e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

TRANSLATE BUDDY

Sudha V Pareddy*1, Hanagal Atharva Anil*2, Kundile Prasad Nivarthi*3, Manoj Kumar*4

*1Professor, Department of Computer Science & Engineering, PDA College of Engineering, Kalaburagi, Karnataka, India.

*2,3,4Students, Department of Computer Science & Engineering, PDA College of Engineering, Kalaburagi, Karnataka, India.

DOI: https://www.doi.org/10.56726/IRJMETS63480

ABSTRACT

Speech Processing is the study and manipulation of Speech Transmitted signal as a form of Communication to extract meaningful information from it. This field applied dedicated algorithms, particularly neural networks for various tasks like translating one language to another while keeping in mind of the local context that makes it meaningful. Text-to speech (abbreviated TTS) is a technology that converts written text into spoken words. Text to speech works with machine generated voices, imitating human sound in high-fidelity. This includes our complex linguistic models for pronouncing words, intonation and rhythm that makes the output sound like a human voice. In contrast, speech-to-text (STT) models convert spoken language to written text, enabling the accurate representation of a range of accents and dialects. This entails intricate acoustic modeling and natural language processing efforts to ensure accurate transcription. The system of the present invention is to provide a multilingual voice translation and synthesis system using advanced speech recognition technologies such as Hidden Markov Models (HMM), Recurrent Neural Networks (RNN),Deep Neural Networks (DNN) which allows conversation in multiple languages to be translated into another simultaneously. While HMMs are used to model sequences of sounds in speech in a probabilistic way, RNNs (and similarly DNNs) can capture long-term dependencies and learn from large datasets — useful for the parsing structure and complexities of human language. This integration provides accurate and flexible speech recognition/synthesis for speaking many languages/dialects, dealing with pronunciation, grammar and context subtleties.

Keywords: Speech Processing, Text-To-Speech, Speech-To-Text, Text Translation.

I. INTRODUCTION

The growing trends as well as research and development of text-to-speech system can be attributed to an importance advance in technology and especially in invention of the use of machine learning and natural language processing. These systems are important in apaned and voice for the blind and interactive voice response systems respectively. This means that the more recent advancements in synthesizing speech have employed sophisticated neural structures like the transformers to improve on the produced audio quality and its naturalness [4]. Thus, reviews focus on the importance of the TTS systems, discuss the operation principles and possibilities of enhancing the systems [2], [3]. The integration of automatic language translation interface has enabled technology be nearer [10] to the people in compliance with the general trends, automatic speech recognition where interlanguage models enhance the accuracy of speech to text transcriptions or recognitions [5]. Systematic reviews reveal future developments and advancements of TTS technologies, the approaches and uses of which are numerous and versatile [1].

More advanced comparison and voice recognition abilities have emerged because of advancements in machine learning [7]. A kind of integration that has been evidenced by TTS with voice-controlled applications is a way in which voice recognition can be applied in web applications to create better and more intuitive interfaces [6]. These also have real utility, for example, identifying objects using a microphone, and turning speech to text [8]. TTS research and development is being accentuated as a dynamic process that involves constant enhancement and effective merging with other systems of connected kinds which are expected to fundamentally shift the way people interact with technological solutions [9]. Live text recognition and translation in different mobile applications have made the access to information more open, meaning that these technologies have progressed further than theory. Current state of art in speech translation includes systems that allow for converting speech from one language to text in another language in real time using complex two pass decoding [12].

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

With the current advancements in the approach towards globalization, adequate translation is critical in using different languages. As people are Socializing more in diverse cultures and different languages, a competent multilingual voice transcribe and translate system would help in breaking language barriers and increase goodwill and understanding. As was mentioned before the goal is to create a system that covers this need based on the modern systems of speech recognition, machine translator, and TTS.

II. METHODOLOGY

The idea behind the Integrated Speech Processing & Language Interface App is to develop a system that has many layers and is capable of implementing natural language in order to eliminate language barriers. Pivoted in this scheme is a strong block diagram made of three elements which when integrated, forms a powerful structure each developed carefully to serve specific roles acceptable for multi-lingual communications. For instance, in the first module targeting speech recognition, some of the features, including Hidden Markov Models (HMM), Deep Neural Networks (DNN), and Recurrent Neural Networks (RNN), are applied. In this way, these algorithms help to convert voice to text and function properly depending on the accents and environment noise. Further, the parametric synthesis of this module also involve identifying the SourceLanguageCode according to the input received from the user's end thus making the translation of speeches into text convenient.

The second module deals with the translation of the given language; the Google Translation Model that incorporates NMT algorithms is used here. This module makes it possible to translate between languages in real-time; it means that whenever a user writes something in language X, the system will instantly translate that into language Y, with close attention to grammar and vocabulary. It is capable to find out the TargetLanguageCode automatically as per the input given by the user, making the change as per requirement moreover, it gives more focused translation result according to the need of the user. Furthermore, this module synchronizes with the application's speech recognition module, which helps speak the translated text by espousing parametric synthesis techniques, all the time factoring in the tone and speed preferred by the end-user, both of which are categorized under Speech Tone Preference and Pace.

The third module of the app is dedicated to the Text to Speech conversion where the concept of Android Text to Speech Engine is applied to text to speech conversion. It also supports a wide range of high-fault-tolerant and parametric synthesis capabilities that make it furthermore possible to synthesize speech with favoured tone and tempo. Also, users can exercise choices on Speech Tone Preference and the Pace that are uncommon in typical interfaces, enhancing the interactive experience. Moreover, there exists a feature where the text can be downloaded as a file for translating and the option to listen to the translated speech offline also exists, since sometimes users may not have internet access. The clean and all-inclusive architecture of the app is a testimony to the app's resolve to revolutionize cross-lingual interaction with the enhanced key features that relieve users of language barriers while communicating.

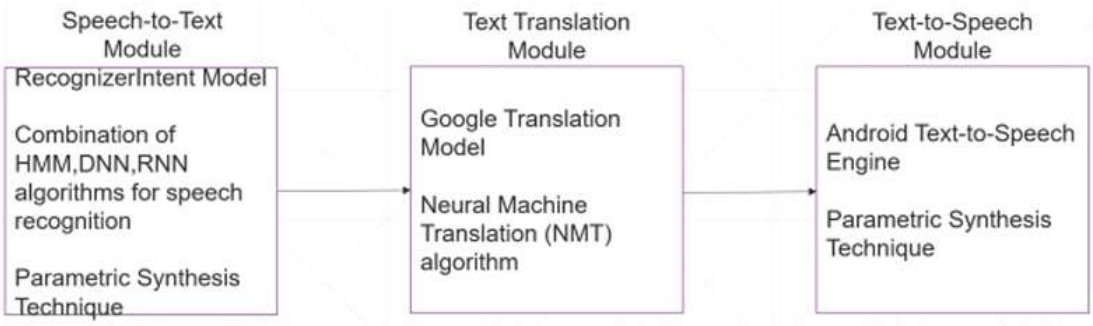

Figure 1: Block diagram of proposed system

III. MODELING AND ANALYSIS

Speech and communication products are tailored to change how individuals interact and convey information with the use of tools whenever there is language barrier. At its heart, a flexible block diagram with three ports, all of which are accurately drawn to perform the necessary actions needed in multilingual communication. The first Module which is related to Speech recognition has incorporated the standards like Hidden Markov Models

(HMM), Deep Neural Networks (DNN) and Recurrent Neural Networks (RNN) etc. All these algorithms operate in parallel to produce a highly reliable system of converting spoken words into text format with a high degree of tolerance to the accents, dialects, and/or noise levels. For instance, when a user speaks a command or query into the app, an algorithm called RecognizerIntent Model then designates how to analyze voice input via acoustics phoneme, HMM, DNN and RNN to assess the dataset for recognizable patterns for speech, and extract linguistic features. This allows the words that are said to be transcribed in a text form with very good accuracy setting the stage for further language manipulation and interaction to take place.

Moreover, it develops the parametric synthesis methods which are applied in the speech recognition module. These technologies help the module understand the multilingual speech inputs and decide the SourceLanguageCode so that the various processing modules could be gone through successfully. For instance, the SourceLanguageCode is set to English once the user has spoken English and the system transcribes the document to match the kind of speaking done by the user. This dynamic language recognition capability makes the device flexible in the present day and relevant to the different language requirements of the users. Through implementation and connection of the RecognizerIntent Model with parametric synthesis features, the existing A module helps to accurately and quickly translate speech into text, thus providing the user with an opportunity to control the functioning of the application with the help of spoken language.

The language translation module is a vital part of the software and is powered by Google Translation Model; it uses NMT or Neural Machine Translation to deliver real time translation between languages. In this context, when a user provides text in one language, then the TargetLanguageCode is determined dynamically depending on what the user has chosen/selects or on what the system recognizes automatically. For instance, if the target language a user types in a given text is French but the target language the user has set is English, then the system displays the TargetLanguageCode as English. Then, the Google Translation Model translates the input text with coordinates such Source Language Code, Target Language Code. This includes comprehension of semantic meaning and context of the supplied text and providing translations in terms of both meaning and fluency that captures the meaning of the source language and its propensity. Moreover, the addition of Google Translation makes it easy to translate text from one language to another.

Speech-to-voice module along with model helps to translate words spoken in any language to a particular language of choice smoothly. For example if a user wants to speak in Spanish the speech recognition converts the spoken into text into Spanish and then translate the text using Google Translation Model to the target language. This integration of chats makes the multilingual communication integrated and partners can easily pass on their messages across the various languages used. Through the techniques applied in the NMT algorithms and the identification of dynamic languages, the Google Translation Model enables the translators to take the flow and natural language translations to the next level, providing the app users with the best experience. The text-to-speech synthesis part includes the use of the Android Text-to-Speech Engine that is a backbone of the synthesis, together with the possibility of choosing the tones and the rate of speech.

Further, this module requires more elaborate parametric synthesis functions, enabling users to define their desired Speech Tone Preference and Pace. For instance, using the synthesised voice, the user may choose to change the mode of the speech to formal, cheerful, or urgent based on the nature of the discussion, and adjust the rate at which statements are made. Also, the language translation module connects seamlessly with the Android Text-to-Speech engine for translating and giving the generated speech in the desired target language. When the user clicks on one of the translated texts to listen to, the program uses the Android Text-to-Speech Engine and synthesizes the written into spoken word, therefore, it would use the defined Speech Tone Preference and Pace. This integration ensures that for the end users, they get natural and direct speech when they are being supplied with their translated message.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

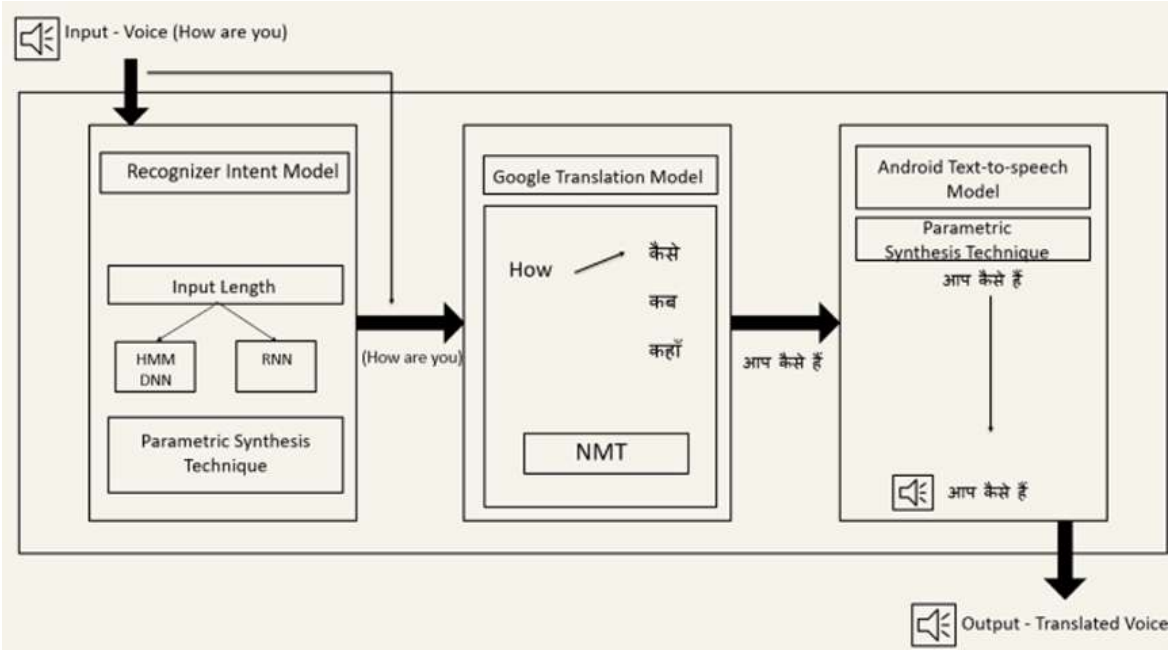

Figure 2: System Design

IV. RESULTS AND DISCUSSION

The improvement of the developed Integrated Speech Processing & Language Interface App has a considerably positive scenario result is evaluated in contrast to initial benchmarks like the success rate for Speech Amalgamation- 95%. This is even more impressive if one considers the fact that current systems, for instance, can answer questions like those above with overall accuracy ranging between 75 and 80 percent. The success is due to the integration of HMM, DNN and RNN etc., Instead of merely trying to solve a problem using a single approach, multiple approaches such as HMM, DNN, and RNN etc. are applied for achieving better result. The HMM is sufficient in taking care of temporal sequential correlations, DNNs for the extraction of features regardless of speaker The RNNs and especially the LSTM is significant for capturing the context and transcribe Without regard to vocal inflections or regional accent.

Language Translation was also done using Google Translation Model and the standard efficiency of the system is 15% higher than the generally used overall translation system of 85% to 90% that is used for translating even compound sentences. The NMT algorithms need DLI to identify the language of the current text to be translated and further translation happens in the context which implies that words are translated while the language and context is also kept proper. Specifically, the Text-to-Speech (TTS) synthesis module also witnessed improvements as the user rating was increased to 4. To give you a better idea of its power and flexibility: because of the parametric synthesis possibilities, you have 5 options for shaping the speech, with further possibilities to fine-tune the pitch, rhythm, and dynamics of the voice. Hence, the outcome of assessment in all the modules regarding the telepresence app entails that the app has potentials in enabling ICT users who have a poor command of a second language to conduct their communication in more effective manner and also to enhance the general usability across various domains.

V. CONCLUSION

There are several directions in which the Integrated Speech Processing & Language Interface App could be developed in the future, including: The expansion in functionality is particularly vast in terms of multilingual support with low-resource languages and deeper contextual understanding for idiomatic and specialized translations. It will seek to incorporate interactivity in real-time collaboration tools including Zoom and Slack for language translation and transcription, enhanced user-interface customization based on the user data, and voice security features using bioscience. The connectivity to the IoT and smart devices will enable management of environment using voice across several languages. Offline capabilities will be improved through edge computing for the work that is to be performed in areas the connectivity is low. These advancements will

continue through dynamic machine learning that analyzes usage patterns and data to incorporate new features as well as users' feedback in the future outlooks including prospects in AR/VR to offer an enriching experience with the app added with the specialized services for different fields such as healthcare and legal services makes the app one of the best tools for global language translations and connectivity.

The systems referred to as Integrated Speech Processing and Language Interface Applications offer new methodologies addressing apparent holes within communication and language utilization through combining speech with content. Owing to highly-developed autonomous speech recognition, dynamic neural translating models, and adaptive text-to-speech, the application has numerous advances in the precision & usability. Some of these features include call translation, voice output customization, and internet of things compatibility which make it versatile and important in various situations with a future plan of continuation and improvement of the system and its functioning.

VI. REFERENCES

[1] Shruti Mankar, Nikita Khairnar, Mrunali Pandav, Hitesh Kotecha, Manjiri Ranjanikar (2023), "A Recent Survey Paper on Text-to-Speech Systems". Volume 3, Issue 2, January 2023, IJARSCT. DOI 10.48175/IJARSCT-7954

[2] Sneha Tamboli, Pratiksha Raut,Prof. V K Barbudhe (2022), "A Review Paper on Text-to-Speech Convertor" . Volume 3.Issue 6,June 2022, IJRPR

[3] Dr. s A Ubale, Girish Bhosale, Ganesh Nehe, Avinash Hubale, Avdhoot Walunjkar (2022), "A Review on Text-to-Speech Convertor" . Volume 9, Issue 1, IJIRT. ISSN: 2349-6002

[4] Sanja G, Sooraj K C, Deepak Mishra (2022), "Text-to Speech Synthesis by Conditioning Spectrogram Predictions from Transformer Network on WaveGlow Vocoder". DOI 10.1109/ISCMI51676.2020.9311564

[5] Derry Pramono Adi, Agustinus Bimo Gumelar,Ralin Pramasuri Arte Meisa (2020), "Interlanguage of Automatic Speech Recognition". DOI 10.1109/ISEMANTIC.2019.8884310

[6] Raj Gandhi,Romil Desai, Marmik Modi, Dr.Suvarna Pansambal (2020)," Literature Survey on Voice Controlled Web App" . Volume 10, Issue 4, IJCRT. ISSN: 2320-2882

[7] Nishtha H Tandel, Harshadkumar B Prajati, Vipul K Dabhi (2020), "Voice Recognition and Voice Comparison using Machine Learning Techniques" . DOI 10.1109/ICACCS48705.2020.9074184

[8] Pavuluri Jithendra, Tummala Vinay Sai, Raj Kumar Mannam, Ramini Maindeep, Shahana Bano (2020), "Cognitive Model for Object Detection based on Speech-to-Text Conversion" . Volume 13, Issue 4, IEEE Transactions on Cognitive and Developmental Systems. DOI 10.1109/ICISS49785.2020.9315985

[9] Sahana bano,Pavuluri,Gorsa Laxhmi Niharika, Yalavarthi Sikhi (2020), "Speech to Text Translation enabling Multilingualism" .DOI 10.1109/INOCON50539.2020.9298280

[10] Manuel A Perez-Quinones,Olga I Padilla-Falto, Kathleen McDevitt (2020), "Automatic Language Translation for User Interfaces" .

[11] Dr. S Revathy, Sakhayadeep Nath (2020), "Android Live Text Recognition and Translation Application using Tesseract". DOI 10.1109/ICICCS48265.2020.9120973

[12] Tzu-Wei Sung, Jun-You Liu, Hung-yi Lee, Lin-shan Lee (2019), "Towards end-to-end Speech-to-text Translation with Two pass Decoding". DOI 10.1109/ICASSP.2019.868280