International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Speech Recognition and Translation with Real-time Video and Closed Captioning

Prof. P.R.Rodge[1], Sanjyot Dahale[2], Sahil Ahire[3] and Omkar Karnik[4]

[1]Head of Department of Computer Engineering at S.S.J.C.O.E, Dombivli (E), Maharashtra
[2,3,4]Students, Department of Computer Engineering at S.S.J.C.O.E, Dombivli (E), Maharashtra

---

Abstract: Today's technology has brought a drastic amount of advancements in the field of entertainment, sports, medicals facilities etc. However, deaf individuals are often hindered to these facilities. In order to narrow the gap between the hearing impaired and normal humans we present a Speech Recognition application. *The idea is to enable hearing impaired and deaf people to see real-time subtitles while hearing a conversation along with translation to the preferred language of the user.*

Key Words:

AR – Augmented Reality, SDK – Software Development Kit, VR – Virtual Reality, API – Application Program Interface

1. INTRODUCTION

Recognition of the spoken natural language is difficult for the hearing impaired and deaf individuals. One of the solutions found for this is by lip reading speaker, but it often requires good viewing ability of the user. By using todays latest technology, we wanted to create a better world for the hearing impaired folks. Our approach involves using Speech Recognition along with the phone's camera which will allow to see and perceive the visual cues of the speaker at the same time '**Live Subtitles**' will be displayed. The subtitles include the option for multilingual support which will translate into the desired language.

2. System Configuration:

We planned to use Android Studio and Camera Module for the integration of computer vision into the application. This API is based on the Open GL ES3 platform for building Android applications. Furthermore, to convert the audio into speech we used Google Text-to-Speech API, this is done to by integrating JAVA classes into KOTLIN. The Google Text-to-Speech also has various dialects as well as it is very efficient and supports more than 100 different languages. To convert the recorded language into the user required one we used the Google Translate API which is very precise and produces the results efficiently in practical condition.

3. Working

3.1 Google Speech-to-text API

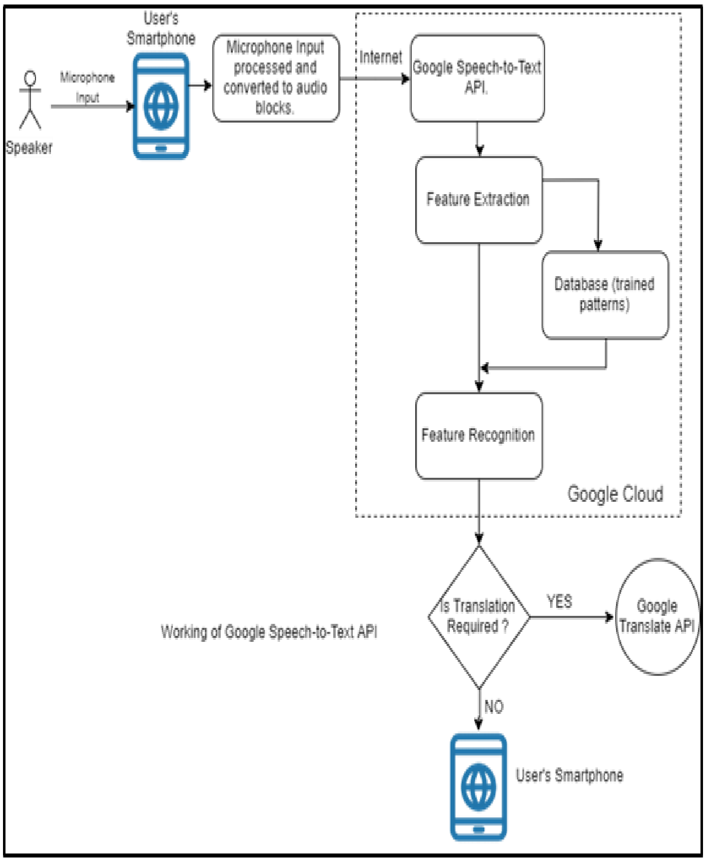
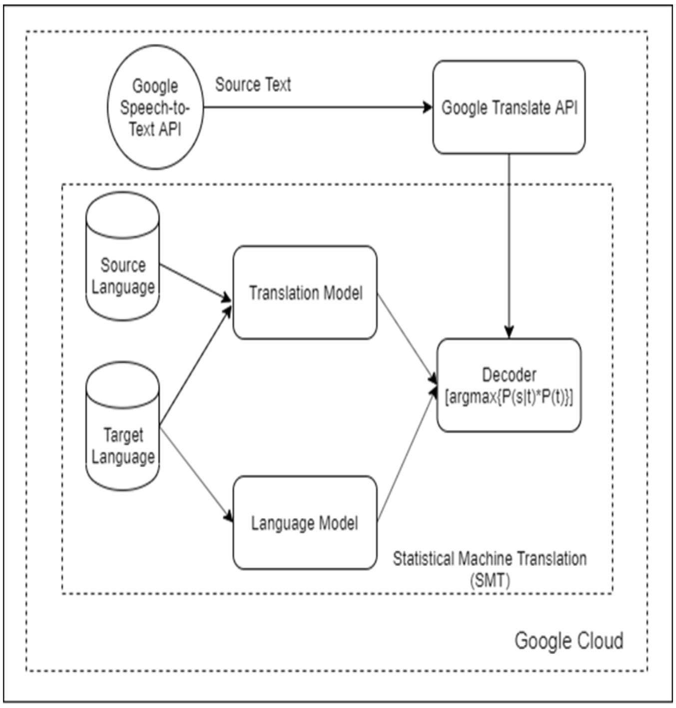


Fig 1: Working of Google Speech-to-text API

As show in the above fig, the microphone records the audio of the speaker and send the processed audio to the Google Cloud Speech-to-Text API. This system is based on Domain Independent Speech Recognition which uses Feature Extraction to extract all the words and patterns from the audio received from the application. These features are used to recognize the phonemes, which are then compared with the Database of trained patterns using Neural Networks. The output of this gives us recognized features which are in text format. This output is then sent back to the application and it is displayed on the smartphone screen as closed captions. If the user chooses to translate the output of the Speech-to-

Text, the output is sent to the Google Translate module for further processing.

3.2 Google Translate API



Working of Google Translate

Fig 2: Working of Google Translate API

As shown in the above fig. the Google Cloud Translate API is based on Statistical Machine Translation. This system uses multiple language databases which contain a corpus of previous approved set of translations. When the user selects the Translation option, he/she also need to provide the language to which the text should be translated. The Translate API uses this language codes to choose the Source language and Target Language Corpuses. The text to be translated is broken down into smaller phrases which are compared with Language model and the Translation model. The Translation model helps the system to translate the source language text into the required language without changing the meaning of the sentence. The language model helps to maintain the grammatical accuracy of the translated sentence. This data is then passed through the Decoder which uses Probability to check the accuracy of the output based on the Translation and Language model outputs by comparing them to the Source Text. If the output has probability higher than the threshold, then it is accepted as the correct output.

4. Evalution

For the actual demonstration, user said a sentence while the App is active. After the App recognizing the audio it displays the live subtitles on the screen. The user translates the text into "Marathi" and "Hindi" language. Initialization of the Application take about 10-15 sec and later the results are displayed with 1-1.5 sec delays. The following Figures 4.1, 4.2 and 4.3 displays the actual demonstration of the Application being used.

Fig 4.1: Recognition of audio and live subtitles being displayed

Fig 4.2: Translation of subtitles into "Hindi" language

Fig 4.3: Translation of subtitles into "Marathi" Language

5. CONCLUSION

This tool enables hearing impaired and deaf people to see real-time "live subtitles'' while hearing a talk on a specific subject, combining them with the real world lecturer's body gestures in the actual environment to provide the user a complete communication experience. Currently, providing support on only two languages this application could be further improved and enhanced by adding more language support, so as to be used in the Global Market.

REFERENCES

[1] Akmeliawati, R., Bailey, D., Bilal, S., Demidenko, S., Gamage, N., Khan, S., ... Sen Gupta, G. (2014). Assistive technology for relieving communication lumber between hearing/speech impaired and hearing people. The Journal of Engineering, 2014(6), 312–323. doi:10.1049/joe.2014.0039

[2] Jimenez, J., Iglesias, A. M., Lopez, J. F., Hernandez, J., & Ruiz, B. (2011). Tablet PC and Head Mounted Display for live closed captioning in education. 2011 IEEE International Conference on Consumer Electronics (ICCE). doi:10.1109/icce.2011.5722919

[3] Shadiev, R., Reynolds, B. L., Huang, Y.-M., Shadiev, N., Wang, W., Laxmisha, R., & Wannapipat, W. (2017). Applying Speech-to-Text Recognition and Computer-Aided Translation for Supporting Multi-lingual Communications in Cross-Cultural Learning Project. 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT). doi:10.1109/icalt.2017.20