Hindi text. Subsequently, the Translation Module within the same layer translates this Hindi text into Telugu. The Text-to-Speech module then synthesizes the translated Telugu text into audio output. The Interface Layer, implemented using the Tkinter library, provides a user-friendly GUI for interaction. This GUI displays the Hindi input, Telugu output, and status updates, allowing the user to control the translation process. The Control Layer manages the overall system flow, employing thread management for concurrent execution of tasks and state control to maintain system state and handle errors. This layered structure enhances modularity, maintainability, and testability, while the use of threads improves system performance and responsiveness.

**LITERATURE REVIEW**

**[1] Title: Direct Speech-to-Speech**

**Translation Using Machine Learning**

This study presents a novel method for translating speech. It transforms speech input into directly translated speech output, eschewing conventional text-based middlemen. Sirish Hong Limbu's research examines the potential of machine learning-driven systems that allow current
speech - speech is intended to define limitations on the pipeline for translation. particularly with regard to accumulation of errors and delays. The inputs and outputs of the system are displayed using spectrograms. Text elimination is a step in between using a sequential neural network design (Seq2Seq). The voice qualities and subtleties of emotion are preserved with this technique. Make the translation seem more natural. This is in spite of obstacles including data scarcity and alignment problems. However, this study demonstrates how effective direct translation systems may be. This proof of concept demonstrates the advantages of integrating neural machine translation (NMT) and automated speech recognition (ASR) technologies into a single framework to transform multilingual communication. The quality of translations keeps becoming better in the interim. Additionally, this method is
a solid basis for upcoming developments in affective, real-time, context-aware solutions.
The framework emphasizes lower latency. Improved preservation of speaker identity and flexibility in a multilingual setting. It offers guidance on how to handle the difficulties brought on by accent changes, background noise, and facial expressions. It establishes the groundwork for revolutionary uses in media, education, and international cooperation.

**[2] Title: Text-to-Speech Conversion**

The system's purpose is to translate textual information into speech. By extracting text from photos and turning it into speech, the system uses optical character recognition (OCR) techniques to help make content more accessible for users, including those who are visually impaired. The procedure entails scanning and reading the numbers and letters in English that are displayed in the picture. After that, it undergoes processing to generate the appropriate speech output.
The primary goal of this technology is to offer a practical and effective way to transform text to speech. Improves text content engagement for those who require or desire auditory feedback. Incorporates speech synthesis and OCR technologies. Natural correctness is guaranteed by the system.
Thus, the whole user experience is enhanced by the louder speech.
This method highlights how crucial accessibility is in technology.

**[3] Title: Human Languages in Source Code: Auto-Translation for Localized Instruction**

This paper presents Code International, a cutting-edge technology that improves accessibility in computer science education by translating source code between human languages.
Understanding that non-native English speakers may find programming that relies heavily on English to be a barrier, theto determine the frequency and trends of non-English usage in code, writers thoroughly examined 2.9 million Java repositories on GitHub. The necessity for localized programming tools was highlighted by the findings, which showed a large use of different human languages in code comments and identifiers. The researchers translated the Karel programming textbook into more than 100 languages to show off CodeInternational's educational value and help classrooms throughout the world adopt it. This method, which aims to reduce language barriers and encourage greater involvement in computer science, emphasizes the significance of linguistic inclusion in coding instruction.

**[4] Title: Multilingual Speech and Text Recognition and Translation Using Image**

This solution, which is especially made for desktop users, combines several features into a single program to overcome linguistic barriers. Voice recognition, language translation, speech synthesis, and visual translation are its four main modules. The program enables expressive and productive conversations by identifying spoken language and translating it into a target language specified by the user. It also takes inputs of written text and translates them into the target language. One noteworthy feature is its capacity to use optical character recognition (OCR) technology to extract text from photos, whether they were taken with a camera or saved in the system. After being retrieved, this text is translated and shown on the system screen, improving comprehension in many languages. By combining speech recognition, text translation, and image-based text extraction, the framework seeks to address the communication difficulties brought on by linguistic differences. The application provides desktop users with a flexible tool for multilingual communication by combining these features, guaranteeing that language barriers are successfully overcome by technology.

**ALGORITHM**

**Input Layer:**

**Microphone:** The system uses the operating system's built-in audio APIs to capture audio input from the microphone.

**Processing Layer:** Speech Recognition (using Google Speech Recognition):

**Acoustic Modeling:** This step converts the captured audio signal into a sequence of features like Mel-Frequency Cepstral Coefficients (MFCCs) that represent the speech characteristics.

**Language Modeling:** A statistical language model (like n-grams or Hidden Markov Models - HMMs) is used to predict the most probable sequence of words in Hindi based on the acoustic features and the grammar of the language.

**Decoding Algorithm:** Techniques like Viterbi decoding or beam search are employed to find the most likely sequence of words that aligns best with the acoustic features and the language model.

**Encoder-Decoder Architecture:** This is a common approach in NMT, where an encoder network processes the source Hindi sentence (represented as a sequence of word embeddings) and generates

a context vector. A decoder network then utilizes the context vector to generate the translated Telugu sentence word by word.

**Attention Mechanism:** Advanced NMT models often incorporate attention mechanisms to focus on relevant parts of the source sentence during translation, improving accuracy.

**Text-to-Speech (TTS):**

Unit Selection Synthesis or Statistical Parametric Synthesis: These techniques are commonly used for TTS. They involve storing pre-recorded speech units (phonemes or syllables) in a database and concatenating them to form the target speech output (Telugu in this case). Statistical models based on linguistic features might also be used to synthesize speech waveforms.

**Conclusion**

By enabling real-time multilingual communication, the AI-powered speech recognition and text translation system created in this project successfully overcomes language boundaries. The system records spoken input, turns it into text, and then quickly and accurately translates it into the target language by combining Automatic Speech Recognition (ASR) with Neural Machine Translation (NMT). This approach is very helpful in a number of industries where smooth communication is crucial, including media, healthcare, international business, and education. Reliable performance is ensured in real-world circumstances by the system's capacity to manage a variety of accents, environmental noise, and linguistic differences. With this project, we illustrated how AI can promote inclusivity and teamwork for a global audience. Scalability and future enhancements, such adding support for additional languages or improving contextual correctness, are made possible by the system's modular design and sophisticated algorithms. This system demonstrates the revolutionary potential of AI in improving human communication and represents a major step toward removing linguistic barriers. Future research and development in speech recognition and translation systems will be well-founded on the knowledge gathered from this effort.

**REFERENCES**

[1] Sireesh Haang Limbu, “Direct Speech to Speech Translation Using Machine Learning”, December 2020

[2] S. Venkateswarlu, D. B. K. Kamesh , J. K. R. Sastry and Radhika Rani, “ Text to Speech Conversion”, 23 September 2020

[3] Chris Piech, Sami Abu-El-Haija, “Auto-Translation for Localized Instruction”, Sep 2019

[4] Sagar Patil, Mayuri Phonde, Siddharth Prajapati, “Multilingual Speech and Text Recognitionmand Translation using Image”, April-2020