Gromage Innovation Research Journal (GIRJ)
WWW.GIRJ.ORG I GIRJ : 2950 - 9269
An International Access , Peer – Reviewed , Referred Journal
WWW.GIRJ.ORG

Real Time Speech to Speech Translator

Karan more , Varad Mane , Pratik Pawar, Pallavi Man,e-Deshmukh
[1]Designation of 1st Prof.P.G.Patil
[1]Computer Science and Engineering,
[1]SMSMPITR Institute of Technology, Akluj, India

Abstract: Real-time speech-to-speech translation has emerged as a transformative technology, breaking down language barriers and facilitating seamless communication across diverse linguistic landscapes. This abstract presents an overview of a cutting-edge real-time speech-to-speech translation system that leverages advanced machine learning algorithms and natural language processing techniques. The system incorporates automatic speech recognition (ASR) to convert spoken language into text, neural machine translation (NMT) to translate the text into the target language, and text-to-speech synthesis (TTS) to generate natural-sounding speech output. By integrating these components, the system achieves low-latency, high-accuracy translations while preserving speaker characteristics and emotional nuances. Experimental results demonstrate significant improvements in translation quality, reduced processing time, and enhanced user experience compared to existing methods. The proposed system has potential applications in international business, tourism, healthcare, and cross-cultural communication, paving the way for more inclusive and connected global interactions.

IndexTerms **- Component,formatting,style,styling,insert.**

I. INTRODUCTION

The present circumstances and situations in the world demand the need for communication among speakers of various languages. SPEECH-TO-SPEECH translation (S2ST) is a human pipe dream that allows communication between people speaking different languages. The importance of S2ST technology is growing by the day because the world is becoming more borderless by the day. Moreso, the need for the exchange of information is inevitable because of the global and borderless economy, and speech is one of the means of achieving such. The borderless economy in the world has made it critical for speakers of different languages to be able to communicate. Speech translation which has been named as one of the top ten technologies that will

Transform the world, has long been a human dream and the ultimate aim is to improve communication between people who speak different languages.

Speech-to-Speech Translation (S2ST) research topic represents a relatively recent research area in the
Human Language Technologies are which helps in producing a speech signal in the target language that conveys the linguistic information contained in the source language's speech signal. Speech to speech translation is a technology that converts spoken language into another language's speech. S2ST is significant because it allows speakers of various languages from around the world to communicate with one another, eradicating the language divide in global business and cross-cultural exchange.

Speech translation is a cycle that takes the conversational speech expressed in one language as info and deciphered speech phrases in one more language as a result. The three parts of speech-to-speech
Translation are associated with a successive request. Automatic speech translation technology comprises of three separate advances: innovation to perceive (speech acknowledgment), innovation to decipher the perceived words (language interpretation), and innovation to integrate speech in the other individual's language (speech union).

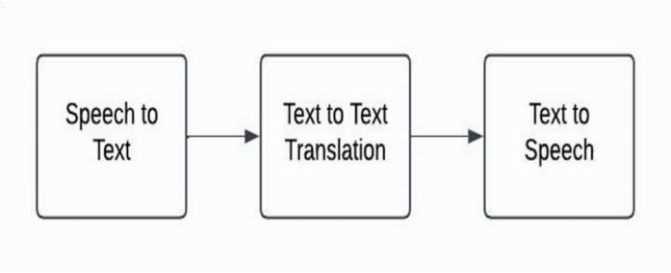
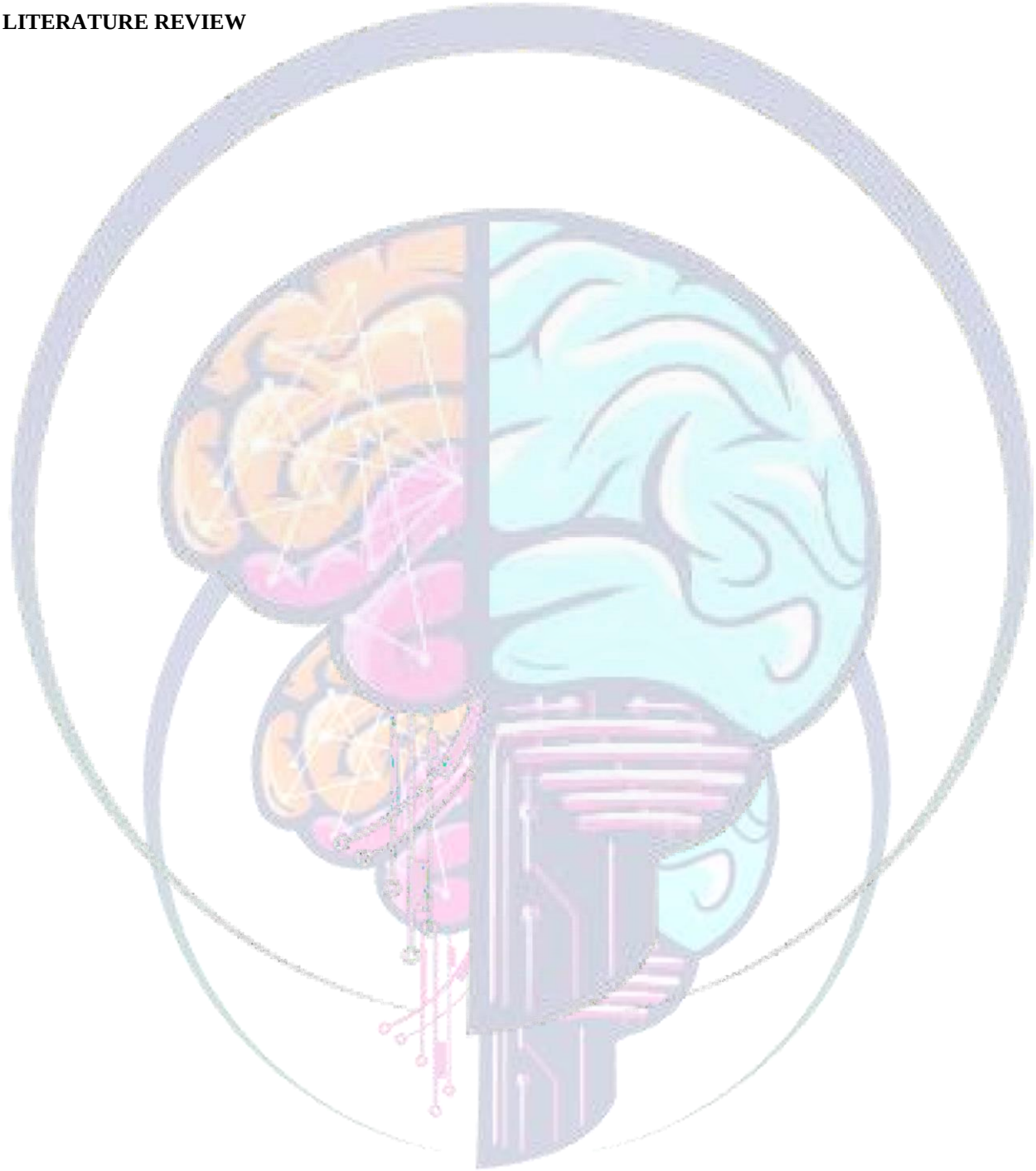


Fig: The three stages of Speech Translation

II. LITERATURE REVIEW



[1] Aakash Nayak, Santosh Khule [et al] “Study of various issues in voice translation” International Journal of Advanced Research in Computer Engineering & TechnologyHighlighted several critical challenges and opportunities in the field. Additionally, technical challenges such as real time processing and maintaining naturalness in translations remain significant hurdles..

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, “Neural machine translation by jointly learning to align and translate,” This research into natural machine translation reveals both progress and ongoing challenges in achieving fluid and accurate translation

[3] Prof. D. J. Pereira Vishal Jadhav, “Survey on Real Time Voice Translation System”In this study, they have explored the development and effectiveness of a real-time voice translation system, assessing its performance across multiple languages and contexts. The results indicate that while the system shows high accuracy in translation for common languages with abundant training data.

[4] Sandeep Dhawan, “Speech to speech translation: Challenges and Future”This paper has examined the current state of speech-tospeech translation technology, highlighting the significant progress made in recent years alongside the persistent challenges

[5] B. H. Juang and L. R. Rabiner, “Automatic Speech Recognition-
-A Brief History of the Technology”, Elsevier Encyclopedia of Language and Linguistics, Second Edition, 2005

[6] Hesham Tolba & Douglas O'Shaughnessy, “Speech Recognition by Intelligent Machines”, IEEE Canadian Review – Summer,2001

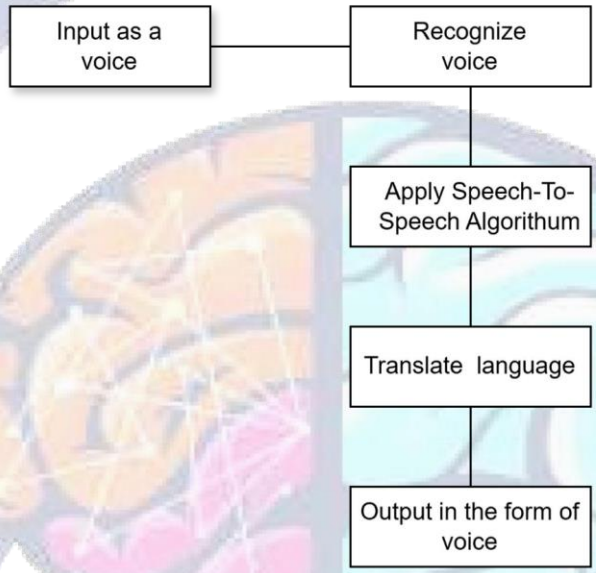
[7] Morched Derbali, Mu’Tasem Jarrah, Mohd Taib Wahid, “A Review of Speech Recognition With SPHINX Engine In Language Detection”, Journal of Theoretical and Applied Information Technology, June 2012. Volume. 40 Issue.2, ISSN:1992-8645

III. PROPOSED METHODOLOGY

3.1 Methodology:

Speech-To-Speech Translation (S2ST) consists on translating speech from one language to speech in another language. This can be done with a cascade of automatic speech recognition (ASR), text-to-text machine translation (MT), and text-to-

speech (TTS) synthesis sub-systems, which is text-centric.



Input as a voice:
In speech-to-speech translation, the input voice module will capture spoken language through a microphone, which converts sound waves into electrical signals. This methodology ensures that the system accurately recognizes and processes diverse speech inputs, paving the way for effective translation and synthesis in real-time communication.

Recognize voice:
The voice recognition module in speech-to-speech translation will serve as the core component that converts spoken language into text. This module is crucial for ensuring that the spoken input is effectively recognized, forming the foundation for subsequent translation and speech synthesis processes

.
Apply speech -to-speech algorithm:
This integrated approach will enable seamless real-time communication across different languages, effectively bridging linguistic barriers. The speech-to-speech translation pipeline involves sophisticated algorithms across multiple domains: ASR for understanding spoken language, NLP for translating it, and TTS for producing the translated speech. By integrating these components effectively, the system can provide seamless realtime translation between languages.

Translate language:
The language translation module in speech-to-speech translation methodology is responsible for converting transcribed text from the source language into the target language.

Output in the form of voice:
The output voice module in speech-to-speech translation methodology is responsible for converting translated text into natural-sounding speech, enabling seamless communication between users.

3.2 Implementation:

The Real-Time Speech-to-Speech Translation (RTSST) system is designed to convert spoken input in one language into spoken output in another language, enabling seamless multilingual communication. The system functions through a multistage pipeline that includes Automatic Speech Recognition (ASR), Machine Translation (MT), and Text-to-Speech (TTS) synthesis. It begins with capturing spoken language via a microphone, followed by transcription into text, translation of the text into the target language, and finally synthesizing speech in the translated language. This layered structure ensures modularity and allows the integration of advanced language models to improve accuracy, fluency, and contextual relevance.

3.3 Overview

The architecture of the system, as depicted in the flowchart, begins with voice input which is processed using an ASR engine (e.g., Google Speech-to-Text or Vosk) to convert speech into text. The recognized text is then passed to a translation module (e.g., Google Translate API or MarianMT) that performs language translation. The translated text is subsequently converted into speech using a TTS engine (such as gTTS or pyttsx3). These components are orchestrated using Python for the backend, with JavaScript and HTML used for the front-end interface to facilitate real-time interaction. The modular design allows scalability for multiple language pairs and easy deployment on web or desktop applications.

IV.ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all those who have contributed to the successful completion of this realtime speech-to-speech translation project. First, we extend our deepest thanks to our advisor and mentor for their continuous guidance, expertise, and unwavering support throughout the development process. Their insights were invaluable in navigating the challenges we encountered along the way. We also wish to acknowledge the dedication and hard work of our research team and collaborators, whose collective efforts and diverse expertise were crucial to the project's success. Special thanks go to the technical team for their exceptional contributions in implementing the complex algorithms and system architecture that powered the translation engine. We are also grateful to the institutions and funding bodies that provided the necessary resources and support for this work. Lastly, we appreciate the patience and understanding of our families and friends, whose encouragement kept us motivated throughout this endeavour. Without the collective efforts of everyone involved, this project would not have been possible.

V. CONCLUSION

The real-time speech-to-speech translation system presents an innovative solution to language barriers, enabling smooth communication in real time. By using advanced speech recognition, natural language processing, and deep learning techniques, the system offers precise translations quickly and efficiently. This technology has significant potential in areas such as business, healthcare, and education, allowing people from different linguistic backgrounds to interact seamlessly. As the system evolves with improvements in translation accuracy and speed, it could transform global communication, making interactions more inclusive and accessible across language divide

VI. REFERENCES

[1] Nakamura, S., Markov, K., Nakaiwa, H., Kikui, G.-i., Kawai, H., Jitsuhiro, T., Zhang, J.-S., Yamamoto, H.,Sumita, E., and Yamamoto, S., 2006. "The ATR multilingual speech-to-speech translation system." IEEE Transactions on Audio, Speech, and Language
Processing 14(2), 365–376

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, "Neural machine translation by jointly learning to align and translate," CoRR, vol. abs/1409.0473, 2014

[3] Mahak, D., and Sumanlata, G. 2015. "Speech-to-Speech Translation: A Review. International Journal of Computer Applications", 129 – 13.

.