

“V-TRANSLATE”- A Speech to Speech Translation System

Shashank Mahale Student, Department of Computer Engineering
Vidyalankar Institute of
Technology Mumbai, India

Ronit Motiwale Student, Department of Computer Engineering
Vidyalankar Institute of
Technology Mumbai, India

Sami Rumani Student, Department of Computer Engineering
Vidyalankar Institute of
Technology Mumbai, India

Sachin Deshpande
Associate Professor,
Department of
Computer Engineering
Vidyalankar Institute of
Technology
Mumbai, India

Abstract- V-Translate, A Speech to Speech Translation Platform, is a system which will take a voice input of a certain uniform language, say English, and will translate that particular language into another language, say Hindi, and provide a voice output of the same. The input will be stored in the database in the form of audio files. It can be used in future by a user by selecting a particular language out of the provided options. Moreover, our application will recognise the input language on its own given it is consistent. This application has 4 modules. First, to take the Voice input and identify the language. Second, to convert this Audio input to Text format. Third, to translate the Text format of the input to the Text format of the desired language. And Fourth, to convert the translated Text to an Audio format and hence provide the output.

Keywords- Audio; Text; translation; modules; consistent; language.

I. INTRODUCTION

In this day and age, with so much information around us and new concepts popping up every day, there are lot of new things to learn. With around four thousand widely spoken languages, there is a huge possibility that language could become a barrier for people who want to learn and attend intellectual speeches. We, on our part have tried to contribute so that language would not become a barrier for communication. V-Translate mainly implies translating a voice input in a certain language to a particular desired language which can be accessed by an Android device like smartphones. It also stores the voice input in the form of audio files in the server's database. This stored data can then be accessed via a search function and can be heard later on in a desired mentioned language. V-Translate provides the speech to speech language translation facility on multiple devices in multiple languages at a time provided these devices are connected to our cloud via the android application. The user of this application will not have to be physically present for example in the college auditorium to listen to a speech going on in the auditorium. They can access the speech and select the desired language via the application which is connected to the cloud. The speech is stored in the form of text files on the cloud. The aim was to provide a vocal output of the selected file in the desired language to the user and to execute those operations in the minimum time possible.

II. PROBLEM STATEMENT

V-Translator application [1] mainly consists of the use of methods and options that can be used to translate speech from one language into another. Communicating with technological devices via voice has become so popular and natural that we may be just in wondering why the world's richest companies are only bringing these services to us now. The cost of getting the technology to work for new languages (much less all the world's 7,000 languages) is still too high. Extensive speech and translation databases must be collected in order to support a new language; also, vocabularies, language models, and acoustic models must be built, trained, and adapted for each language, domain, and task. Most long-tail languages are not well researched, and for many, no data can be found on the internet. We tried our hand at making a system which can convert speech from one language [2] into another efficiently. It also stores the voice input in the form of audio in the server's database. This stored data is then accessed via a search function and can be heard later on in a desired mentioned language.

III. PROPOSED SYSTEM

We have created a system which solves the problem of not understanding the speaker and missing out on important topics just because of the language barrier. This project focuses on applying the concepts of Artificial Intelligence to translate the speech to a desired language. The user has to use our application on their device for this to happen. It is as good as translating the speech from one language to another at your fingertips. The user won't have to carry out any tedious tasks as we have kept the interface as simple as possible.

We have created tunnels so as to link these different processes and directly provide the user with a final output. In our application, the user only has to get the speech as an input and we will provide them with a translated voice output.

Moreover, we not only provide user with the translation of the speech that they had given as an input but we also store the speech inputs provided by different users in our database and then these speeches can be accessed later on by different user via the application and heard in their desired language. This will help those users who have missed out on a particular speech conducted in say, their local auditorium and want to listen to it in their desired language. By implementing this project, we hope to solve the problems of those people who, say for example, are weak in English and want to listen to certain speeches or sessions in Hindi so as to understand things in a better way.

Our Project consists mainly of 3 important modules:

- *Speech -to-Text Conversion:*
In this module, the speech given by the speaker is recorded and converted into text in the same language that the speaker is speaking in. Our system analyses the person's specific voice and uses it to fine tune the recognition of that person's speech, resulting in more accurate transcription.
- *Text-to-Text Translation:*
In this module, the converted text is translated to the desired language with the use of Google API's. Due to the limited processing and internet speed, the real time conversion becomes less feasible.
- *Text-to-Speech Conversion:*
The translated text is then converted into an audio form so as to provide it as an output to the user.

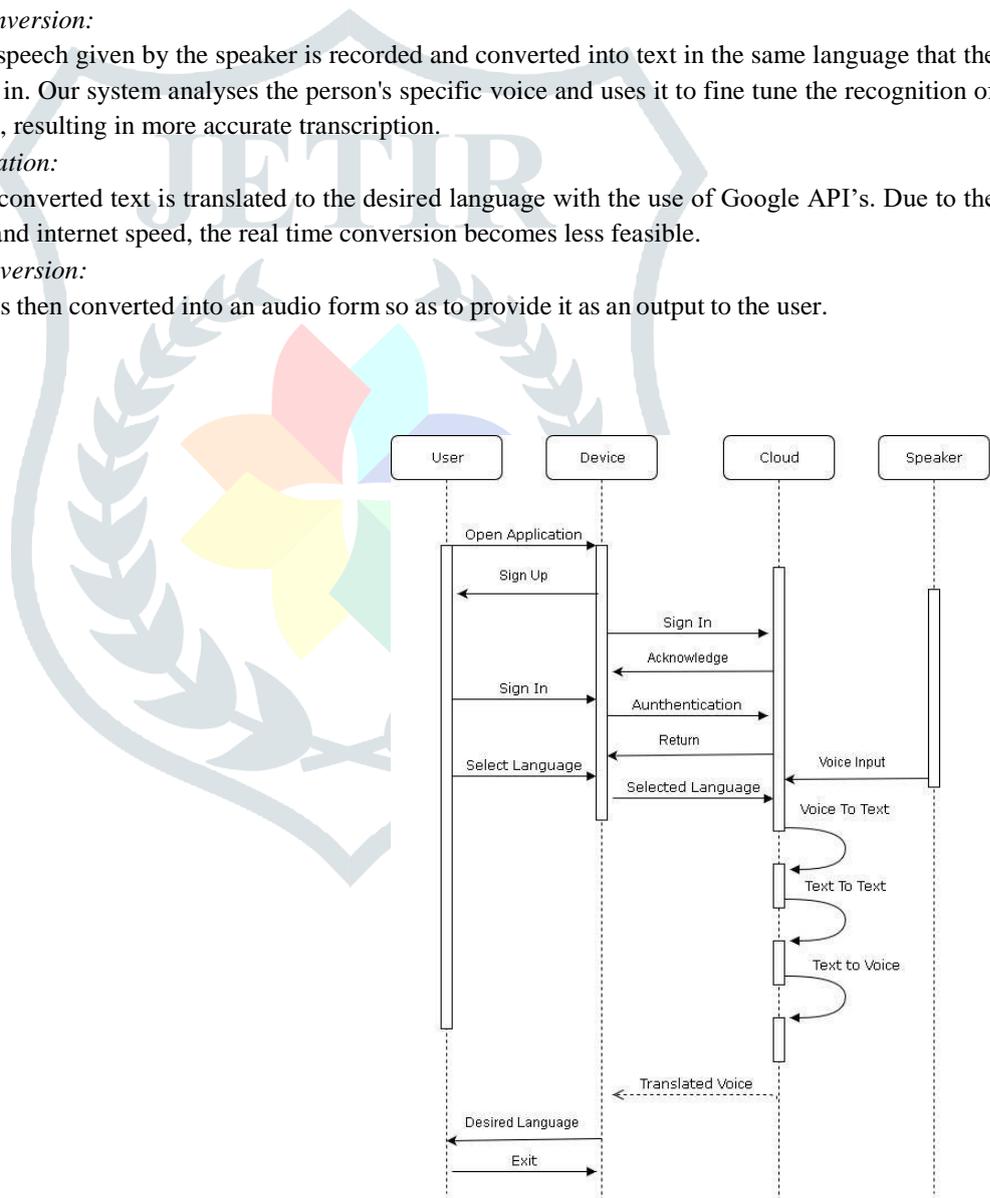


Figure 1:Sequence Diagram of the Working of System

IV. SYSTEM OVERVIEW

The android application is the one that interacts with the user and gets the Audio input. This audio input obtained by the application is divided into blocks of 15 second audios and sent to the cloud. The cloud is where all the processing takes place. It converts the audio format into a text format and translation of this text takes place. The cloud then converts the translated text back into the audio form and, after some involvement with the Google Bucket, it is sent to the user on the application. The speed of the process depends on the strength of the network to which the device is connected.

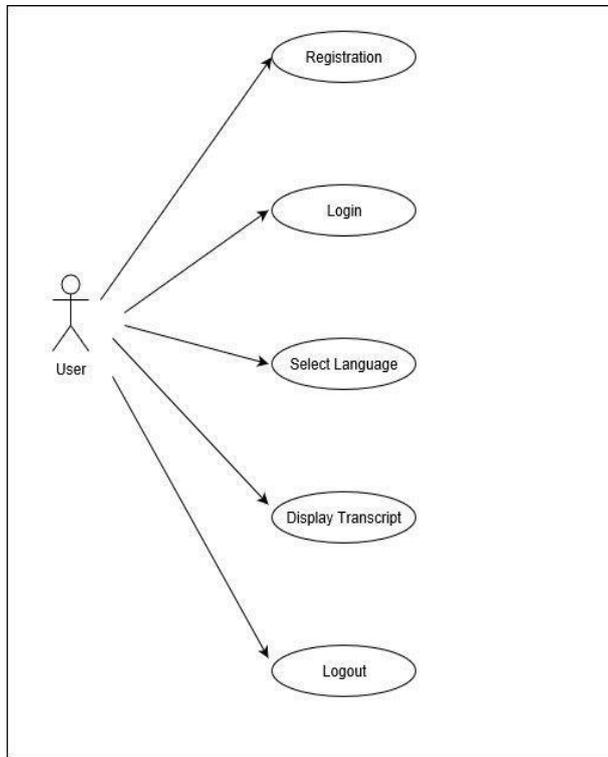


Figure 2: Use Case Diagram of the System

V. IMPLEMENTATION OF PROPOSED SYSTEM

Users will first have to sign up in our application. After the said process is done, their information is stored on the cloud. The authentication, monitoring, etc. is done through the cloud. After logging in, user gets an option of providing the voice input and asked about the language desired for translation.

We make use of API's. API (Application Programming Interface) allows application to communicate with one another. These API's make use of LSTM-RNN algorithm for its functioning. First, audio gets recorded in source language through our app. We record audio such that it forms batches of 15 seconds. These 15 second batches[4] are called blocks. These blocks of audio files will be uploaded to the cloud and stored in the bucket. The blocks are stored in the bucket in the form of Blobs. Blob is a Binary Large Object. We access google cloud from the desktop using Cloud SDK. In virtual instance on Cloud, which we access through SSH keys, we process these audio files by first converting them to text[3] using the Speech to Text API and then translating text[6] into the required language using Text to Text translation API which is then converted into audio using Text to Speech[5] API. The audio and text files of the target language is stored as objects which is accessed by the application to provide to the end user.

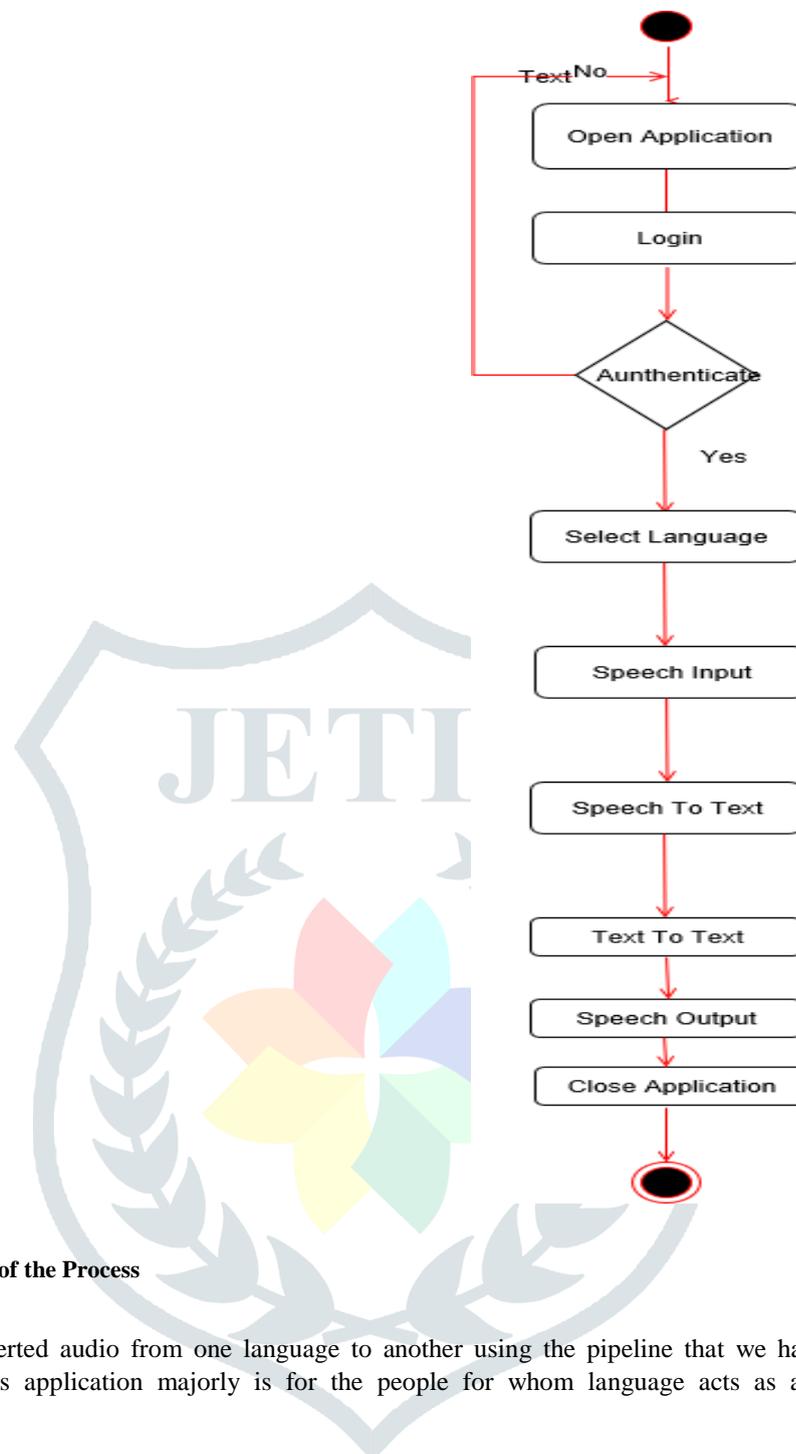
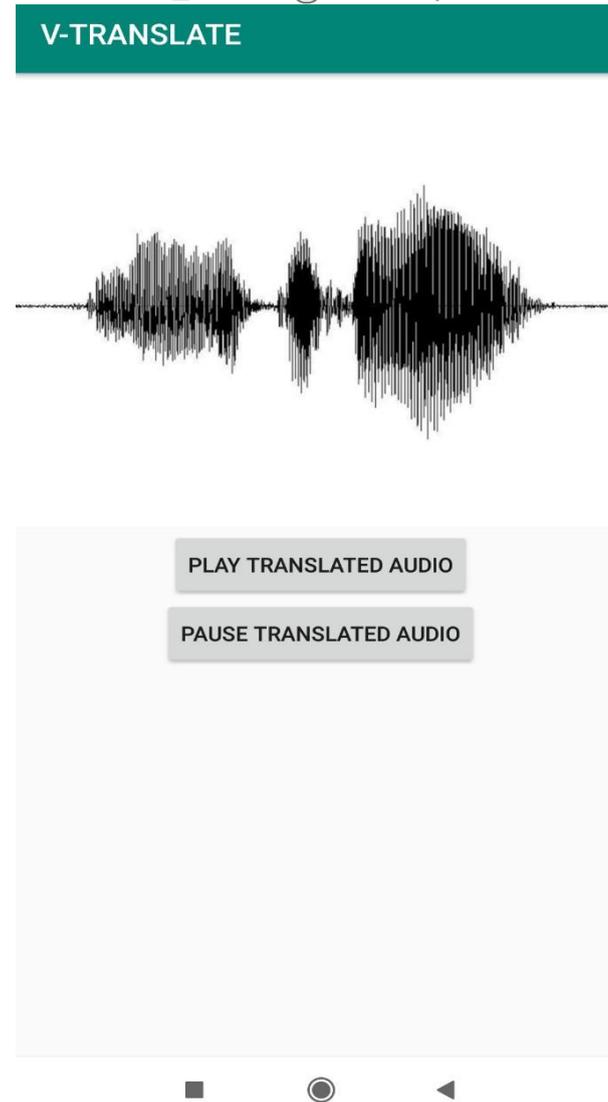


Figure 3: Activity Diagram of the Process

RESULT

We successfully converted audio from one language to another using the pipeline that we have embedded in our Android Application. This application majorly is for the people for whom language acts as a barrier in grasping knowledge.

While executing the given process we found out that we were able to convert speech from one language to another successfully but with a few minor issues. The division of audio input in fifteen second batches lead to a possibility of words that lie on these fifteen second boundaries to break up. As a result, they get lost while the translation takes place. To overcome this problem, we came up with a solution of recording the audio sentence by sentence instead. As a result, it will not break a word into two parts. Noise still remains an issue. An acoustic foam panel was used effectively to reduce the noise as much as possible. Using high quality microphones shall overcome the effects of this issue.



VI. CONCLUSION

In this paper, we have proposed a system, V-Translate, a speech to speech translation system which provides the users the ability to listen to a particular piece of audio in their desired language. The system ensures that it authenticates a user before using the service. It is a free application and is designed in such a way that advancements are still possible to be made to the overall system and it can be upped a notch with proper investment in place. We are working on developing it broadening it further by working on adding some extra features, like sentiment analysis of the converted text so as to study the type of words frequently used, and to translate longer duration speeches.

ACKNOWLEDGEMENT

We are grateful toward all those who helped us throughout the process of writing this paper. Their worthy and insightful inputs and constructive criticism have been of utmost significance.

REFERENCES

- [1] Krasimir Angelov, Bjorn Bringert, Aarne Ranta, Speech-Enabled Hybrid Multilingual Translation for Mobile Devices, Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics, pages 41–44, Gothenburg, Sweden, April 26-30 2014.
- [2] J.Poornakala, A.Maheshwari, Automatic Speech-Speech Translation System from English to Tamil Language, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 5, Issue 3, March2016.
- [3] <https://cloud.google.com/text-to-speech/docs/>
- [4] <https://cloud.google.com/compute/docs/>
- [5] <https://cloud.google.com/speech-to-text/docs/>
- [6] <https://cloud.google.com/translate/docs/>

