

MULTILINGUAL REAL TIME SPEECH TO TEXT CONVERTER

# A COURSE LEVEL PROJECT REPORT

***Submitted by***

**III-year students of Bachelor of Technology in Computer Science and Engineering**

|  |  |
| --- | --- |
| **AMBATI POLI REDDY** | **- 99220040238** |
| **CHINEPALLI NITHISH KUMAR REDDY** | **- 99220040254** |
| **VALLALE SRI SAI CHANDH VIGNESH** | **- 99220040218** |
| **A ROKESH** | **-9921004619** |
|  |  |
|  |  |

***in partial fulfillment of the course***

**213CSE3301-DEEP LEARNING**

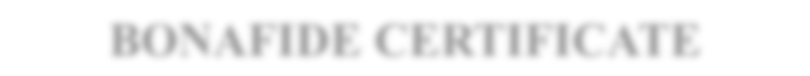
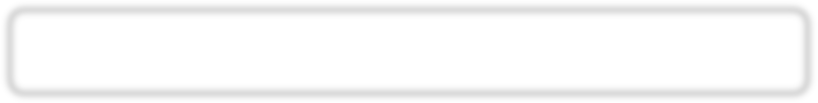
**in the**

**School of Computing**

**Department of Computer Science and Engineering**

**Academic Year 2023 – 2024 (Even Semester)**





**BONAFIDE CERTIFICATE**

# SCHOOL OF COMPUTING

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

Certified that this project report **“MULTILINGUAL REAL TIME SPEECH TO TEXT CONVERTER”** is the Bonafide work of **“AMBATI POLI REDDY,CHINEPALLI NITHISH KUMAR REDDY,VALLALA SRI SAI CHANDH VIGNESH, A ROKESH** who carried out the project work under my supervision.

**Faculty In-charge Head of theDepartment**

Submitted for the Project Viva-voce / Review held at Kalasalingam Academy of Research & Education, Krishnankoil on ………………………………

**Internal Examiner External Examiner**

# TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **CHAPTER No.** | **TITLE** | **PAGE No.** |
| 1. | **ABSTRACT** |  |
| 2. | **INTRODUCTION** |  |
| 3. | **LITERATURE SURVEY** |  |
| 4. | **HARDWARE AND**  **SOFTWARE COMPONENTS** |  |
| 5. | **METHODOLOGY OF WORKING** |  |
| 6. | **PRO’S AND CON’S** |  |
| 7. | **CONCLUSION** |  |
| 8. | **REFERENCES** |  |

## ABSTRACT:

The “Multilingual Real-Time Speech-to-Text Converter” is an innovative tool designed to facilitate communication across different languages in real time. In our increasingly globalized world, language barriers can hinder effective interaction in various

settings, such as business meetings, educational environments, and everyday conversations. Th is project addresses this challenge by enabling users to speak in one language and instantly rec translation in another language, displayed as text and spoken aloud.

## The system works by allowing users to select both their input language and the desired

## output language before recording their speech. Once the user speaks, the tool captures the audio, processes it using advanced speech recognition technology, and translates it into text

## in the chosen target language. This functionality is crucial for creating an inclusive environment where individuals from diverse linguistic backgrounds can engage and communicate freely.

## In addition to translating speech into text, the converter offers text-to-speech

## capabilities, giving users the option to listen to the translated output. This

## feature is powered by popular text-to-speech engines like `pyttsx3`, which works

## offline, and `gTTS`, which provides high-quality, natural-sounding voices online.

## Users can personalize their experience by adjusting the volume and the speed at which

## the translated text is read aloud.

## The applications for this technology are vast. It can significantly enhance collaboration in international business settings by enabling clear communication among team members from different countries. In education, it provides support for students learning in a language that is not their native tongue, making it easier for them to understand lectures and participate in discussions. Furthermore, in public spaces, such as airports or train stations, it can assist travelers by providing vital information in their preferred language.

## Overall, the “Multilingual Real-Time Speech-to-Text Converter” represents a step forward in overcoming language barriers, making communication more accessible and effective

## for everyone, regardless of their linguistic background. This project has the potential to improve understanding and connection among people worldwide, fostering greater collaboration and inclusivity.

## INTRODUCTION:

## In today’s world, people from different countries and cultures interact frequently,

## whether for work, learning, travel, or social events. However, language differences

## still create barriers that can make communication challenging. Our project,

## the “Multilingual Real-Time Speech-to-Text Converter”, is designed to help

## people bridge this gap by offering a way to understand each other in real time, even

## if they don’t speak the same language.

## The goal of this project is to take a live speech input from a user in one language and instantly convert it into text in another language, chosen by the user. For example,

## if a person speaks in English, the tool can translate it to Spanish, French, or any other supported language. This allows people to have a natural flow of conversation

## without needing to understand each other’s language.

## To make this even more useful, the tool also offers the option to hear the translated

## text in audio form. Users can choose from different text-to-speech (TTS) engines,

## such as `pyttsx3` for offline speech or gTTS for more natural, online voices. Users

## can adjust settings to customize how they hear the translation, including the volume

## and how fast the words are spoken.

## This project is designed to help in many situations, from business meetings

## and educational settings to travel and public spaces where people may need

## to understand each other. For instance, in an international meeting, participants

## can understand each other’s ideas immediately. In education, it can support students

## who learn in a different language than their native one. And in public spaces, it can

## make announcements and important information accessible to people who

## speak different languages.

## In short, the “Multilingual Real-Time Speech-to-Text Converter” is a tool to break

## down language barriers and help people communicate better. With this project, we

## aim to make understanding and connecting across languages easier and more

## accessible for everyone.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **LITERATURE SURVEY:** | | | | | |
| **SNO** | **TITLE** | **AUTHORS** | **JOURNAL** | **VOLUME** | **ACCEPT ED**  **DATE** |
| 1. | Speech-to-Text and Text-to-Speech Recognition Using Deep Learning | 1.V. Madhusudhana  2.T. Vaishnavi | International Conference on Edge Computing and Applications | Volume:9 | 16 August 2023 |
| 2. | Speech to text conversion for multilingual languages | 1.Yogita H 2.Sushama D. Shelke | International Conference on Communications and Signal Processing | Volume:7 | 24 November 2022 |
| 3. | Automatic Speech  Recognition : Speech to Text  Conversion | 1.Mr. Riyazahmed Jamadar  2.Mehul Pawar | International Research Journal of Modernization in Engineering Technology and Science | Volume 1: Long Papers, pages 4918 -  4935 | May-2023 |
|  | | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 4. | Multilingual Speech to text Converstion | 1.Saloni  2.W.Singh | Advances in Mathematics: Scientific Journal | - | May  2020 |

## HARDWARE AND SOFTWARE COMPONENTS:

## Hardware Components:

## 1.Computational Infrastructure:

## Servers/Workstations: Powerful CPUs or GPU are necessary to handle the processing demands of real-time speech recognition and translation.

## Memory Resources: Adequate RAM is needed to ensure smooth operation,especially when handling multiple simultaneous requests.

## 2.Microphones:

## - High-quality microphones are essential for capturing clear audio input from users, which is critical for accurate speech recognition.

## 3. Audio Input Devices:

## In addition to microphones, any other necessary audio interfaces or sound cards that enhance audio capture quality.

## 4. Storage Systems:

## Local storage or cloud-based solutions to save audio recordings, user settings.

## 

## Software Components:

## 1.SpeechRecognition:

## A Python library for performing speech recognition, allowing application to

## convert spoken language into text efficiently.

## 2.gTTS (Google Text-to-Speech):

## A Python library for converting text to speech using Google’s TTS API, enabling

## userto hear the translated text in different languages.

## 3.Pyttsx3:

## An offline text-to-speech library that allows the application to generate audio output without needing an internet connection, providing users with flexibility.

## 4.Python Standard Library Modules:

## Modules such as os for file and system management, which support various

## operations needed by the application.

## 5.User Interface Frameworks:

## Any additional libraries (like Flask or Django) if a more complex web application is

## Required beyond.

## METHODOLOGY OF WORKING:

## METHODOLOGY:

## 1. User Interface Design:

## The UI should allow users to select their input language and the desired output language for translation before recording their speech. It should also include controls for volume adjustment and speech rate.

## 2. Speech Input Capture:

## Implement functionality to capture real-time audio input from users

## 3. Language Selection:

## Allow users to choose the input and output languages from a predefine list.

## 4. Speech Recognition:

## Processing the audio input to extract speech features.

## 5.Translation:

## Sending the recognized text to the translation API. Receiving the translated text

## back from the API for further processing.

## 6. Text-to-Speech Conversion:

## Implement the text-to-speech functionality using gTTS or Pyttsx3. This step

## will convert the translated text back into spoken language.

## 7. Output Display:

## Present the translated text to the user on the interface, along with an option to play

## the audio of the translated speech.

## 8. Data Storage:

## Implement a system for storing user interactions, recorded audio, and translations.

## This can be done using local storage or cloud solutions, allowing users to access

## their previous sessions or feedback data.

## Working:

## 1. User Interaction:

## The user opens the application and is greeted with the main interface, where they

## select their input language and the language they wish to translate to.

## 2. Recording Speech:

## The user clicks the "Record" button, and the application captures their speech

## through the microphone.

## 3. Speech Recognition:

## Once the user stops recording, the application processes the audio using

## the SpeechRecognition library, converting the speech into text format.

## 4. Translation:

## The recognized text is then sent to the Google Generative AI translation model,

## where it is translated into the user’s selected output language.

## 5. Text-to-Speech Output:

## After receiving the translated text, the application utilizes either gTTS or Pyttsx3

## to convert this text into audio. The audio is played back to the user, and they can

## adjust the volume and speech rate as needed.

## 6. Results Display:

## The translated text is displayed on the user interface, along with playback controls

## for the audio output. Users can review the text and listen to the audio translation.

## 7. Data Handling:

## Application stores the interaction data for future reference, which can help improve

## model over time and provide users with access to their previous translations needed.

## 8. Continuous Improvement:

## Based on user feedback and testing, the application undergoes updates to enhance

## accuracy, user experience, and the range of supported languages.

## PRO’S AND CON’S:

## PRO’S:

## 1. Enhanced Communication:

## The system facilitates seamless communication between people who speak different languages, making it easier to understand and collaborate in diverse settings.

## 2.Real-Time Processing:

## By converting speech to text and translating it in real time, the application allows for

## dynamic conversations without delays, which is crucial in business meetings and live

## interactions

## 3. User-Friendly Interface:

## With an intuitive interface built using Streamlit, users can easily navigate the application, select languages, enhancing the overall user experience.

## 4. Versatile Applications:

## Project can be utilized in various fields, such as education, customer support, travel,

## and international business, broadening its impact and usefulness.

## 5. Accessibility Features:

## By providing text-to-speech functionality, the application is accessible to individuals with hearing impairments or those who prefer audio outputs, promoting inclusivity.

## 6. Customizable Settings:

## Users have the option to adjust the volume and speech rate, allowing for a personalized experience that caters to individual preferences.

## 7. Data Collection and Improvement:

## The ability to store user interactions can help in refining and improving the accuracy of

## speech recognition and translation models over time through machine learning.

## CON’S:

## 1. Accuracy Limitations:

## The effectiveness of the speech recognition and translation components can vary based on the quality of the input audio, accents, dialects, and background noise, potentially leading to misinterpretations.

## 2. Dependence on Internet Connectivity:

## If the application relies on cloud-based services for translation or TTS, users may experience delays or unavailability during poor internet conditions.

## 3. Language Support Constraints:

## While the system aims to support multiple languages, not all languages may have equal

## in terms of speech recognition and translation accuracy, limiting its usability for some

## 4. Processing Delay:

## Although designed for real-time interaction, there may still be minimal delays during processing of audio input, translation, and audio playback, which could disrupt

## flow of conversation.

## 5. Privacy Concerns:

## Capturing and storing user audio and translation data may raise privacy concerns.

## Users must trust that their data is handled securely and responsibly.

## 6. Complexity in Integration:

## Integrating different libraries and services (like SpeechRecognition, gTTS, and translation APIs) may introduce technical challenges, requiring careful management to ensure compatibility and performance.

## 7. Resource Intensive:

## Running real-time speech recognition and translation models can be resource-intensive, requiring significant computational power, which may not be accessible for all users or in all environment

## 

## CONCLUSION:

The “Multilingual Real-Time Speech-to-Text Converter” is a powerful tool designed to facilitate seamless communication across language barriers. By integrating advanced speech recognition, translation, and text-to-speech technologies, it enables users to engage in real-time conversations effortlessly. The user-friendly interface and customizable settings enhance the overall experience, promoting accessibility and inclusivity.

While challenges such as accuracy, internet dependency, and privacy must be addressed, the potential impact of this project is significant. As globalization continues to grow, this converter will play an essential role in fostering understanding and collaboration among diverse cultures, making communication more accessible and effective for everyone.

## REFERENCES:

1. V. Madhusudhana Reddy, T. Vaishnavi (16 August 2023). Speech-to-Text and Text-to-Speech Recognition Using Deep Learning. International Conference on Edge Computing and Applications.
2. Yogita H. Ghadage; Sushama D. Shelke (24 November 2016). Speech to text conversion for multilingual languages. International Conference on Communications and Signal Processing.
3. Shahana Bano, Pavaluri Jithendra (November 2020). Speech to Text Translation enabling Multilingualism. IEEE International Conference for Innovation in Technology.
4. Mr. Riyazahmed Jamadar, Mehul Pawar (May-2023). AUTOMATIC SPEECH RECOGNITION: SPEECH TO TEXT CONVERTER. International Research Journal of Modernization in Engineering Technology  and  Science.
5. SALONI AND W. SINGH (2020). MULTILINGUAL SPEECH TO TEXT CONVERSION. Advances in Mathematics: Scientific Journal.
6. Vinnarasu A, Deepa V Jose (October 2019). Speech to text conversion and summarization for effective understanding and documentation. International Journal of Electrical and Computer Engineering.
7. Pranab Das, Vijay Prasad (November 2015). VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT. Journal of Applied and Fundamental Science.
8. Ayushi Trivedi,Navya Pant, Pinal Shah,Simran Sonik and Supriya Agrawal (17-03-2018). Speech to text and text to speech recognition systems-A review. IOSR Journal of Computer Engineering.
9. Vijay Kumar, Hemant Singh, Animesh Mohanty (April 3, 2020). Real-Time Speech-To-Text / Text-To-Speech Converter with Automatic Text Summarizer Using Natural Language Generation and Abstract Meaning Representation. International Journal of Engineering and Advanced Technology.
10. Rustam Shadiev, Wu-Yuin Hwang, Nian-Shing Chen, Yueh-Min Huang (November 2014). Review of Speech-to-Text Recognition Technology for Enhancing Learning. This article is from Journal of Educational Technology & Society.