Speech To Speech Translation

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

[1]Professor, Ms. Priyanka G. Padmane, Dept. of Computer Technology, Priyadarshini College of Engineering, Nagpur
[2]Mr. Ayush Pakhale, Dept. of Computer Technology, Priyadarshini College of Engineering, Nagpur
[3]Mr. Sagar Agrel, Dept. of Computer Technology, Priyadarshini College of Engineering, Nagpur
[4]Ms. Ankita Patel, Dept. of Computer Technology, Priyadarshini College of Engineering, Nagpur
[5]Mr. Sarvesh Pimparkar, Dept. of Computer Technology, Priyadarshini College of Engineering, Nagpur
[6]Mr. Prajwal Bagde, Dept. of Computer Technology, Priyadarshini College of Engineering, Nagpur

---

Abstract— *The term "machine translation" refers to the automatic translation of one natural language into another. The fundamental goal is to bridge the language divide between people, communities, or countries who speak different languages. There are 18 official languages and ten widely used scripts. 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. We propose an English to Hindi machine translation system based on recurrent neural networks (RNN), LSTM (Long short-term memory), and attention mechanisms, as English is a worldwide language and Hindi is the language spoken by the majority of Indians.*

Keywords—RNN, LSTM, Speech to text, text to Speech, Multi-linguistic.

I. INTRODUCTION

Since 1940, machine translation was a work in progress. Since then, google translate has been increasing in popularity. Text or speech is translated from one human language to another by a machine translation system. Machine translation is required to turn a document or text from another language into our own tongue. It breaks down linguistic barriers. NLP is a branch of computer science that aims to close this gap. Neural Machine

Translation is conceptually simple and requires little domain knowledge. A large neural network has been taught to generate extremely lengthy word sequences. Unlike traditional machine translation systems, the model explicitly stores big phrase databases and language models. The collaboration is responsible for first condition precedent of the MT system. The cultural significance for translation in societies in which more than one language is spoken gives rise to the significance of Machine Translation. In addition, the concept of an attention mechanism is employed.

Hindi is a commonly spoken language and India's primary official language, but English is being spoken worldwide and hence is a globally recognised language. In India, English as a verbal was introduced during the British period. As a result, both English and Hindi are widely spoken languages. As a result, a translator is required to convert one language to another. We'll be studying English to Hindi translation here. In India, there is now a growing understanding of the importance of using regional languages such as Hindi for document drafting and other uses.

In this setting, developing an MT system that really can translate English into various regional languages has become critical. Furthermore, many sites are all in English, which is of little value to rural people because they do not speak English and hence cannot understand the information provided on the site. As a result, a translator is required who can translate English to Hindi, which is commonly comprehended by the general public.

II. RELATED WORK

The work is primarily concerned with rule-based machine translation. It is built on a bilingual database and corpus management system. The parser and morphology tools in the system architecture analyse the grammar of the language specification and then convert it into the chosen language. The strategy proposed in the research

[1] necessitates a thorough comprehension of both the source and target languages' grammatical structures. Statistics are used in statistical machine translation. This is based on the concept of information theory. The probability distribution is used to guide the translation. To reduce errors, the strategy proposed in the paper

[2] employs the Bayesian decision rule and statistical theory. There is a choose a problem among phrases and a language modelling challenge in the approach outlined in this study.

[3] For conversion, a hybrid approach is utilised, which combines rule-based and statistical machine translation. The splitting, parser, verb endings tagger, sentence rules, reorder, lexical database, and translator are all part of the

architecture. In this work, the source language is broken into words by a splitter, and then the grammar and semantic structure are analysed by a parser. To denote singular, plural, case, and gender, the declension tagger inflects nouns, adjectives, and pronouns. The source language is then reordered, and the target language is translated using lexical rules. The neural google translate is used in the study.

[4] The coder, decode, residual connection, and other components of the architecture are covered in this work. The conditional probability of converting a source sentence to a target sentence is modelled in this method. This method yields a more precise translation.

Deep Learning is a relatively new method that is widely used in a variety of machine learning applications. It helps the system to learn in the same way that humans do and to enhance its performance through training. By combining supervised and unsupervised learning, Deep Learning algorithms can represent features. Feature extraction is the term for this capacity. The feature could be something as simple as colour or connection with the performance. It might be far more difficult, such as facial recognition. Deep learning is being employed in image processing, big data analysis, speech recognition, and machine translation, among other applications [2]. Machine learning includes a subset called deep learning. Compared to standard machine translation, neural machine translation provides a more precise translation as well as a better representation. DNN can be used to make traditional systems more efficient by enhancing them with DNN. For a better machine translation system, different deep learning algorithms and libraries are necessary. RNNs, LSTMs, and other algorithms are utilised 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 maximise 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.

III. PROBLEM STATEMENT AND OBJECTIVE

A. Problem Statement

Our project's goal is to automate the application in order to overcome the language barrier that exists between countries and states within a country. The above-mentioned programme will perform the various aspects of the application. The suggested system's goal is to create a system that can conduct translation, text to speech conversion, speech recognition, and text extraction. The suggested method will be created for a small set of English words.

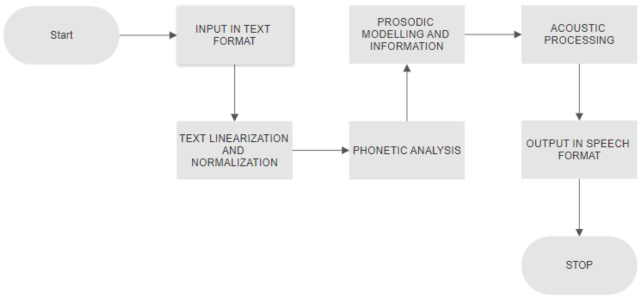
B. Objectives

- Our main goal is to integrate all different functions, such as speech recognition, text translation, text synthesis, and text extraction from images, into a single application that is easy to use.

- Voice output

- Easy to use

IV. FLOW CHART



V. ARCHITECTURE

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 analyser and generate English text transcriptions lies at the heart of the engine.

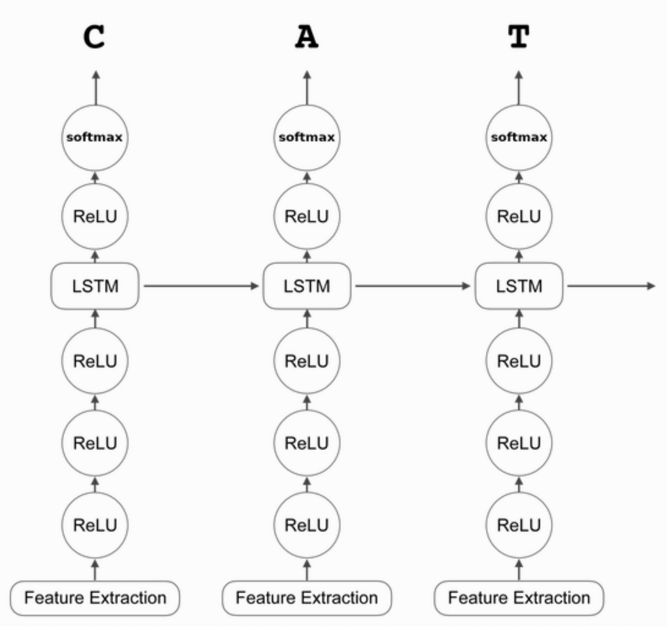


Fig 1: Complete RNN model

VI. FLOW CHART

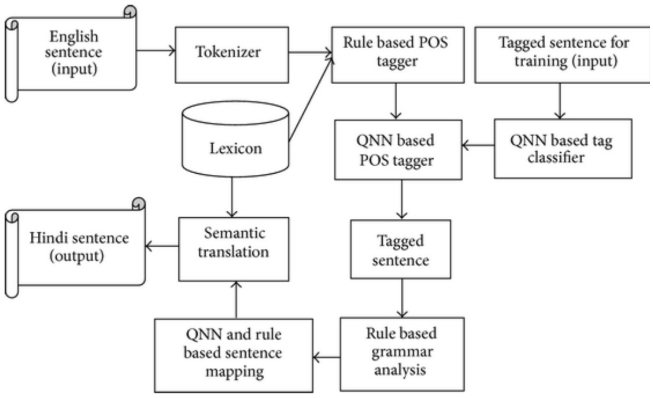


Fig 2: Architecture

VII. PROPOSED WORK

Pre-processing, POS-tagging, Rule creation, Rule Match, Rule Extraction, Word2vec, and Translation are some of the methodologies utilised in machine translation. Using these seven components, the system can perform three tasks related to text processing. The first step is to use the Stanford library for tagging. The second step is to retrieve the reordering rules using exact matching or fuzzy matching using cosine similarity applied to the rules, and the third step is to translate using a bilingual corpus and the retrieved reordering rules, as well as certain categorised words. The following is a description of how all modules work: Our system consists of four modules: text to speech, speech to text, image extraction, and language translator, all of which are interconnected.

Converting Text to Speech The basic goal of a text to voice conversion system is to transform any text into speech, whether it is random or chosen. By tuples recorded speech kept in a database, speech could be obtained as an output.

There are primarily two components: the first is input text processing, and the second is voice language conversion.

Speech synthesis is the process of converting text to speech, and the computer system used for this is known as a speech synthesiser. It consists of two fundamental jobs, one of which is text normalisation and the other is tokenization. Each word is given a phonetic transcription in this procedure. Make phrases, clauses, and sentences out of it. Analysis and Detection of Text

It analyses the supplied text and organises it into a list of words in text analysis. It also looks up the word in the database. Normalization and Linearization of Text

Text normalisation is the process of converting text into a form that can be spoken.

The text normalisation procedure converts uppercase and lowercase letters while also removing punctuation marks. It's better for comparing characters that have the same meaning. "Can't" and "Cannot," "I've" and "I've," "Don't" and "Don't not," and so on. Normalization is divided into four phases: abbreviation conversion, part of a word, number conversion, and acronym conversion.

Processing of Sound Finally, the speech waveforms for each word and sentence is created using both phonemes and prosody. There are two processing methods: the first is synthesis of recorded speech segments. The term "chunk" refers to a cluster of words. The second step is signal processing-based formant synthesis. The next module we'll look at is extracting text from photos. If necessary, the text collected from the images is given to the Text to Speech module.

Recognition of speech: - The voice recognition engine uses audio as input, converts it to text, and then returns the text to the user. There is a forward end and a back end to the voice recognition process. The audio is processed on the front end, which isolates sound segments and converts them to numeric values. This value is utilised in signal to categorise the vocal sound. 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. The Lexicon database contains all of the language's terms and teaches you how to pronounce them. 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. Sign Up: - If a user doesn't have any account he/she needs to create by adding basic details.

3. Home Screen: - Home screen is like a dashboard for a user where he will have options for input.

4. Audio and live mic input. :- Input will have two types first as a pre-recorded audio another will be live input in English.

5. STT model: - Input will be passed to the model for translation.

6. English to Hindi model: - After successful translation of

English input the output will be in Hindi for the given input.

A. Details of hardware and software

Hardware Requirements:

- Hard disk – 500 GB

- System – *I5* Processor

- RAM-4 GB

Software Requirements:

- LANGUAGE –

- Python

FRONT END: HTML, CSS ▪ APP- Java ▪ Web – python ▪ Database – SQLite ▪ Framework – flask

VIII. 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.

IX. 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.

REFERENCES

[1] [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. 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, Sept 2016.

[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 Yoshua Bengio. 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 and Applications, 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 neural networks. IEEE Transactions on Signal Processing, 45(11):2673 2681,1997.

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