

**Sinhgad Institutes**

**Savitribai Phule Pune University**

**Project Review on  
"Translation as a Service"**

**By**

**Akhil A. (B191058504)**

**Balraj Upadhyay (B191058507)**

**Ojas Bhat (B191058512)**

**Abhang Tiwade (B191058594)**

**Pralhad Yadav (B191058600)**

**Under the Guidance of  
Prof. W.P.Rahane**

**In partial fulfillment of  
S.T.E.S's**

**NBN SINHGAD TECHNICAL INSTITUTE CAMPUS, PUNE-41**

**B.E (INFORMATION TECHNOLOGY)**

**Savitribai Phule Pune University**

**May-June 2022-2023**



## Sinhgad Institutes

**NBN SINHGAD TECHNICAL INSTITUTE CAMPUS, PUNE-41**

**B.E (INFORMATION TECHNOLOGY)**

---

### CERTIFICATE

This is to certify that the Project entitled

**"Translation as A Service"**

Submitted By

| <u>Name of Student</u> | <u>Exam Seat No.</u> |
|------------------------|----------------------|
| Akhil A.               | B191058504           |
| Balraj Upadhyay        | B191058507           |
| Ojas Bhat              | B191058512           |
| Abhang Tiwade          | B191058594           |
| Pralhad Yadav          | B191058600           |

Date: January 2023

Place: Pune

Prof. W. P. Rahane

**Internal Guide**

Prof. R. M. Samant

**H.O.D Of Information Technology**

Prof. S.P. Patil

**Principal**

## ACKNOWLEDGEMENT

A project is defined as a piece of work that needs skill, effort and careful planning. But during the course of project we found that it not only sharpened our logic skill but also taught us the value of joint effort and hand work. A successful project is the result of good team work, which contains not only the people who put their login and hand work but also those who guide them.

We would like to take this opportunity to thank our project guide **Prof. W.P. Rahane**, who channelized our raw ideas and gave us the encouragement to pursue our goals to realize this project.

We would like to express our deep gratitude to our **H.O.D. Prof. R. M. Samant**, who gave this opportunity to develop this project.

Also, we would like to give thanks to our **Hon' Principal Prof. S. P. Patil**.

Akhil A. (B191058504)

Balraj Upadhyay (B191058507)

Ojas Bhat (B191058512)

Abhang Tiwade (B191058594)

Pralhad Yadav (B191058600)

## **Abstract**

There has been a rise in work from home working culture, where everyone needs to attend online meetings. Many employees from different locations join the meeting, so the issue of the language barrier comes in to resolve this issue a translator is used, but information about the company is confidential. A key tool for addressing this issue is Translation as a Service. The main function of Translation as a Service is to automatically translate the person's speech into the desired language. Neural Machine Translation is the algorithm used in this model. The goal of Neural Machine Translation is to take a sentence in one language as input and return the sentence translated into a different language. After using the Translation as a Service model employees can translate real-time meetings which will help in reducing the language barrier.

# Contents

|                                 |           |
|---------------------------------|-----------|
| Abstract . . . . .              | 1         |
| <b>List of Figures</b>          | <b>7</b>  |
| <b>List of Tables</b>           | <b>8</b>  |
| <b>1 Introduction</b>           | <b>10</b> |
| 1.1 Overview . . . . .          | 11        |
| 1.2 Motivation . . . . .        | 12        |
| 1.3 Problem Statement . . . . . | 12        |
| 1.4 Objectives . . . . .        | 13        |
| 1.5 Project Planning . . . . .  | 14        |
| <b>2 Literature Survey</b>      | <b>15</b> |
| 2.1 Literature Survey . . . . . | 16        |

|          |   |           |
|----------|---|-----------|
| 2.1.1    | Literature Survey on Attention Is All You Need . . . . .  | 16        |
| 2.1.2    | Literature Survey on Google's Neural Machine Translation<br>System: Bridging the Gap between Human and Machine<br>Translation . . . . . | 17        |
| 2.1.3    | Literature Survey on Applications Research of Machine Learn-<br>ing Algorithm in Translation System . . . . .                           | 18        |
| 2.1.4    | Literature Survey Unsupervised Neural Machine Translation<br>With Cross-Lingual Language Representation Agreement . . .                 | 19        |
| 2.1.5    | Literature Survey on Overview of Neural Machine Transla-<br>tion for English - Hindi . . . . .  | 20        |
| 2.1.6    | Literature Survey on Paper 6 . . . . .  | 21        |
| 2.1.7    | Literature Survey on Paper 7 . . . . .  | 21        |
| 2.1.8    | Literature Survey on Paper 8 . . . . .  | 21        |
| 2.2      | Summary of Literature Survey . . . . .  | 21        |
| 2.3      | Gap Analysis . . . . .  | 21        |
| <b>3</b> | <b>System Design</b>  | <b>22</b> |
| 3.1      | System Architecture . . . . .   | 23        |
| 3.2      | Data Flow Diagrams . . . . .  | 24        |
| 3.2.1    | DFD Level 0 . . . . .   | 24        |

|          |                                     |           |
|----------|-------------------------------------|-----------|
| 3.2.2    | DFD Level 1 . . . . .               | 25        |
| 3.2.3    | DFD Level 2 . . . . .               | 26        |
| 3.3      | UML Diagrams . . . . .              | 27        |
| 3.3.1    | Use Case Diagram . . . . .          | 27        |
| 3.3.2    | Activity Diagram . . . . .          | 29        |
| 3.3.3    | Sequence Diagram . . . . .          | 31        |
| 3.3.4    | Class Diagram . . . . .             | 32        |
| <b>4</b> | <b>Other Specifications</b>         | <b>33</b> |
| 4.0.1    | Advantages . . . . .                | 34        |
| 4.0.2    | Limitations or Challenges . . . . . | 34        |
| 4.0.3    | Applications . . . . .              | 34        |
| <b>5</b> | <b>Summary</b>                      | <b>35</b> |
| <b>6</b> | <b>References</b>                   | <b>36</b> |
|          | <b>Bibliography</b>                 | <b>37</b> |

# List of Figures

|     |                               |    |
|-----|-------------------------------|----|
| 3.1 | System Architecture . . . . . | 23 |
| 3.2 | DFD Level 0 . . . . .         | 24 |
| 3.3 | DFD Level 1 . . . . .         | 25 |
| 3.4 | DFD Level 2 . . . . .         | 26 |
| 3.5 | Use Case Diagram . . . . .    | 28 |
| 3.6 | Activity Diagram . . . . .    | 30 |
| 3.7 | Sequence Diagram . . . . .    | 31 |
| 3.8 | Class Diagram . . . . .       | 32 |



# List of Tables

1.1 Project Planning . . . . . 14

# Notations and Abbreviations

**NMT: Neural Machine Translation**

**NMTS: Neural Machine Translation System**

**CLE: Cross Lingual Embeddings**

**LSTM: Long Short Term Memory**

**UNMT:- Unsupervised Neural Machine Translation System.**

# **Chapter 1**

## **Introduction**

## 1.1 Overview

There has been a rise in online meetings because of work from home culture. There are employees from different countries who are not always comfortable with English, therefore company's translator easily translates the language into comfortable languages. There are several languages across the world and each language is different from others in one sense or the other. Like English, Hindi, Spanish, French, Russian, Japanese, Korean, etc. Also, the company's data is confidential so the translator should be of the company's trust, otherwise, companies can suffer because of data loss. To resolve this issue translator was introduced using machine learning techniques. There are various AI-based translators like Google Translate, Microsoft Translate, Amazon Translate, and many more.

## **1.2 Motivation**

Recently online meetings have taken a rise, but during the meetings, if there are people from many countries then it is very hard to explain everyone everything in their native language. So, to improve the functionalities Translation as a Service model is used by which the user input is taken as voice it is converted to text. After that text is converted into the required language after that we convert the text to voice which the people can hear the translated language output which will help them understand what the presenter wants to present. The input will be taken from the inbuilt microphone and the output will be given back from the speaker.

## **1.3 Problem Statement**

To build a deep learning model using Neural Machine Translation which can convert one language speech to required language and will give output in form of speech which will be useful in meetings

## **1.4 Objectives**

The project is aimed to design a deep-learning model with the following features

1. Capturing the Speech
2. Importing pre-trained model
3. Converting Speech to Text
4. Translating text to required language
5. Converting text to Speech

## 1.5 Project Planning

| Sr. No. | Activity   | Start Date          | End Date          |
|---------|--|---------------------|-------------------|
| 1       | Topic selection: TaaS  | 5 August 2022       | 8 August 2022     |
| 2       | Feasibility study and research   | 5 August 2022       | 16 August 2022    |
| 3       | Research paper summary   | 12 August 2022      | Ongoing           |
| 4       | Designing of the system architecture, and approval of topic                                    | 16 August 2022      | 20 August 2022    |
| 5       | Workbook printout  | 20 August 2022      | 20 August 2022    |
| 6       | Project Review 1: Presentation of the system architecture model, and further approach on topic | 3 September 2022    | 3 September 2022  |
| 7       | Further research on the topic of speech to text and text to speech translation                 | 4th September 2022  | Ongoing           |
| 8       | Software design planning   | 4th September 2022  | 10 September 2022 |
| 9       | Development of the translator model using NMT Neural machine translator (Basic model)          | 15th September 2022 | Ongoing           |
|         | Project Review II  | 21 November 2022    | 21 November 2022  |

Table 1.1: Project Planning

## **Chapter 2**

### **Literature Survey**

The main aim of the speech-to-speech model attempts is to provide one language to the required translated languages and giving again the output in speech. This chapter gives better insights into the project through the analysis done on various research papers related to the Translation There are various techniques as well as various models/ algorithms to perform the translation of languages



## 2.1 Literature Survey




### 2.1.1 Literature Survey on Attention Is All You Need

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best-performing models also connect the encoder and decoder through an attention mechanism.

They propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely.

Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train.

**Methodology Used:** Recurrent neural networks, long short-term memory with an attention mechanism

**Reference:** Attention Is All You Need - Ashish Vaswani  Google Brain  
avaswani@google.com Noam Shazeer  Google Brain noam@google.com Niki  
Parmar  Google Research nikip@google.com

## **2.1.2 Literature Survey on Google's Neural Machine**

### **Translation System: Bridging the Gap between Human and Machine Translation**

NMT systems are known to be computationally expensive both in training and in translation. NMT systems lack robustness, particularly when input sentences contain rare words. These issues have hindered NMT's use in practical deployments and services, where both accuracy and speed are essential.

This paper consists of a deep LSTM network with 8 encoder and 8 decoder layers using residual connections as well as attention connections from the decoder network to the encoder. To improve parallelism and therefore decrease training time, our attention mechanism connects the bottom layer of the decoder to the top layer of the encoder. To improve handling of rare words, we divide words into a limited set of common sub-word units ("word pieces") for both input and output.

This method provides a good balance between the flexibility of "character"-delimited models and the efficiency of "word"-delimited models, naturally handles translation of rare words, and ultimately improves the overall accuracy of the system.

**Methodology used:** Neural Machine Translation System

**Reference:** Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation - Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi  
yonghui,schuster,zhifengc,qvl,mnorouzi@google.com

### 2.1.3 Literature Survey on Applications Research of Machine Learning Algorithm in Translation System

Machine translation has significantly evolved over time, especially in terms of accuracy levels in its output. Machine translation is a study spanning many fields. Machine translation rules system can be divided into two categories, binary translation rules and ternary translation rules.

The design and implementation of the Chinese English neural machine translation system mainly includes pre-processing module, encoding and decoding module and attention module. The sequence of each module in accordance with the system includes pre-processing module, encoding Overall, divide and conquer the longest noun phrase and sentence frame by the divide and conquer strategy, and in part use the method of multi sequence coding to translate the longest noun phrase and sentence frame.

These are mostly used by individual consumers, such as travelers, students, etc. Real-time translation applications most commonly offer:

**References:** [1] Haque A U, Mandal P, Meng J, et al. Wind speed forecast model for wind farm based on a hybrid machine learning algorithm[J]. International Journal of Sustainable Energy, 2015, 34(1): 38- 51. [2] Bahar P, Alkhoul T, Peter J T, et al. Empirical investigation of optimization algorithms in neural machine translation[J]. The Prague Bulletin of Mathematical Linguistics, 2017, 108(1): 13-25

## **2.1.4 Literature Survey Unsupervised Neural Machine Translation With Cross-Lingual Language Representation Agreement**

Unsupervised NMT is a challenging scenario, however it may allow to overcome some of the most pressing problems related to MT for low-resource languages. Although the field of UNMT is fairly new, it has been evolving rapidly.

Current UNMT models achieve results that are promising, yet we think that the performance of UNMT can be further improved with a better initialization.

Cross-lingual embeddings (CLEs) allow to learn multilingual word representations.

CLEs are important components that facilitate cross-lingual transfer in current NLP models and proved to be very useful in downstream NLP tasks (such as MT). Importantly, CLEs help to bridge the gap between resource-rich and low-resource languages, because CLEs can be trained even without using any parallel corpora

### 2.1.5 Literature Survey on Overview of Neural Machine Translation for English - Hindi

In this paper they used Word2Vec model that processes text. Word2Vec creates vectors that are distributed numerical illustrations of word features, features such as the context of singular words.

This algorithm has two flavors CBOW and Skip-Gram. The purpose of Skip-gram is to predict source context-words from the target words. Then we use POS-Tagging in which if the word is present in sample database then we directly assign the POS-tag from the database. Then we can use exact matching or fuzzy matching. For the Translation reordered rule is extracted from the databases then that rule is applied on source sentence which translates words from source language to required languages.

The evaluation of this method is, if the model is of short messages then you get upto 85 percent accurate result but for lengthy messages the result is of 60 percent accurate results.

**Reference:** Shashi Pal Singh, Ajai Kumar, "Machine Translation using Deep Learning: An Overview", International Conference on Computer, Communications and Electronics, 2017. Hugo Larochelle, Yoshua Bengio, Jérôme Louradour, Pascal Lamblin, "Exploring strategies for training deep neural networks", Journal of Machine Learning Research 1 (2009) submitted 12/07.

### **2.1.6 Literature Survey on Paper 6**

### **2.1.7 Literature Survey on Paper 7**

### **2.1.8 Literature Survey on Paper 8**

## **2.2 Summary of Literature Survey**

## **2.3 Gap Analysis**

- The more dataset we create for this model, the more accurate the model will be.
- The subtask we are performing are very domain specific so a lot of study is done on the topic.

## **Chapter 3**

# **System Design**

### 3.1 System Architecture

The system architecture diagram represents a high level model of the whole project showcasing the important modules of the program. The request for translation for the audio originated from the user, which would be done by the user interface and the designed API. This request is then transmitted to the server where, it is first of all validated by server and further will be processed through the three models, i.e., the speech synthesizer, the translator model, and finally by the text to speech synthesizer. This after successful processing is transmitted back to the user, where is able to hear the audio is in the desired language.

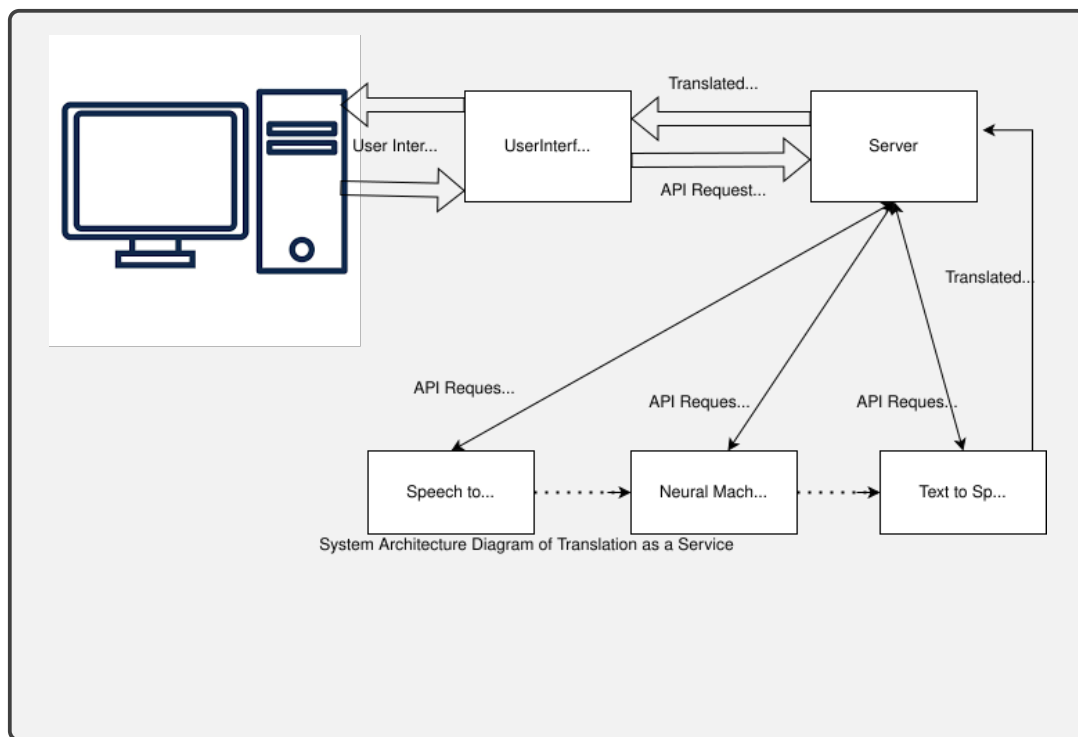


Figure 3.1: System Architecture



## 3.2 Data Flow Diagrams

### 3.2.1 DFD Level 0

The DFD Level 0 diagram show case the abstract view of the project, where the user shall be providing the audio source and then receive a translated audio output in the desired language.

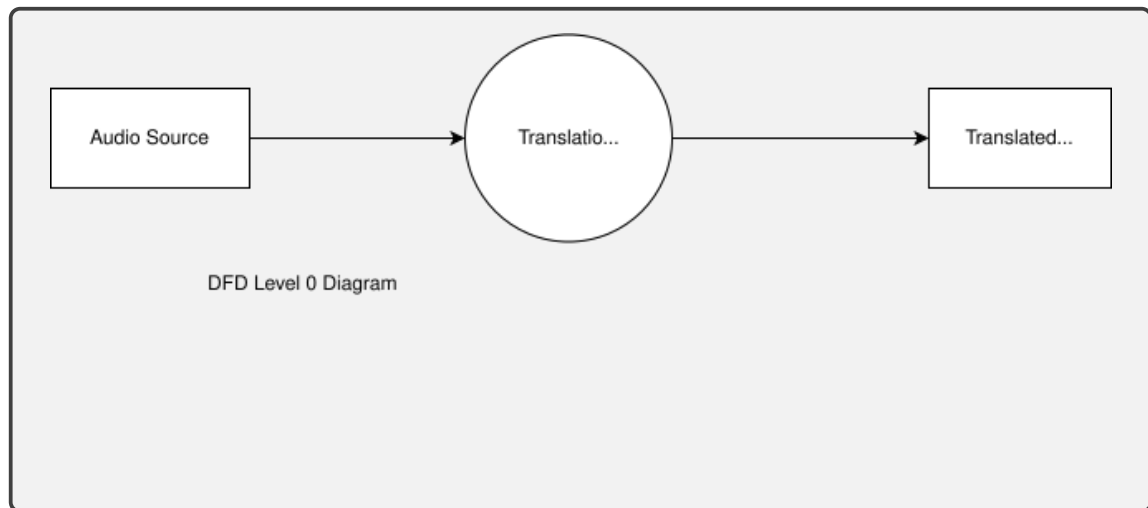


Figure 3.2: DFD Level 0

### 3.2.2 DFD Level 1

The Level 1 diagram, represents the module where the software resides and how the server interacts with the model for translation.

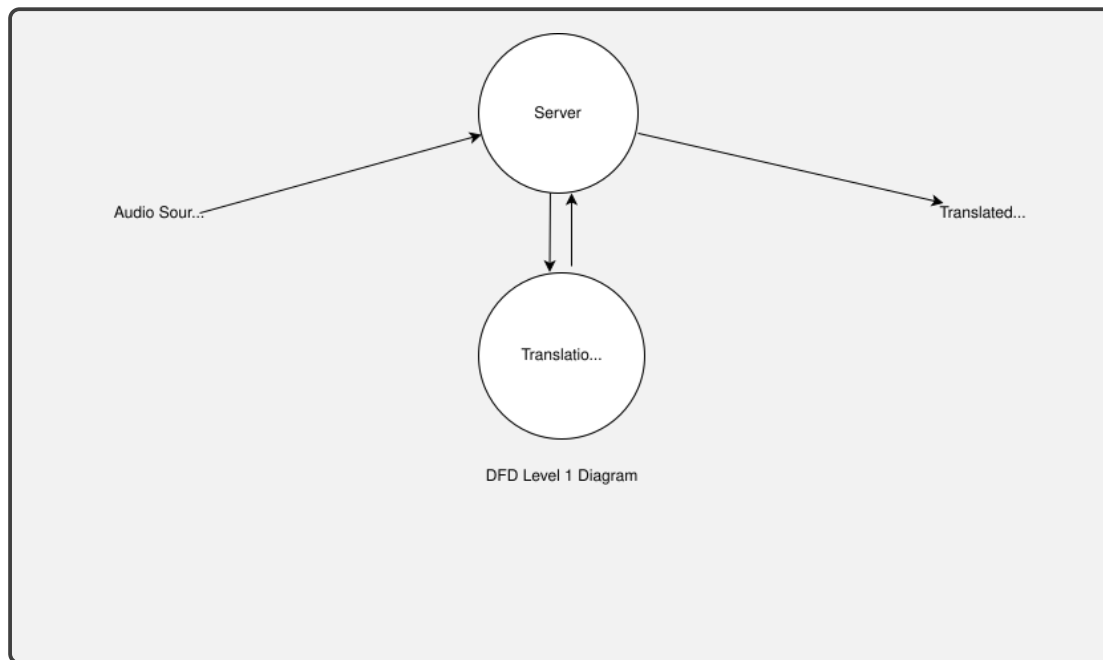


Figure 3.3: DFD Level 1

### 3.2.3 DFD Level 2

The DFD Level 2 represents each module present at the server and how the data is being processed from one to another. It has three models, the speech to text synthesizer, the translator, and the text to speech synthesizer.

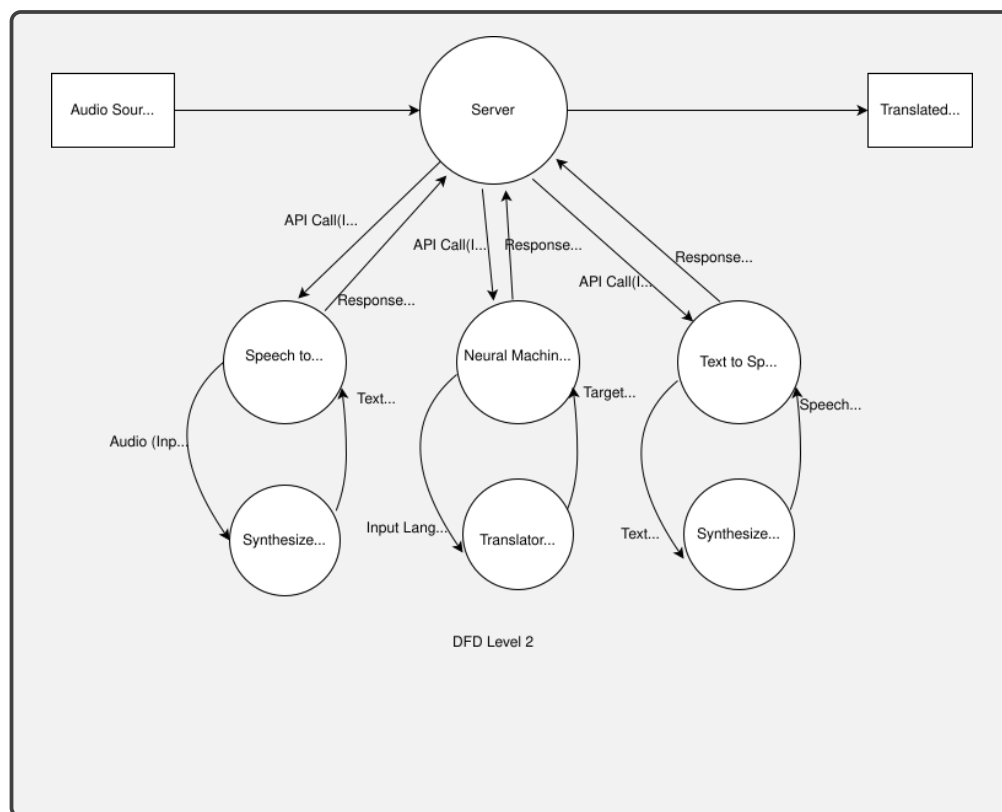


Figure 3.4: DFD Level 2

## **3.3 UML Diagrams**

### **3.3.1 Use Case Diagram**

The use case diagram shows how the all actors, i.e., user and system interact with each other to make the translation successful.

The user request through the user interface by providing the appropriate input data, audio source. The server is sent a. API request for the translation. The server initiates the required sequences, and checking all the interaction and transmissions are working appropriately. The system provides the services to carry out the translation.

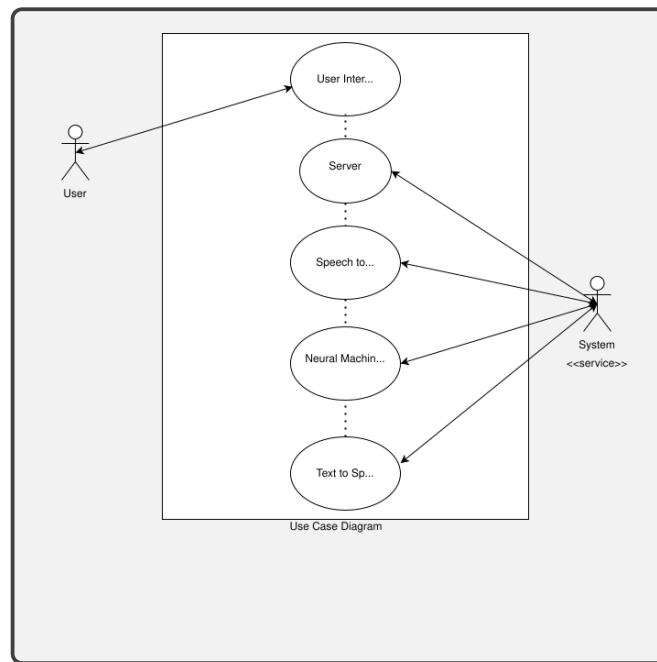


Figure 3.5: Use Case Diagram

### 3.3.2 Activity Diagram

The activity diagram show case all the activities that occur from initialing the sever with the deployed model, to the successful interaction and translation of the audio into the desired language.

In this first the server is initiated with the model deployed in it, and it checks for the status of the server. If the status is in working state, then further actions are proceeded. Otherwise, it checks for the status until it gets activated. Then the server listens for the request for services.

The user starts the application section, and the connection with the server is established, for the further services. After that the user send the input data, i.e., the audio file, along with the current language and the desired language names. The server then listens to the requests and check if it is a valid request and the inputs are in appropriate format. If it is valid request then further processing is done. Otherwise, error message is displayed and the server again listens for the request and data.

The server then transmits the request to the three models sequentially, along with checks for any error. At each stage the data is processed and transformed in the desired format.

Finally, if all the stages proceed without any error, then the audio file is provided as output the user, in the desired language.

The request session interaction goes on until, the user terminates the session.

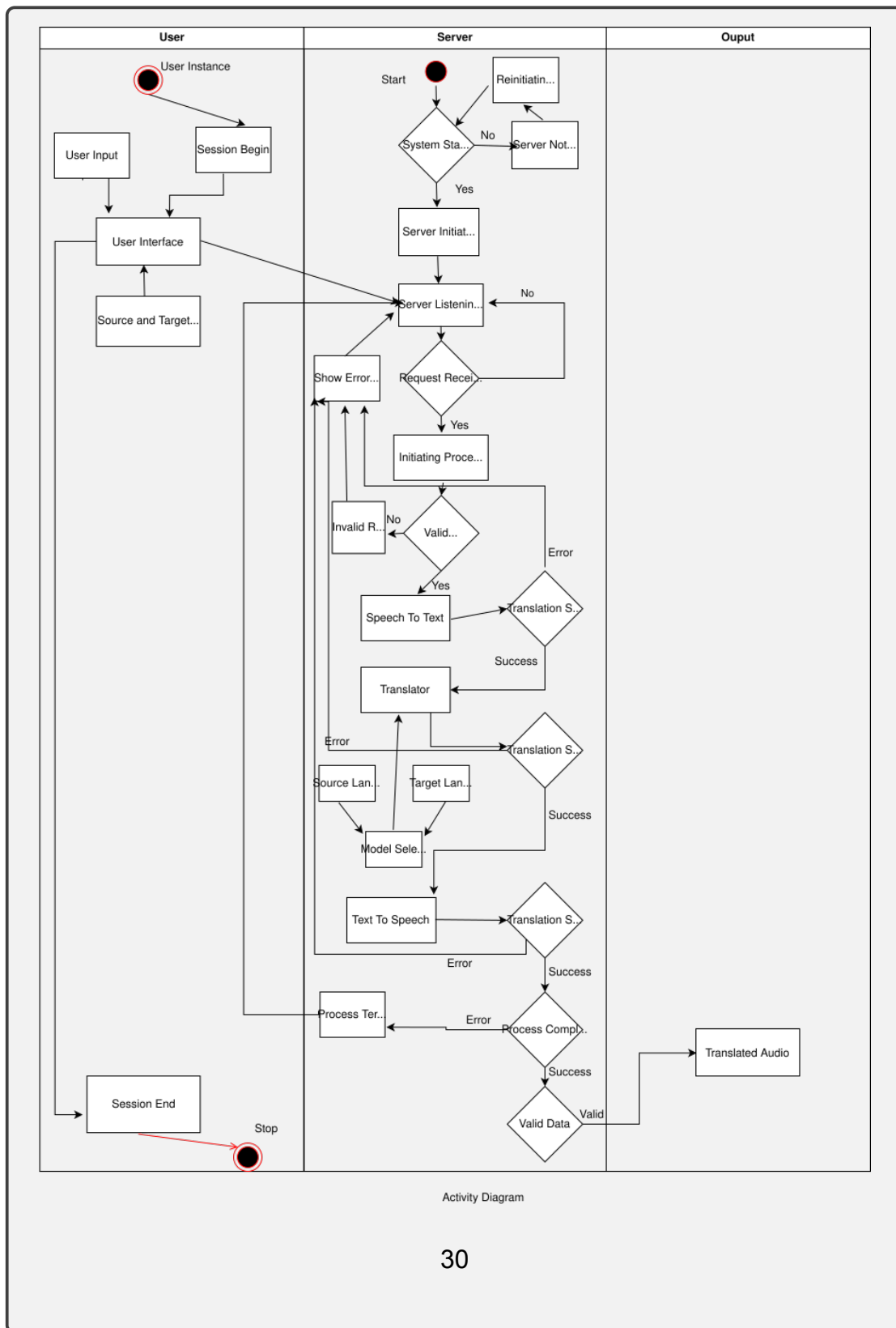


Figure 3.6: Activity Diagram

### 3.3.3 Sequence Diagram

Sequence diagram show the interaction of the user and the server, along with the session duration and all the actions that occur between them.

The session request to the server last from sending the input data to server for processing to the translated output received at the user end.

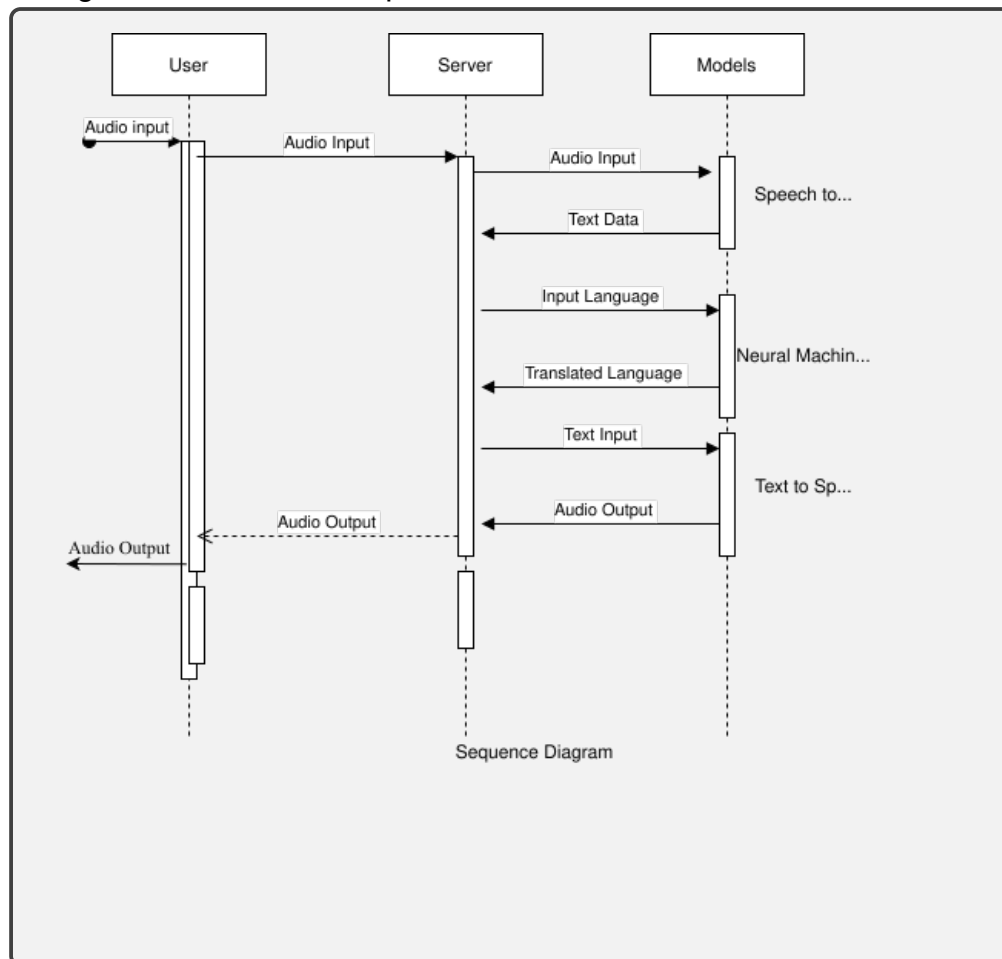


Figure 3.7: Sequence Diagram



### 3.3.4 Class Diagram

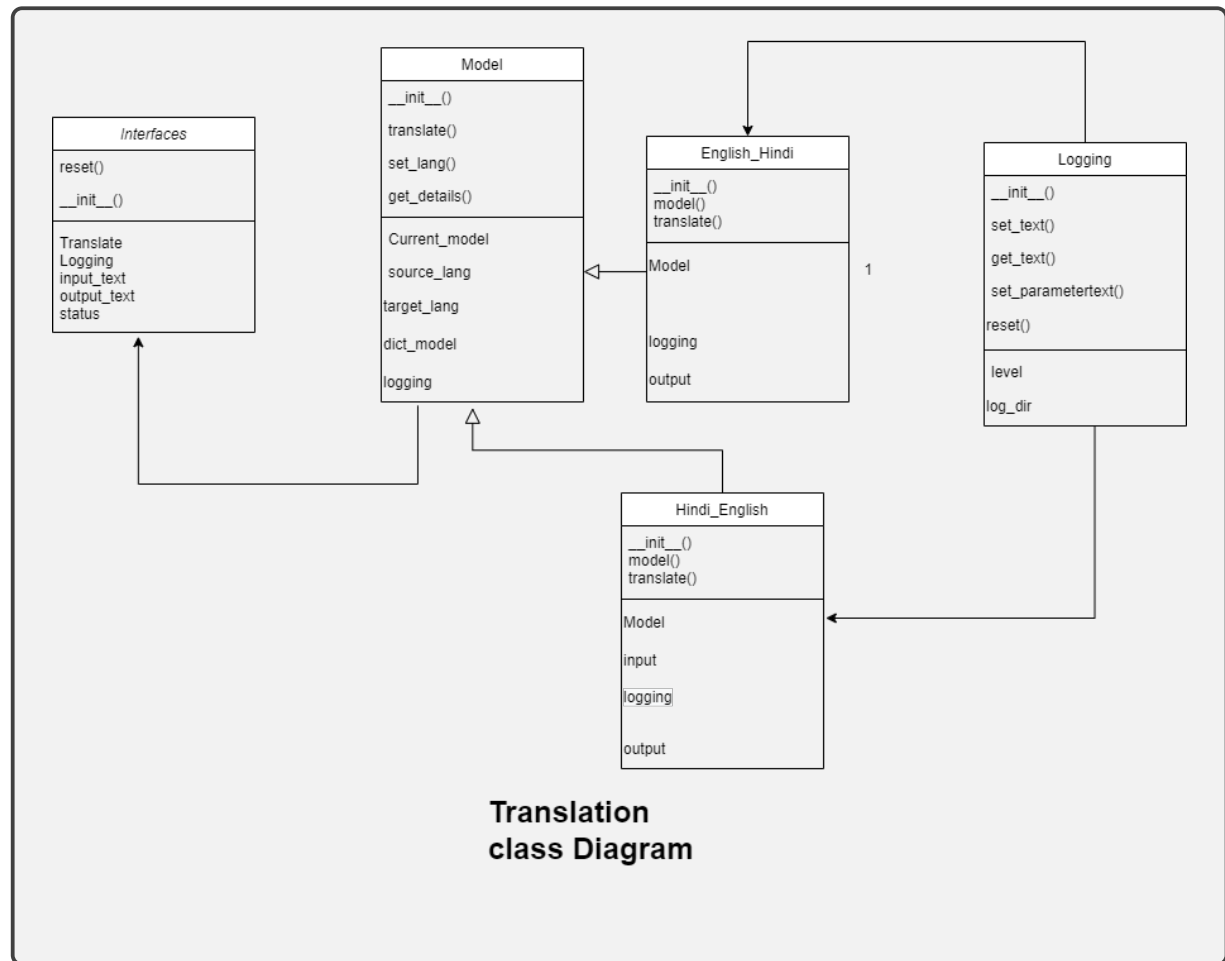


Figure 3.8: Class Diagram

## **Chapter 4**

### **Other Specifications**

### **4.0.1 Advantages**

- Provides an efficient way to translate one language speech to required language speech
- Allows you to communicate with a wide range of employees with different languages for better results.

### **4.0.2 Limitations or Challenges**

- Assigning weight for extracted text
- Latency of the model
- Accuracy of the result in other language

### **4.0.3 Applications**

- During office meetings
- For communicating with different people

# **Chapter 5**

## **Summary**

This report studied papers. The highlights of the papers and observations are reported in chapter 2. The gap has been analyzed, based on which problem statement is designed along with its objectives. The detailed plan of all the activities is mentioned in section 1.5.

This report addresses the problems of companies with employees from different countries for their communication. Given an input, with speech, the goal is to extract text out of it, then translate the text to the desired language and give back the audio file for the required language. The accuracy of the individual components of our design is good, however, with a huge score for improvement.

## **Chapter 6**

## **References**

# Bibliography

- [1] Attention Is All You Need - Ashish Vaswani□ Google Brain  
avaswani@google.com Noam Shazeer□ Google Brain noam@google.com  
Niki Parmar□ Google Research nikip@google.com
- [2] Google's Neural Machine Translation System: Bridging the Gap between  
Human and Machine Translation - Yonghui Wu, Mike Schuster, Zhifeng Chen,  
Quoc V. Le, Mohammad Norouzi  
yonghui,schuster,zhifengc,qvl,mnorouzi@google.com
- [3] H. Sun, R. Wang, K. Chen, M. Utiyama, E. Sumita, and T. Zhao,  
“Unsupervised bilingual word embedding agreement for unsupervised neural  
machine translation,” in Proc. 57th Annu. Meeting Assoc. Comput.  
Linguistics, Florence, Italy, Jul. 2019, pp. 1235–1245. [2] T. Mikolov, Q. V. Le,  
and I. Sutskever, “Exploiting similarities among languages for machine  
translation,” 2013, arXiv:1309.4168

- [4] Haque A U, Mandal P, Meng J, et al. Wind speed forecast model for wind farm based on a hybrid machine learning algorithm[J]. International Journal of Sustainable Energy, 2015, 34(1): 38- 51. [2] Bahar P, Alkhoul T, Peter J T, et al. Empirical investigation of optimization algorithms in neural machine translation[J]. The Prague Bulletin of Mathematical Linguistics, 2017, 108(1): 13- 25