

Personalized Speech Translation using Google Speech API and Microsoft Translation API

Sagar Nimbalkar¹, Tekendra Baghele², Shaifullah Quraishi³, Sayali Mahalle⁴, Monali Junghare⁵

¹Student, Dept. of Computer Science and Engineering, Guru Nanak Institute of Technology, Maharashtra, India

²Student, Dept. of Computer Science and Engineering, Guru Nanak Institute of Technology, Maharashtra, India

³Student, Dept. of Computer Science and Engineering, Guru Nanak Institute of Technology, Maharashtra, India

⁴Student, Dept. of Computer Science and Engineering, Guru Nanak Institute of Technology, Maharashtra, India

⁵Student, Dept. of Computer Science and Engineering, Guru Nanak Institute of Technology, Maharashtra, India

Abstract - Speech translation innovation empowers speakers of various language to impart. It subsequently is of enormous estimation of mankind as far as science, cross-cultural exchange and worldwide trade. Henceforth we proposed a that can connect this language hindrance. In this paper we propose a personalized speech translation system using Google Speech API and Microsoft Translation API. Personalized in the sense that user have the choice to select the languages which will be used by the system as a input and a output. The system proposed, can be broken down into three separate segments automatic speech recognition to transcribe the source speech as a text, machine translation to translate the transcribed text into the target language, and text-to-speech synthesis to generate speech in the target language from the translated text. During the execution of the system the speech flawlessly translated in the ideal language and the most intriguing thing is that it is completely done on the cheap by leveraging inexpensive hardware, free translation API s, and some open-source software.

Key Words: Speech translation, google speech API, Microsoft translation API, speech recognition, machine translation.

1. INTRODUCTION

If you've ever tried to communicate with someone who speaks a different foreign language, you know that it can be extremely difficult— even with the assistance of cutting edge translation sites where we need to pay a lot sums of money to fulfill our task. This project will demonstrate to you proper methodologies to transform a \$35 little PC -Raspberry Pi [1] into a component rich language translator which not only just backs up your voice acknowledgment and local speaker playback, but on the other hand is equipped translating your voice into many different language. The incredible part is that it is possible at little to no cost by utilizing reasonable equipment, free translation APIs [2], and some open source programming done on Raspberry Pi which runs on Linux Command line [3]platform. Upon successful completion of this project, mysteriously talk in another language, whatever you say gets translated into another desired language.

Personalized Speech Translator by Using Google speech API and Microsoft translation api is a speech-to-speech translation system that enables communication between people speaking in different languages. A Multilingual speech to speech translation is one of the most serious problems

inherent in globalization.[6] Speech to speech translation systems are often used for applications in a specific situation, such as supporting conversations in non-native languages. The demand for trans-lingual conversations, triggered by IT technologies has boosted research activities on Speech to speech technology. Most of them were consist of three modules namely speech 7recognition, machine translation and text to speech translation. Speech recognition is an interdisciplinary sub-field of computational linguistics that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers. For fulling this requirement in our system we used Google speech API. It can convert the speech into written text which can be further processed for getting the desired output. Apart from being free it only provides 50 requests per day. The second module is Machine Translation. It is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one natural language to another. It can be especially useful for performing tasks that involve understanding and speaking with people who don't speak the same language.[5] For fulling this requirement in our system we used Microsoft Translation API, Google do have it's own translation API but it's not free whereas Microsoft Translation API is free. So this module will convert text written in one language into another language that is the target language. Now the last module is text to speech translation, the output text generated by the second module is converted into speech, this process is simply carried with the help of python language in our system.

2. EXISTING METHODOLOGIES

The Speech translation is one of the process which is being practiced to be carried out through some sort technologies since very long. So there are variety of technologies and systems that offer speech translation. These technologies varies with respect to hardware used, soft-wares used and definitely the methodologies used. The ultimate goal is to produce a system which instantly translate continuous speech. Most of the existing methodologies uses machine translation for speech translation.

Machine translation, sometimes referred to by the abbreviation MT, is a sub-field of computational linguistics that investigates the use of software to translate text or

speech from one language to another. On a basic level, MT performs simple substitution of words in one language for words in another, but that alone usually cannot produce a good translation of a text because recognition of whole phrases and their closest counterparts in the target language is needed. Solving this problem with corpus statistical, and neural techniques is a rapidly growing field that is leading to better translations, handling differences in linguistic typology, translation of idioms, and the isolation of anomalies.

There are many technologies which uses machine translation for speech recognition some of them are -

2.1 MICROSOFT SPEECH API

Microsoft Speech API (SAPI) allows access to Windows' built-in speech recognition and speech synthesis components. The API was released as part of the OS from Windows 98. Text Target language Text to speech (TTS) Voice target language Voice Target language forward. The most recent release, Microsoft Speech API 5.4, supports a small number of languages: American English, British English, Spanish, French, German, simplified Chinese, and traditional Chinese. Because it is a native Windows API, SAPI isn't easy to use unless you're an experienced C++ developer.

2.2 GOOGLE WEB SPEECH API

In early 2013, Google released Chrome version 25, which included support for speech recognition in several different languages via the Web Speech API. This new API is a JavaScript library that lets developers easily integrate sophisticated continuous speech recognition feature such as voice dictation in their Web applications. However, the features built using this technology can only be used in the Chrome browser; other browsers don't support the same JavaScript library.

2.3 JAVA SPEECH API

The Java Speech API (JSAPI) is a specification for cross-platform APIs that supports command-and-control recognizers, dictation systems, and w w w . c o m p u t e r . o r g / s o f t w a r e | speech synthesizers. Currently, the Java Speech API includes javadoc-style API documentation for the approximately 70 classes in and interfaces to the API. The specification includes a detailed programmer's guide that explains both introductory and advanced speech application programming with JSAPI, but it doesn't yet offer the source code or binary classes required to compile the applications.

3. SYSTEM HARDWARE DESIGN

The hardware consists of the following parts: Raspberry pi 3 [model B] mounted with 16 GB SD card, USB Headset, Internet connection via Ethernet or Wi-Fi, laptop. The Fig 1 gives the block diagram of system hardware design.

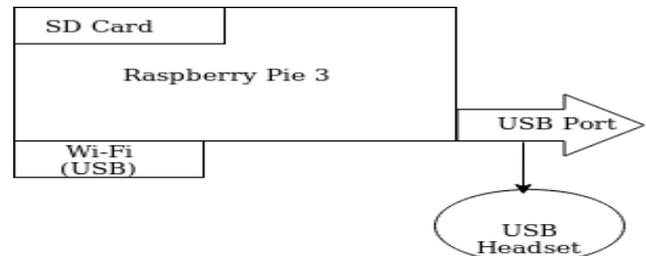


Fig -1: System Hardware Design

3.1 RASPBERRY PIE 3

Raspberry Pi is a low cost, credit-card sized computer that plugs into a computer monitor or TV, and uses a standard keyboard and mouse. It is a capable little device that enables people of all ages to explore computing, and to learn how to program in languages like Scratch and Python. It's capable of doing everything you'd expect a desktop computer to do, from browsing the internet and playing high-definition video, to making spreadsheets, word-processing, and playing games. It has various models, we are specifically using raspberry pie model 3b for our project. Why we are using raspberry pie is because of it's size since a speech translator should be compact in size, handy, and the one which runs on low power or batteries, it satisfy all this requirements.

For our system an 16 GB SD card is flashed with NOOBS OS and is put into the slot. A Wi-Fi adapter is put in USB slot for network connection. Power supply is enabled and raspberry pi is connected to the laptop via a USB cable. SSH client putty is used to work with raspberry pi via command line interface.



Fig -2: Raspberry Pie 3

3.2 USB HEADSET

A USB headset is a convenient way to communicate with other people through the computer. The headset consists of headphones and an attached microphone that connects to your computer or laptop through a USB port. Various

programs will sync up with the USB headset and allow you to speak to your friends, family, or co-workers while you perform other tasks on your computer. We require this since we are using a raspberry pie 3, in this 3.5mm jack doesn't support a microphone and only can be used as a earpiece. But we have to use this project by giving speech as a input and receiving speech as a output. Hence we used a USB headset in



our project.

Fig -3: USB Headset

3.3 WIRELESS ADAPTER

A wireless adapter is a hardware device that is generally attached to a computer or other workstation device to allow it to connect to a wireless system. Before the advent of consumer devices with built-in Wi-Fi connectivity, devices required the use of wireless adapters to connect to a network. Wireless adapters are also known as WiFi adapters. We can also use Ethernet in place of Wi-Fi adapter in our system.



Fig -4: Wireless Adapter

4. SYSTEM SOFTWARE DESIGN

Personalized Speech Translator by Using Google speech API and Microsoft translation API is a speech-to-speech translation system that enables communication between people speaking in different languages. A Multilingual speech to speech translation is one of the most serious problems inherent in globalization.[6] Speech to speech translation systems are often used for applications in a specific situation, such as supporting conversations in non-native languages. The demand for trans-lingual conversations, triggered by IT technologies has boosted research activities on Speech to speech technology.

Most of them were consist of three modules namely speech recognition, machine translation and text to speech translation.

4.1 SPEECH RECOGNITION

Speech recognition is an interdisciplinary subfield of computational linguistics that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers. For fulling this requirement in our system we used Google speech API. It can convert the speech into written text which can be further processed for getting the desired output. Apart from being free it only provides 50 requests per day.

4.2 MACHINE TRANSLATION

The second module is Machine Translation is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one natural language to another. It can be especially useful for performing tasks that involve understanding and speaking with people who don't speak the same language.[5] For fulling this requirement in our system we used Microsoft Translation API, Google do have it's own translation API but it's not free whereas Microsoft Translation API is free. So this module will convert text written in one language into another language that is the target language.

4.3 TEXT-TO-SPEECH TRANSLATION

Now the last module is text to speech translation, the output text generated by the second module is converted into speech, this process is simply carried with the help of python language in our system.

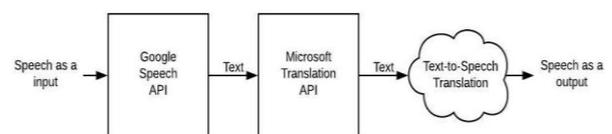


Fig -4: System Software design

4.4 API s USED

API s that we have used in our system-

4.4.1 GOOGLE SPEECH API

Google speech API is speech recognition API which converts speech to text format which is used to translate through Microsoft translation API. The Google Speech-To-Text API isn't free, however. It is free for speech recognition for audio less than 60 minutes. For audio transcriptions longer than that, it costs \$0.006 per 15 seconds. While free, Google's speech API only allows 50 requests per day.

that translates text into the desired language. Despite this approach being more efficient and more focused as compared to other available technologies more research is needed to develop a system like one if you stick in your ear, you can instantly understand anything said to you in any form of language." We aren't there yet, but definitely will be reaching there soon.

There are so much options available for solving the speech translation problems. These options are available a wide ranges varying in technology, hardware architecture, softwares, compatibility and lot more. But there is no such system available which is handy or just like a earpiece which when plugged in your ear, you can instantly understand anything said to you in any form of language." We aren't there yet, but the massive number of ongoing researches will definitely lead us their to achieve such an amazing an ideal speech translation device.

ACKNOWLEDGEMENT

We would like to thank our project guide Prof. Neerajan Chitare, who has been an inspiration. He always has motivated us and helped us in understanding various concepts. We are also grateful to the HOD of Computer Science and Engineering department of our college Guru Nanak Institute of Technology, who taught us the basics of working with raspberry pi.

REFERENCES

- [1] Rithika.H 1 , B. Nithya santhoshi, "Image Text To Speech Conversion In The Desired Language By Translating With Raspberry Pi" in 2016 IEEE International Conference on Computational Intelligence and Computing Research.
- [2] Liang Gu, Member, IEEE, Yuqing Gao, Fu-Hua Liu, and Michael Picheny, Fellow, IEEE, "Concept-Based Speech-to-Speech Translation Using Maximum Entropy Models for Statistical Natural Concept Generation" in IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 2, MARCH 2006.
- [3] Asim Smailagic, Dan Siewiorek, Richard Martin, Denis Reilly, "CMU Wearable Computers for Real - Time Speech Translation".
- [4] Barif Nursetyo, De Rosal Ignatius Moses Setiadi,"LatAksLate: Javanese Script Translator based on Indonesian Speech Recognition using Sphinx-4 and Google API" in 2018 International Seminar on Research of Information Technology and Intelligent Systems ISRITI.
- [5] Tiago Duarte, Rafael Prikladnicki, Fabio Calefato, and Filippo Lanubile, "Speech Recognition for Voice-Based Machine Translation" published by the IEEE society in 2014.
- [6] M.D. Faizullah Ansari 2, R.S. Shaji 1 , T.J.SivaKarthick, S.Vivek , A.Aravind, "Multilingual Speech to Speech Translation System in Bluetooth Environment" in 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCCCT).

- [7] Sangeetha Rajesh, Lifna C.S, "UI Design for Language Translator Module in Swasthya Slate (m-Health Tool)" in 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI).
- [8] Nam-Yong Han, Un-Cheon Choi and Youngjik Lee, "ASangeetha Rajesh, Lifna C.S, "UI Design for Language Translator Module in Swasthya Slate (m-Health Tool)" in 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI).
- [9] Ravinder Kumar, Prem Kumar Ch, "Universal Language Translator Using Raspberry Pi" in International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 3, Issue 1, January 2016.

BIOGRAPHIES



I am Sagar Nimbalkar. About me, logical problem solving has always appealed to me and this explains my interests in mathematics, programming and computing in general. The decision to read computer science was therefore a simple one.



I am Tekendra Baghele, currently pursuing engineering in computer science at Guru Nanak Institute of Technology



I am Shaifullah Quraishi, currently pursuing engineering in computer science at Guru Nanak Institute of Technology



I am Sayali Mahalle, currently pursuing engineering in computer science at Guru Nanak Institute of Technology



I am Monali Junghare, currently pursuing engineering in computer science at Guru Nanak Institute of Technology