International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 04 | April - 2022 Impact Factor: 7.185 ISSN: 2582-3930

A Preliminary Work on Speech to Speech Translation

Priyanka Padmane[1], Ayush Pakhale[2], Sagar Agrel[3], Ankita Patel[4], Sarvesh Pimparkar[5], Prajwal Bagde[6]

[1] *Professor, Dept. of Computer Technology, Priyadarshini College of Engineering, Nagpur - India*
[2][3][4][5][6] *Student Dept. of Computer Technology, Priyadarshini College of Engineering, Nagpur - India*

Abstract - The automatic translation with one human language into another is referred to as "machine translation." The main purpose is to bridge the linguistic gap between people who speak different languages, communities, or countries. There are 18 main language and ten scripts that are frequently used. Because the majority of Indians, particularly remote peasants, cannot understand, read, or write English, an excellent language translator is required. Machine translation systems that transform source text to another will assist Indians in living in a more enlightened society without language barriers. Because English is a worldwide language and Hindi is the language spoken by the majority of Indians, we propose an English to Hindi machine translation system based on recurrent neural networks (RNN), LSTM (Long short-term memory), and attention processes. The automatic translation of one natural language into another is referred to as "machine translation." The main purpose is to bridge the linguistic gap between people who speak different languages, communities, or countries. There are 18 official languages and ten scripts that are frequently used. The majority of Indians, particularly isolated peasants, do not understand, read, or write English, necessitating the implementation of an effective language translator. Machine translation systems that convert text from one language to another will help Indians live in a more enlightened society that is free of language barriers.

***Key Words*:** *RNN, LSTM, Speech to text, text to Speech, Multi linguistic.*

1. INTRODUCTION

Nowadays in communication the language barrier has created problem for successful communication for this we introduced this application. Speech recognition and text translation are mainly used for converting the speech to text and text to speech for understanding the language which are spoken by user during communication, because of this person can recognize the speech are spoken by another person.

Machine translation has been a work in progress since 1940. Google Translate has been growing in popularity since 1940. A machine translation system converts text or speech from one human language to another. To convert a document or content from another language into our own tongue, machine translation is required. It dismantles linguistic barriers. NLP, or natural language processing, is an area of computer science that tries to bridge this divide. The principle of neural machine translation is simple, and it requires little domain knowledge. It has been taught to generate extraordinarily long word sequences; a large neural network was used. The model specifically contains large phrase databases and linguistic models, unlike standard machine translation systems. The partnership is in charge of the MT system's first condition antecedent. Machine Translation is significant because of the cultural significance of translation in civilizations where more than one language is spoken. Furthermore, the idea of an attention mechanism is used.

Hindi is a widely spoken language in India and the country's principal official language, while English is spoken all over the world and is thus an internationally recognized language. During the British colonial period in India, English was introduced as a speech language. As a result, both English and Hindi have a large following. As a result, a translator is needed to translate from one language to another. Here, we'll be learning how to translate from English to Hindi. In India, the necessity of employing regional languages like Hindi for document drafting and other purposes is becoming more widely recognized.

In this context, establishing a machine translation system capable of translating English into a variety of regional languages has become crucial. Furthermore, many sites are entirely in English, which is of little benefit to rural residents who do not know English and hence cannot comprehend the information presented. As a result, a translator who can translate from English to Hindi, a language that is widely understood, is necessary.

2. RELATED WORK

The focus of the research is rule-based machine translation. It is based on a multilingual database management and corpus management system. The system architecture's parser and morphological tools examine the grammar of the language specification before converting it to the target language. The technique presented in the study [1] demands a detailed understanding of the grammatical structures of both the source and target languages. In statistical machine translation, statistics are used. This is based on the information theory idea. The translation is guided by the probability distribution. The Bayesian decision rule and statistical theory are used in the technique suggested in the paper [2] to minimize errors. In the approach given in this paper, there is a choose a problem among phrases as well as a language modelling obstacle. [3] A hybrid approach is used for conversion, combining rule-based and statistical machine translation. The architecture includes the splitter, parser, verb ends tagger, sentence rules, reorder, lexical database, and translator. In this project, a splitter breaks the source language into words, and a parser analyses the grammar and semantic structure. The declension tagger inflects nouns, adjectives, and pronouns to indicate singular, plural, case, and gender. The source language is then reordered, and lexical rules are used to translate the destination language. The study [4] makes use of neural google translate. This paper discusses the architecture's coder, decoder, residual connection, and other components. This method models the conditional likelihood of transforming a source sentence to a

target sentence. A more precise translation is obtained using this procedure.
Deep learning is used in a variety of applications, including image processing, big data analysis, speech recognition, and machine translation [2]. Deep learning is one of the subsets of machine learning. Neural machine translation delivers a more precise translation as well as a better representation than ordinary machine translation. By upgrading existing systems with DNN, traditional systems can be made more efficient.
Deep Learning is a relatively new machine learning technology that is widely employed in a range of applications. It assists the system in learning in the same way that humans do, as well as improving its performance through training. Deep Learning algorithms can represent features by mixing supervised and unsupervised learning. This capability is known as feature extraction.
For a better machine translation system, different deep learning algorithms and libraries are necessary. RNNs, LSTMs, and other algorithms are utilized to train the system that will transform the sentence from source to target. Using the appropriate networks and deep learning algorithms is a solid choice because it modifies the system to maximize the translation system's accuracy when compared to others.

Advantages of Neural Machine Translation

- (SMT) models need a fraction of the memory that these models do.
- Deep Neural Nets surpass earlier state-of-the-art algorithms on shorter sentences when big parallel corpora are available.
- To solve the rare-word issue in larger sentences, NMT techniques can be paired with word-alignment algorithms.

3. PROBLEM STATEMENT AND OBJECTIVE

A. Problem Statement

The purpose of our project is to automate the application in order to overcome the language barrier that exists between nations and within countries. The application's many elements will be handled by the above-mentioned program. The purpose of the proposed system is to construct a system that can translate, convert text to speech, recognize speech, and extract text. A limited selection of English words will be used to test the proposed strategy.

B. Objectives

- Our main goal is to combine all of the many functionalities, such as speech recognition, text translation, text synthesis, and text extraction from images, into a single, user-friendly program.
- emission of voice
- Simple to use

4. COMPONENTS OF SPEECH TO SPEECH TRANSLATION

Speech recognition technology which recognizes the user's speech input and converts into source language text or the technology to recognize speech. b) Machine translation which translates source language text into the target language text or the technology to translate the recognized words. c) Speech synthesis or Text-to-speech synthesis which converts translated text into speech or the technology to synthesize speech in the other person's language. In addition, the technology understands natural language and UI-related technology also plays an important role in this speech-to-speech translation system.

5. APPLICATION

As machine translation applications are reaching significantly high accuracy levels, they are being increasingly employed in more areas of business, introducing new applications and improved machine-learning models.

Machine Translation in Industry for Business Use Although big players like Google Translate and Microsoft Translator offer near-accurate, real-time translations, some "domains" or industries call for highly-specific training data related to the particular domain in order to improve accuracy and relevancy. Here, generic translators would not be of much help as their machine-learning models are trained on generic data. These applications are used by small, medium and large enterprises. Some organizations offer multi-domain translation services– that is, customizable solutions across multiple domains–and other organizations offer translation solutions only for a specific domain. These solutions, although automated for the most part, still depend on human translators for pre- and post-editing processes.

Some fields that warrant domain-specific machine

translation solutions are:

- Government
- Software & technology
- Military & defense
- Healthcare
- Finance
- Legal
- E-discovery
- Ecommerce

Benefits of translation technology for healthcare

Reduced costs Healthcare translation technology can significantly reduce costs for hospitals and providers in their interpretation needs, while also boosting productivity. "This sort of technology is a low-hanging fruit CFOs and senior admins hardly recognize," "That cost saving can be leveraged to be used for more critical, clinical applications that are much more sensitive to cost cutting.

Mobility Another benefit of healthcare translation technology is its mobility. Consider the back-up hospitals can face at an emergency department admissions desk due to lack of available interpreters. Having a translation product brought to the ER when needed can reduce wait times for patients. Instead of having to wait for an interpreter to be found, the technology is already available in the hospital. “It would increase the quality of patient care, throughput, and overall healthcare experience, which means patient satisfaction goes way up.

6. FLOWCHART

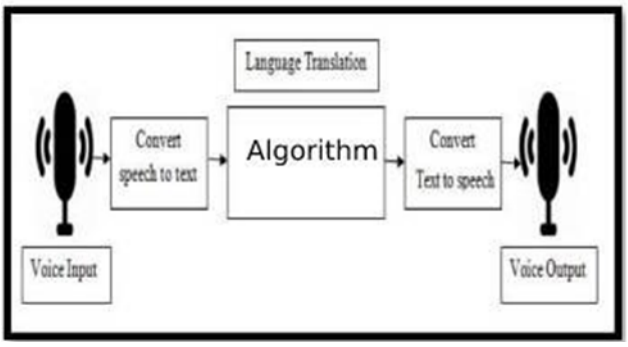


Fig 1: Flowchart

7. ARCHITECTURE

The figure 2 depicts the architectural diagram for speech-to-speech system. First the user has to give the voice input in any of the four languages that are English, Kannada, Hindi and Telugu which then converts speech to text using the Google API. Then it gives into language translation where the preferred language is translated and the meaning of the sentence remains same which is done by computational intelligence. After the language translation it converts text to speech and finally at the last step, we will get the converted voice output. Following steps involves in Speech-to Speech translation.

1)Voice Input:

The voice input is taken as input to the system. The inputvoice may be any of the local languages here we taken for the four languages Kannada, English, Hindi and Telugu.

2)Speech to Text Conversion:

Speech is a particularly attractive modality for human-machine interaction: it is "hands free"; the acquisition requires only modest computer equipment and it arrives at a very modest bit rate. Recognizing human speech, especially continuous speech (connected), without heavy training (independent of the speaker), for a vocabulary of sufficient complexity (60,000 words) is very difficult. However, with modern methods, we can easily process speech signals and convert them into text. The output of this phase is the text that was spoken by the caller.

3)Language Translation:

Language Translation is done by Google Translator.

4)Text to speech Conversion:

The output of the previous phase is Text. The existing speech synthesizer will be adopted which will convert text- to-speech. This speech will be easily understandable in the preferred language selected.

5)Voice output:

After the language translation it converts text to speech and finally at the last step, we will get the converted voice output

The purpose of the project is to produce a voice recognition engine that is simple, open, and widely used.

Simple in the sense that the engine should not run-on server-class hardware. The code & models are open, as they are released under the Firefox Public License. The engine should be ubiquitous in the sense that it should run on a variety of platforms and provide bindings for a variety of languages.

The engine's architecture was inspired by the work presented in Deep Speech: Scaling up edge speech recognition. However, the engine is no longer identical to the one that inspired it in the first place.

A recurrent neural network (RNN) trained to absorb speech spectrum analyzer and generate English text transcriptions lies at the heart of the engine.

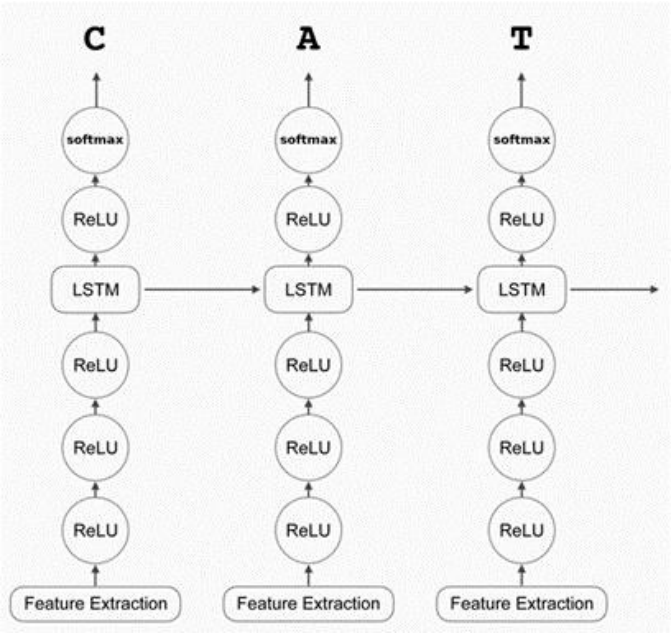


FIG 2: Complete RNN Model

8. FLOWCHART

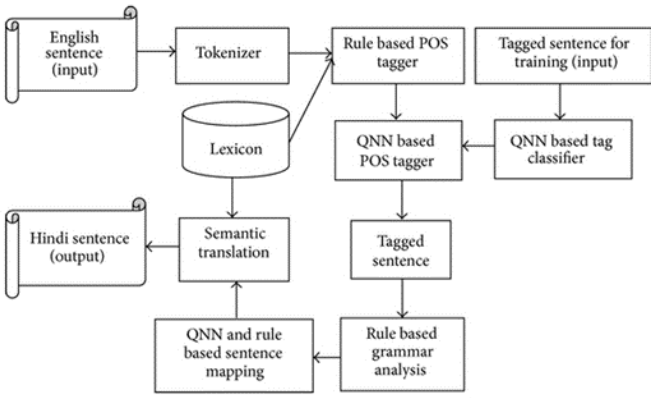


FIG 3: Flowchart

9. PROPOSED WORK

1)Input The voice input is taken as input to the system.

User can give the voice input in any of the four languages they are Kannada, English, Hindi, and Telugu. Classification of Speech Recognition Speech depends on various factors that are speaker, bandwidth and size of the vocabulary. Speech recognition is classified into various types based on utterance, speaker mode and vocabulary size. Classification of Speech Recognition based on the Utterances There are four different types of speech available -Isolated Word, Connected Word, Continuous Speech, Spontaneous Speech. Classification of Speech Recognition based on the Speaker Mode Speech recognition system is broadly classified into main categories based on speaker models, namely, speaker dependent and speaker independent. Classification of Speech Recognition based on the Vocabulary Size The vocabulary size is divided into five categories namely, small vocabulary, medium vocabulary, large vocabulary, Very-large vocabulary, Out-of-Vocabulary. [12]

2)Speech to Text Conversion Input from microphone:

Speech word samples are going to be extracted from no. of samples & unbroken in separate word. The microphone is employed to store the signal into the system. The input is human voice that is sampled at rate of 16,000 per second. It should to run in live mode. Speech generation: Speech generation devices, also known as voice output communication aids, are augmented and alternative electronic communication systems used to replace the speech or writing of people with severe speech to communicate verbally. Speech generation devices are important for people who have limited means of verbal interaction, as they allow individuals to become active participants in communication interactions. They are particularly useful for patients with amyotrophic lateral sclerosis (ALS), but have recently been used in children with predicted speech deficiency. Speech Processing: Speech processing has been defined as the study of speech signals and their processing methods, as well as the intersection of digital signal processing and natural language processing. Speech processing technologies are used for digital speech coding, spoken language dialogue systems, speech-to-text synthesis, and automatic speech recognition. Information (such as speaker identification, gender or language identification, or speech recognition) can also be extracted from speech. Speech can be a more intuitive way to access information, control things and communicate, but there may be viable alternatives: speech is not necessarily the "natural" way to interact with a computer. Speech is hands free, without eyes, fast and intuitive. Text output: Text output is generated after the speech processing.

3) Language Translation

Google's translation system works primarily on text, but has a built-in feature that allows you to pick up a microphone input and then play back the sound from the speakers. Google Translate uses the classic speech-to-speech translation style with the use of a speech identifier for text speech, text translation, and speech synthesis to generate the audio associated with the text. It should be noted that Google's translation service is a difficult candidate to beat in terms of correct translations due to its well-designed implementation with huge amounts of input data from many different sources. [13] A basic machine translation system makes a simple word substitution to convert text from one language to another. The translation process involves decoding the meaning of the source text and re-encoding the meaning in the target language. Word meaning refers to the concept or sense of the source text. Machine Translation Systems can be built using several approaches like Rule-based, Statistical-based, and Example based systems. In Rule-based translation to translate from English-to-Telugu and Example-based translation to translate from Telugu-to-English. In Rule-based translation, we perform source language text reordering, to match target language syntax, and word substitution. On the other hand, in Example-based translation, we map examples of text in source language to the corresponding representations in destination language, to achieve translation.

4)Text to speech Conversion:

The output of the previous phase is Text. The existing speech synthesizer will be adopted which will convert text-to-speech. This speech will be easily understandable in the preferred language selected.

5) Voice output:

After the language translation it converts text to speech and finally at the last step voice output is generated.

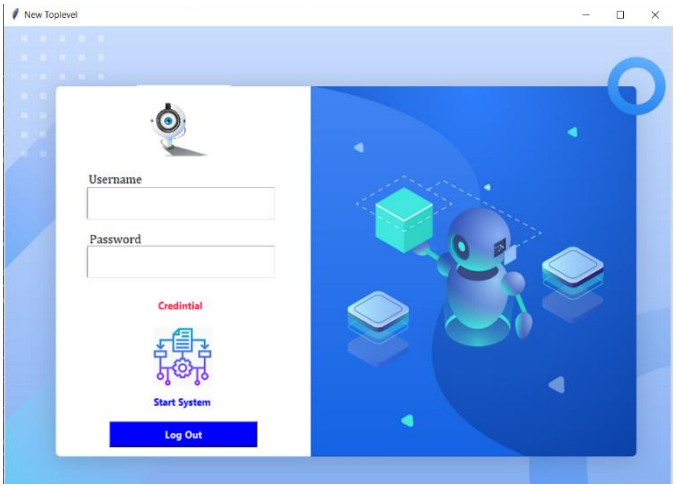
The back end is a search engine that receives data from the front end and searches it across the databases listed below:

The acoustic model is made up of acoustic sounds that have been taught to distinguish different speech patterns.

Text Translation: Text translation is the process of taking a piece of text and converting it into another language. English is the primary language. The text is divided into words, which are then searched in the dictionary, with the appropriate matched text/word shown.

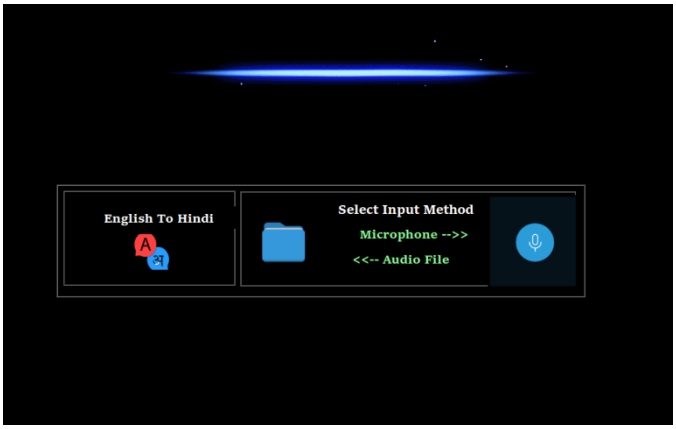
Project Modules:

1. Login: -



User Will login into the system by entering his details and our model will authenticate it with its credentials.

2. Home Screen: -



Home screen is like a dashboard for a user where he will have options for input.

A. Details of hardware and software

Hardware Requirements:

- Hard disk – 500 GB
- System – I5 Processor
- RAM-4 GB

Software Requirements:

- LANGUAGE –
- Python
- Java

FRONT END: HTML, CSS

- APP- Java
- Web – python
- Database – SQLite
- Framework – flask

10. FUTURE WORK

This technology is being implemented for a desktop application, but it can also be used for a mobile phone in the future. As a result, customers can more efficiently use this system by simply pressing a button on their mobile device rather than relying on a desktop for language conversion.

11. CONCLUSION

We implemented the system in this suggested system for users who are experiencing language barrier issues, and the user interface is also user pleasant so that users can easily engage with it. As a result of the fact that this system does not require the use of a dictionary to comprehend the meaning of words, it reduces the user's task of understanding languages for communication.

12. REFRENCES

[1] M.A.Anusuya, S.K.Katti, “Speech Recognition by Machine: A Review”, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009 [2] Shyam Agrawal, Shweta Sinha, Pooja Singh, Jesper Olsen, “Development of text and speech database for Hindi and Indian English specific to mobile communication environment”. [3] D.Sasirekha, E.Chandra, “Text to speech: a simple tutorial”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, And March 2012. [4] A. A. Tayade, Prof.R.V.Mante, Dr. P. N. Chatur,“Text Recognition and Translation Application for Smartphone.

[2] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” arXiv, 2015.

[3] J. Redmon, A. Farhadi, “YOLO9000: Better, Faster, Stronger,” arXiv, 2016.

[4] M. Swathi and K. V. Suresh, "Automatic Traffic Sign Detection and Recognition: A Review," 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), Chennai, 2017, pp. 1-6, https://ieeexplore.ieee.org/document/8186650.

[5] S. Saini and V. Sahula. A survey of machine translation techniques and systems for Indian languages. In 2015 IEEE International Conference on Computational Intelligence Communication Technology, pages 676–681, Feb 2015.

[6] S. Chand. Empirical survey of machine translation tools. In 2016Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pages 181–185, Sept2016.

[7] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. CoRR, abs/1409.3215, 2014.

[8] Kyunghyun Cho, Bart Van Merri¨enboer, Dzmitry Bahdanau, and YoshuaBengio. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259, 2014.

[9] LR Medsker and LC Jain. Recurrent neural networks. Design andApplications, 5, 2001.

[10] Sepp Hochreiter and J¨urgen Schmidhuber. Long short-term memory.Neural computation, 9(8):1735–1780, 1997.[9] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neuralnetworks. IEEE Transactions on Signal Processing, 45(11):2673–2681,1997.

[11] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficultyof training recurrent neural networks. In International Conference onMachine Learning, pages 1310–1318, 2013.