



Voice Translator Using Speech Recognition

Jami venkata krishna sai, Department of CSE, GIET University

Baswa sai kiran, Department of CSE, GIET University

Revalla Rudrateja, Department of CSE, GIET University

Bidush Kumar Sahoo, Department of CSE, GIET University bidush.sahoo@gmail.com

Abstract: A voice translator can refer to either a technology or a software that translates spoken language in English to a completely different language in real time. Speech recognition software can record spoken words, which are then processed through advanced language databases and machine learning, all of which are combined into what's known as a voice translator. It promotes comprehension among language shooters during its application at various activities, such as in travel, business meetings, and at international conferences. In order to ensure an adequate interpretation and also avoid mistranslations, voice translators most of the time resort to other techniques such as natural language processing. As an added advantage, a significant number of these devices are capable of improving their performance over time as they are able to learn according to user comments and interactions. Voice translators allow for easy use on the go as they usually come complete with microphones and speakers for hands-free use. Many of the modern systems, however, are not voice translators' which means they are able to "learn "and enhance themselves in the process of improving translation quality and adapting to the language preferences of the users. Portable devices usually have built- in high-quality microphones, sound-canceling devices and loudspeakers, so they can be used both in quiet and loud environments efficiently.

1. Introduction:

Voice translators are advanced technological devices or software applications designed to convert spoken language into another language. It helps communication between people who speak different languages. By capturing spoken words, processing them through speech recognition and natural language processing (NLP) algorithms [1,2], translating them via sophisticated language databases or machine learning models, and outputting the translated text audibly [3,4], voice translators enable seamless multilingual conversations.

2. Methodology

Step1: START

Step2: if user choose any source language

Step 2.1: if user choose any target language

Step 2.1.1: Click on start button

Step 2.2: you speak what you selected

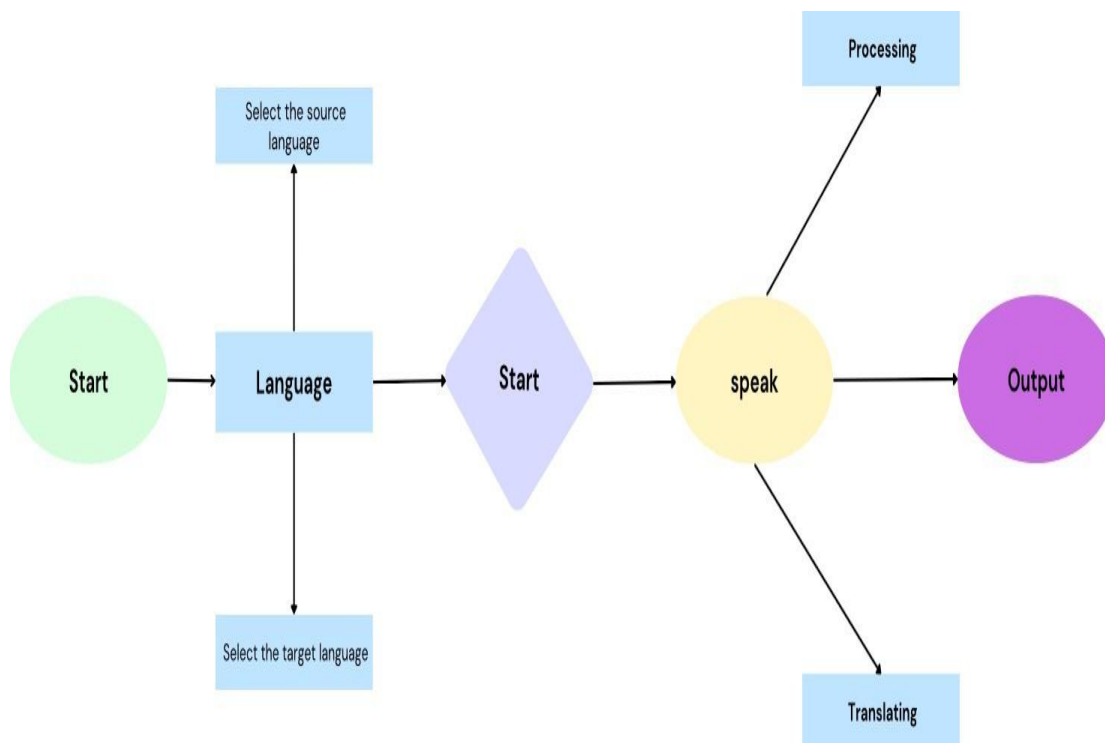
Step 2.2.1: it will take processing

Step 3: After processing it will take translating

Step 3.1: it will come output

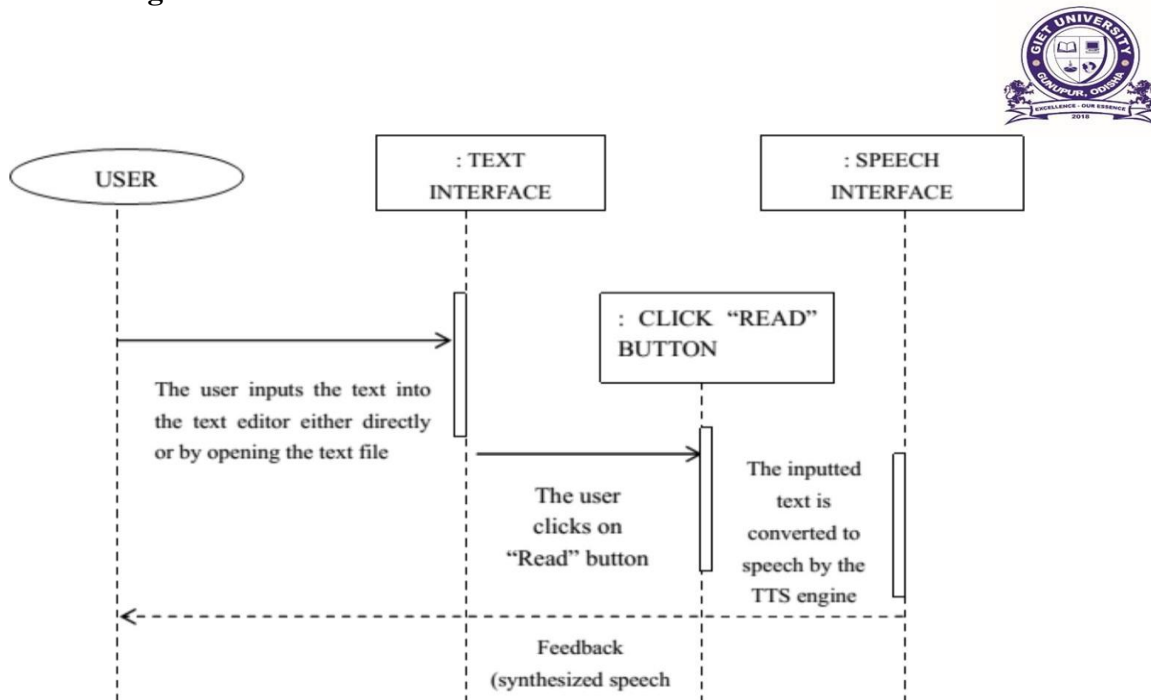
Step 4: Stop

• Structure Chart

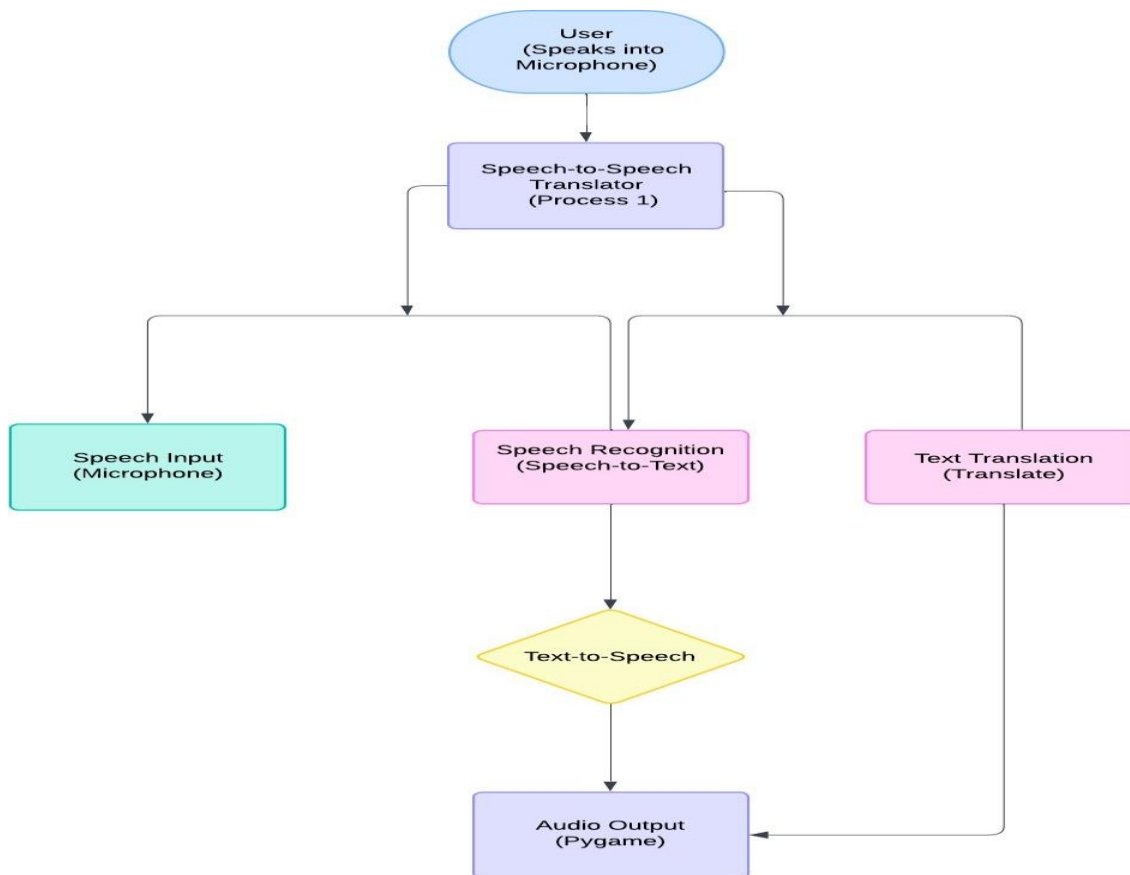




• **Interaction Diagram**

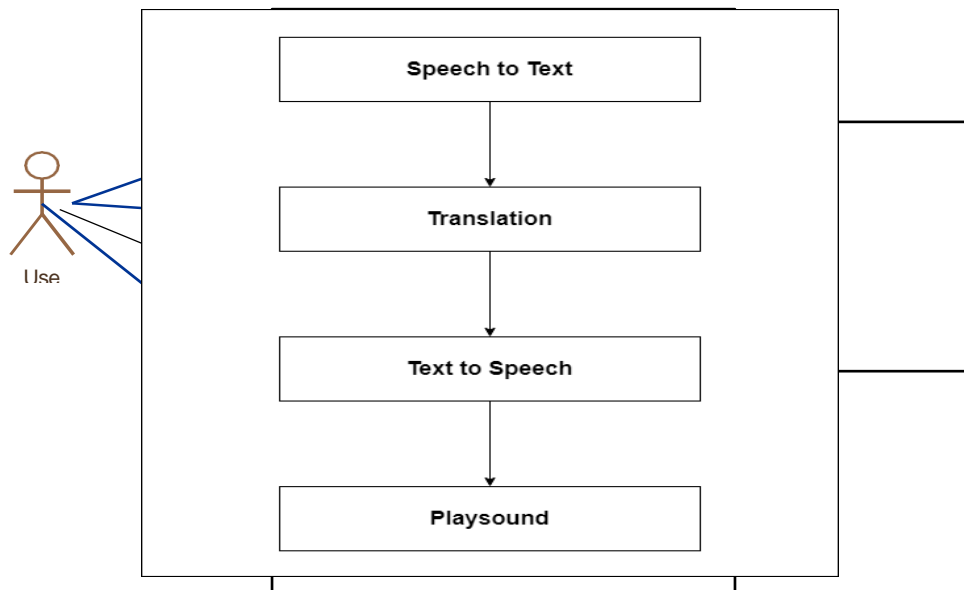


• **Data Flow Diagram**

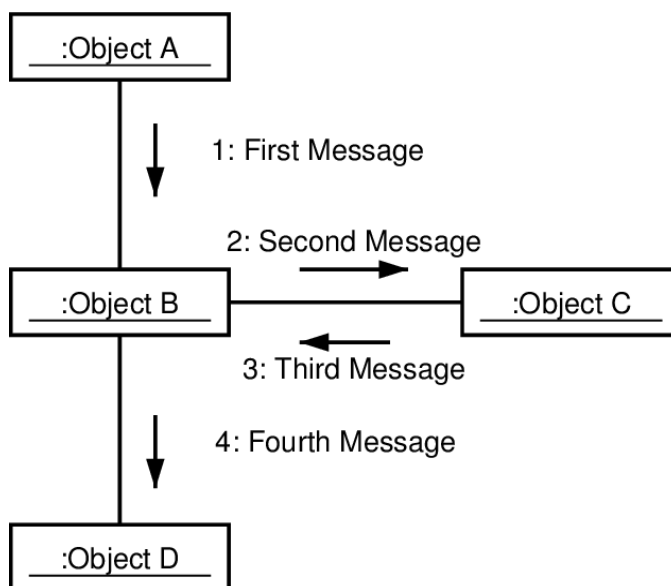


• **UML (Use Case Diagram)**

Text To speech Use Case Diagram



• **Collaboration Diagram**





3. Discussion

Some Screenshots of the work are provided below:

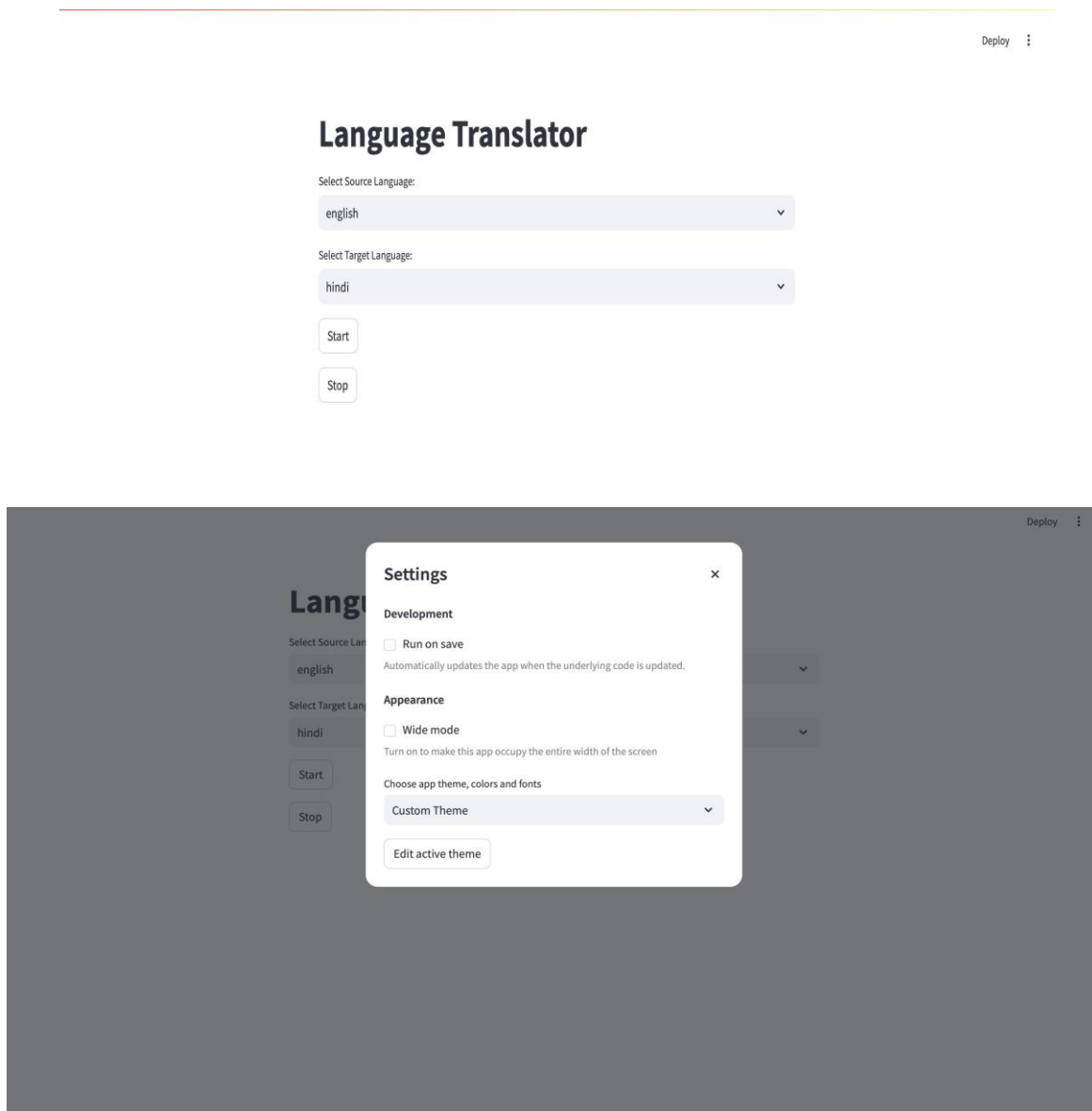


Fig.1. Screenshot of Language translator running

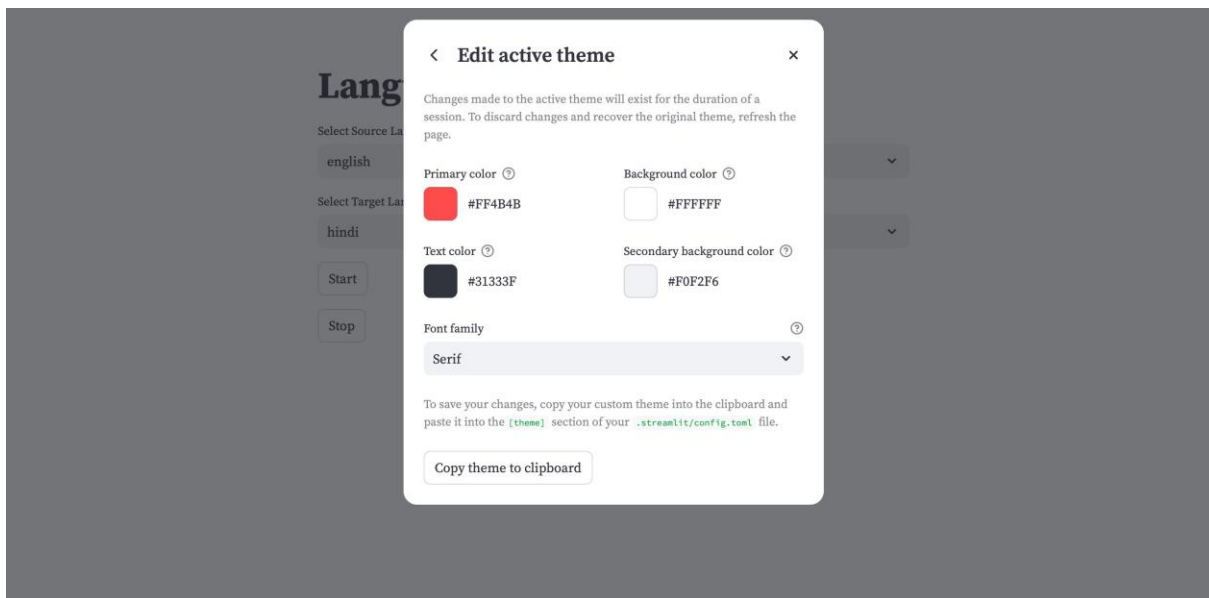


Fig.2. Screenshot of theme modification in language translator



4. TESTING

- **Unit Testing**

Each test case should follow this structure:

- **Test Case ID:** This is a unique number that identifies the test case.
- **Description:** This is a brief explanation of the test.
- **Input:** This is the data that is provided to the function being tested.
- **Expected Output:** This is the result that you expect the function to return.
- **Actual Output:** This is the result that the function actually returns when the test is run.
- **Test Result:** This is whether the test passed or failed, based on whether the actual output matches the expected output.

A **Voice Translator** is a system that allows spoken language to be converted into another language, typically involving three main processes:

1. **Speech-to-Text (STT):** It takes spoken audio and turns it into written text.
2. **Text Translation:** takes the written text and translates it into a different language.
3. **Text-to-Speech (TTS):** It takes the translated text and converts it back into spoken audio.

Testing and **unit testing** are essential to ensure each of these components operates accurately and efficiently. This includes validating the system's ability to recognize diverse speech patterns, translate accurately, and produce understandable audio output.

Overview of Testing and Unit Testing

- **Testing:** Ensuring each part of your software works as intended. This can include different types of testing, such as integration testing (multiple components), system testing (full application), and unit testing (individual functions).



- **Unit Testing:** Testing individual functions or components in isolation to verify they work correctly. Unit tests are often automated to ensure that any code changes do not break existing functionality.

Unit Test Structure for Voice Translator Project

A good **unit test case** structure includes:

- **Test Case Description:** Describe what the test is checking.
- **Preconditions:** List any requirements that need to be in place before the test can be performed.
- **Test Input:** Specific inputs or conditions for the test.
- **Expected Output:** What the function should return or produce.
- **Actual Output:** The output generated when running the test.

Test Case ID	Function	Description	Input	Expected Output	Actual Output	Result
TC-001	Speech-to-Text	Convert speech to text	test_audio.wav	"Hello, how are you?"	(Result from function)	Pass/Fail
TC-002	Text Translation	Translate text from English to Spanish	"Hello, how are you?", "es"	"Hola, ¿cómo estás?"	(Result from function)	Pass/Fail
TC-003	Text-to-Speech	Convert text to audio and create an output file	"Hola, ¿cómo estás?"	File output_audio.mp3 exists	(File existence check)	Pass/Fail



5. Conclusion

The Speech-to-Speech Translator developed in this project successfully bridges language barriers by enabling real-time audio translation. Using Python as the core language, along with libraries such as Speech Recognition, Streamlit, and Pygame, this system allows users to speak in one language and receive an audio translation in another. This project demonstrates the potential of combining Natural Language Processing (NLP) and Text-to-Speech technologies to facilitate communication, offering a practical solution for use cases like travel, international business, and social interactions between speakers of different languages.

6. Reference

1. S. K. Mukundan, “‘ Shreshta Bhasha ’ Malayalam Speech Recognition using HTK,” vol. 1, no. 1, pp. 1–5, 2014.
2. J. S. Pokhariya and S. Mathur, “Sanskrit Speech Recognition using Hidden Markov Model Toolkit,” vol. 3, no. 10, pp. 93–98, 2014.
3. L. Rabiner and B.-H. Juang, Fundamentals of Speech Recognition. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
4. X. Yao, P. Bhutada, K. Georgila, K. Sagae, R. Artstein, and D. Traum, “Practical Evaluation of Speech Recognizers for Virtual Human Dialogue Systems,” 2008

GIET UNIVERSITY JOURNAL, 2025, VOL.4, NO.1, PP. – 95-105

Manuscript received: 13th November 2024, revised: 18th December 2024

Accepted: 7th January 2025

ISSN: 2583-5955

