

REAL-TIME SPEECH TRANSLATION APPLICATION USING GOOGLE APIs

¹R. KARTHICK, ²M.H. KSHEERABDHINATH, ³N.K. KUMARESH BABU, ⁴R.R. ROHITH, ⁵K.J. RISIKESAN

¹⁻⁵ Computer Science and Engineering, KLN College of Engineering, Madurai, India,

Article Information

Received : 05 Nov 2024
Revised : 10 Nov 2024
Accepted : 13 Nov 2024
Published : 15 Nov 2024

Corresponding Author:

M.H. Ksheerabdhinath
Email: ksheerab03@gmail.com

Abstract— The Real Time Speech Translation Application is designed to bridge language barriers through seamless, two-way communication. The app captures voice input, converts it into text via Google's STT API, and translates it using a custom model. The translated text is then synthesized back into audio through TTS, allowing users to communicate across languages effortlessly. Using Firebase for real-time data exchange, the application supports efficient communication between devices, accommodating diverse language preferences and providing a user-friendly interface. This solution demonstrates the potential of machine learning and cloud technologies to foster inclusive communication, addressing multilingual needs dynamically.

Keywords: *custom translation model, firebase, machine learning integration, multilingual communication, speech to text, text to speech*

Copyright © 2024: R. Karthick, M.H. Ksheerabdhinath, N.K. Kumares Babu, R.R. Rohith, K.J. Risikesan. This is an open access distribution, and reproduction in any medium, provided Access article distributed under the Creative Commons Attribution License the original work is properly cited License, which permits unrestricted use.

Citation: R. Karthick, M.H. Ksheerabdhinath, N.K. Kumares Babu, R.R. Rohith, K.J. Risikesan "Real Time Speech Translation Application using Google APIs ", Journal of Science, Computing and Engineering Research, 7(11), November 2024.

I INTRODUCTION

A translation application enables real-time communication by converting spoken language into text and translating it into a target language, providing an accessible solution for cross-linguistic interactions. The translation process involves capturing speech input, converting it into text using Speech-to-Text (STT) technology, and translating the text into the desired language through a custom translation model. This translated text is then converted back into speech using Text-to-Speech (TTS) technology, delivering an audio response to the recipient.

The application incorporates Firebase for real-time communication, allowing seamless message exchange between devices. It also integrates advanced machine learning techniques, where models such as recurrent neural networks (RNNs) are used to capture the nuances in speech patterns, and transformers help improve translation accuracy. Combining these methods enhances the system's adaptability to different languages and dialects, ensuring clear, context-aware translations.

II RELATED WORKS

Speech translation is the process of converting spoken language from one language to another, enabling real-time communication across language barriers. This involves multiple steps: first, capturing and converting the spoken input into text using Speech-to-Text (STT) technology. Next, the text is translated into the target language using machine translation, which can leverage neural networks and transformer models for

accuracy. Finally, Text-to-Speech (TTS) technology converts the translated text back into audio, delivering the response in the recipient's language. Speech translation systems are increasingly vital in global communication, supporting applications in travel, customer service, education, and international business.

III PROPOSED SYSTEM

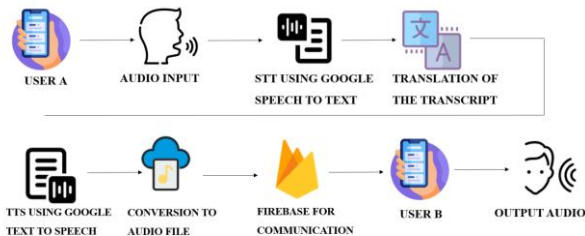
The proposed system is designed to function as a **multilingual chat application** that allows users to communicate seamlessly through text and audio messages, enabling efficient communication regardless of language barriers. This system is a significant advancement from traditional chat applications as it integrates advanced **language translation capabilities**, allowing users to send **audio messages** that can be translated into a language of the recipient's choice before delivery. The system uses Google's Speech-to-Text (STT) API to convert audio messages into text. When a user records an audio message, the STT service listens to the input, processes the speech, and transcribes it into text form. This transcription serves as the input for the next step in the process, which is language translation. After the audio message is transcribed into text, the system uses Recurrent Neural Networks (RNN) for the machine translation step.

Once the translation is complete, the system converts the translated text back into speech using Google's Text-to-Speech (TTS) API. The TTS system generates a natural-sounding

voice in the recipient's chosen language, allowing them to listen to the translated message as an audio file.

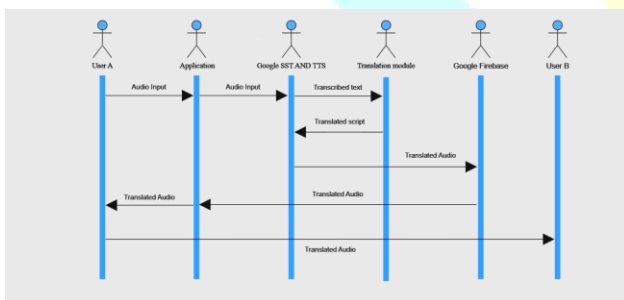
Firebase plays a crucial role in enabling real-time message delivery and synchronization between users. Whether users are sending text, audio, or translated messages, Firebase ensures that the messages are instantly updated across devices. The Firebase Cloud Firestore is used to manage the chat data, providing a seamless experience for both sender and receiver.

III ARCHITECTURE DIAGRAM



In the figure, the diagram illustrates the workflow of a speech-to-speech translation application designed to facilitate real-time, cross-lingual communication between two users, identified as User A and User B. This solution leverages a series of advanced cloud-based technologies, specifically Google's Speech-to-Text (STT), Text-to-Speech (TTS) services, in combination with Firebase for efficient data transfer. This integrated approach enables seamless audio-based conversations, allowing users to communicate naturally across different languages in near real-time.

IV SEQUENCE DIAGRAM



User A Initiates Audio Input: The process starts when User A provides an audio input to the Application. This input might be a spoken sentence or phrase in one language.

Application Processes Audio Input: The Application receives the audio input from User A. It then sends this audio data to Google Speech-to-Text (SST) for transcription, and Text-to-Speech (TTS) for voice processing.

Google SST Transcribes Text: Google SST converts the audio input into transcribed text. This transcription captures the spoken content in text form, enabling further processing in the next stage.

Text Sent to Translation Module: The transcribed text is then sent to the Translation Module, which translates the text from the original language into the target language. The translated script is created and sent back to Google SST and TTS for audio output in the target language.

Translation Module Creates Translated Audio: The Translation Module also generates translated audio based on the translated script, allowing for an audio version of the translated content.

Audio Stored and Delivered via Google Firebase: The translated audio is sent to Google Firebase for temporary storage or real-time streaming, enabling efficient distribution to the intended recipient. • **User B Receives Translated Audio:** Finally, User B receives the translated audio output, allowing them to listen to the translated message in their own language.

IV RESULTS

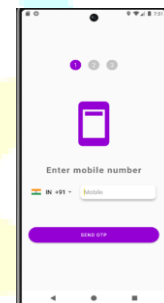
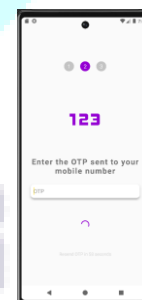


Figure 4.1

In the Figure 4.1, user registers his account in that application by choosing their country code followed by entering their mobile phone number. And then, the user has to click the button SEND OTP so that it will proceed to the second activity.



The User has to enter the OTP sent to their phone number in-order to verify the authentication of their account and to check whether they are not a robot, we have integrated the Firebase Authentication that prompts us to tick the boxes which has the images like stairs, cars, buses, traffic light, etc...

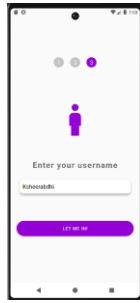


Figure 4.3

In Figure 4.3, user is asked to enter a username, using which they will be able to login each time they come into the application. To ease the hassle, once you register your account, your username will be automatically typed in the username space.



Figure 4.4

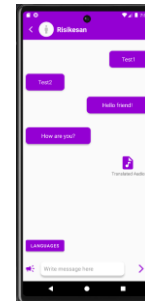
In the Figure 4.4, the chat user-interface is visible. This is the space where you can interact with each other. You can send text message as well as voice input through the speaker icon.



Figure 4.5

In the Figure 7.5, User can choose or select the language to which they desire to convert it into. There are 11 languages, namely English, Japanese, Tamil, Spanish, Tagalog, Arabic, Russian, German, Telugu, Urdu, French. The user has to choose a language and it will take a few seconds to get translated. Before doing that, the user has to record an audio input by pressing the speaker icon. And then, the user has to speak what they desire or want to.

Finally, after the translation process, the audio file is sent to the User B, which in this case is Risikesan. Both the Users, Ksheerabdh and Risikesan can access that audio file and listen to the translated audio by clicking the audio file icon, which plays the translated audio. This process can be done any number of times.



V CONCLUSION

The integration of speech recognition, translation, and text-to-speech technologies in this voice-based multilingual chat application facilitates seamless communication across language barriers. By leveraging a custom translation model and real-time processing, the application ensures accurate and timely translation of user messages, creating a fluid and interactive experience for all users. Through an intuitive interface, users can easily select their preferred language, send messages, and receive translations, allowing for natural and effective cross-linguistic conversations.

This application is designed to support a diverse user base, accommodating both casual and professional communication needs. With its user-centric design and robust functionality, the app provides a dependable tool for users to engage in meaningful conversations across languages, bridging cultural and linguistic divides. The technology is optimized to deliver translations quickly, maintaining the flow of conversation and minimizing delays, which is critical for user satisfaction in real-time communication scenarios.

VI FUTURE ENHANCEMENTS

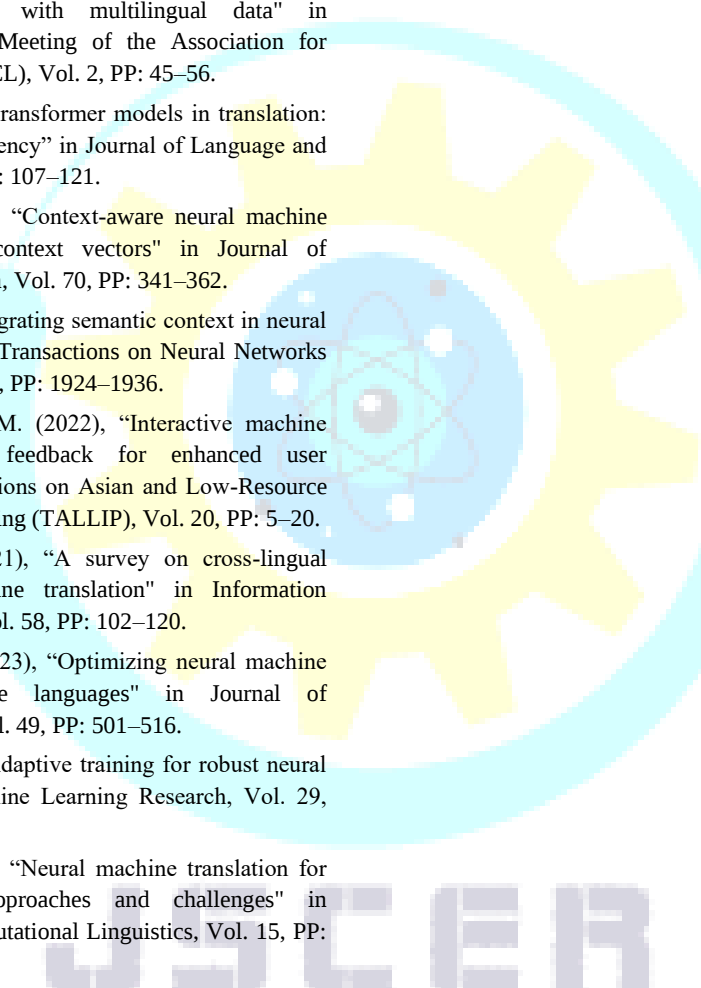
Future enhancements for this multilingual voice-based chat application could further elevate its functionality, accessibility, and user experience. One key area for improvement would be to implement adaptive translation models that fine-tune based on user feedback and evolving language patterns, ensuring continuously improved translation accuracy and relevance. Expanding the app to include support for more languages and dialects, including regional variations, would broaden its usability across diverse cultural backgrounds.

Introducing personalized user profiles could allow the app to track individual language preferences, commonly used phrases, and conversation history, creating a more tailored and seamless experience for frequent users. Additionally, incorporating offline functionality, where basic translation features can be accessed without internet connectivity, would make the app more versatile, especially in areas with limited connectivity.

Adding a gamified experience, such as progress badges for practicing new languages or communication milestones, could enhance user engagement. Features like real-time translation accuracy feedback and suggestions for language learners could

help users refine their communication skills while using the app.

REFERENCES

- 
- [1] Kumar, S., & Singh, A. (2021), "Advancements in neural machine translation: A comprehensive review" in *International Journal of Artificial Intelligence and Applications*, Vol. 12, PP: 57–74J
- [2] Chen, Y., Wang, Z., & Liu, X. (2020), "Towards robust neural machine translation for real-world text" in *Computational Linguistics*, Vol. 21, PP: 303–321.
- [3] Nguyen, T. H., & Chiang, D. (2022), "Improving low-Resource neural machine translation with multilingual data" in *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, Vol. 2, PP: 45–56.
- [4] Geng, Z., & Yu, J. (2023), "Transformer models in translation: Enhancing accuracy and efficiency" in *Journal of Language and Linguistic Studies*, Vol. 14, PP: 107–121.
- [5] Park, H., & Kim, M. (2021), "Context-aware neural machine translation using dynamic context vectors" in *Journal of Artificial Intelligence Research*, Vol. 70, PP: 341–362.
- [6] Li, R., & Liu, Y. (2020), "Integrating semantic context in neural machine translation" in *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32, PP: 1924–1936.
- [7] Xu, J., Zhang, Y., & Sun, M. (2022), "Interactive machine translation with real-time feedback for enhanced user experience" in *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, Vol. 20, PP: 5–20.
- [8] Huang, J., & Zhao, L. (2021), "A survey on cross-lingual language models for machine translation" in *Information Processing & Management*, Vol. 58, PP: 102–120.
- [9] Ramesh, K., & Gupta, V. (2023), "Optimizing neural machine translation for low-resource languages" in *Journal of Computational Linguistics*, Vol. 49, PP: 501–516.
- [10] Zhou, Q., & Li, B. (2020), "Adaptive training for robust neural machine translation" in *Machine Learning Research*, Vol. 29, PP: 190–205.
- [11] Patel, R., & Shah, D. (2022), "Neural machine translation for low-resource languages: Approaches and challenges" in *International Journal of Computational Linguistics*, Vol. 15, PP: 145–161.
- [12] Singh, A., & Mehta, P. (2021), "Enhancing translation accuracy with contextual embeddings in neural networks" in *Journal of Artificial Intelligence and Language Processing*, Vol. 28, PP: 230–245.
- [13] Zhang, L., & Chen, S. (2023), "Cross-lingual transfer learning in machine translation for under-resourced languages" in *Journal of Applied Machine Learning*, Vol. 18, PP: 90–110.
- [14] Garcia, M., & Lopez, R. (2020), "Hybrid approaches in neural machine translation: Combining rule-based and statistical models" in *Language Technology Research*, Vol. 7, PP: 77–93.
- [15] Banerjee, S., & Roy, A. (2021), "Evaluating robustness in neural machine translation systems for domain-specific applications" in *Computational Intelligence Review*, Vol. 12, PP: 42–59.
- [16] Kim, J., & Park, S. (2022), "Exploring multilingual transformers for enhanced translation quality" in *Transactions on Language and Linguistics*, Vol. 9, PP: 183–201.
- [17] Wang, X., & Zhou, T. (2023), "Domain adaptation techniques in neural machine translation for technical texts" in *Journal of Language Processing*, Vol. 21, PP: 299–314.