



ISBN	978-81-929866-5-4
Website	icca.co.in
Received	14 – March– 2016
Article ID	ICCA018

VOL	05
eMail	icca@asdf.res.in
Accepted	02 - April – 2016
eAID	ICCA.2016.018

## An Application for Performing Real Time Speech Translation in Mobile Environment

Buela Kutti<sup>1</sup>, S Hari<sup>2</sup>, G Anbuselvi<sup>3</sup>

<sup>1,3</sup>Asst. Professor, CSE, <sup>2</sup>Meenakshi College of Engineering, Chennai

**Abstract:** This paper presents the method of applying speaker-independent and bidirectional speech-to-speech translation system for spontaneous dialogs in real time calling system. This technique recognizes spoken input, analyzes and translates it, and finally utters the translation. The major part of Speech translation comes under Natural language processing. Natural language processing is a branch of Artificial Intelligence that deals with analyzing, understanding and generating the languages that humans use naturally in order to interface with computers in both written and spoken contexts using natural human languages instead of computer languages. Speech Translation involves techniques to translate the spoken sentences from one language to another. The major part of speech translation involves Speech Recognition which is the translation of spoken speech to text and identifying the context and linguistic structure of the input speech. In the current scenario, the machine does not identify whether the given word is in past tense or present tense. By using the algorithm, we search for a word to check if it is past or present by searching for the sub strings, as “ed”, “had”, “Done”, etc., This paper gives us an idea on working with API's to translate the input speech to the required output speech and thus increasing the efficiency of Speech Translation in cellular devices and also a mobile application that will help us to monitor all the audios present in mobile device and translate it into required language.

**Index Terms** — Speech translation, Speech Recognition, Natural Language Processing, Automatic Speech Recognition, Text-To-Text Translation, Voice Synthesis.

### 1. INTRODUCTION

A single language is not spoken by everyone in this multilingual world. There are from 6800 to 6900 distinct languages in the modern world. An individual may hardly know 4-5 languages to converse with people around him. Therefore people face a huge language barrier in their day to day life. This paper provides an idea to eliminate the language barriers among people speaking different languages and helps them to communicate in their own language. Consider a person from Chennai who knows only Tamil and wants to communicate to a government employee in New Delhi who knows only Hindi and English. Here the communication is not possible until one knows the language of the other. One cannot be a master of all languages and at such situations Speech Translation comes into act as a helping hand to resolve the communication problem.

The users can simply pick up a smart mobile phone and use voice dialing and speech commands in order to initiate a dialog translation session. It emphasizes the robust processing of spontaneous dialogues posing difficult challenges to human language technology. It deals with spontaneous one-way speech. The system is mainly intended to translate between Indian languages such as Bengali, Gujarati, Hindi, Kannada, Malayalam, Punjabi, Tamil, Telugu, and Urdu.

The main idea was derived from the proceedings of the Lok Sabha and the Rajya Sabha where Members of the Parliament from different states, speaking different languages present their opinions and proposed works in their mother tongue which is automatically

This paper is prepared exclusively for International Conference on Computer Applications 2016 [ICCA 2016] which is published by ASDF International, Registered in London, United Kingdom under the directions of the Editor-in-Chief Dr Gunasekaran Gunasamy and Editors Dr. Daniel James, Dr. Kokula Krishna Hari Kunasekaran and Dr. Saikishore Elangovan. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honoured. For all other uses, contact the owner/author(s). Copyright Holder can be reached at copy@asdf.international for distribution.

2016 © Reserved by Association of Scientists, Developers and Faculties [www.ASDF.international]

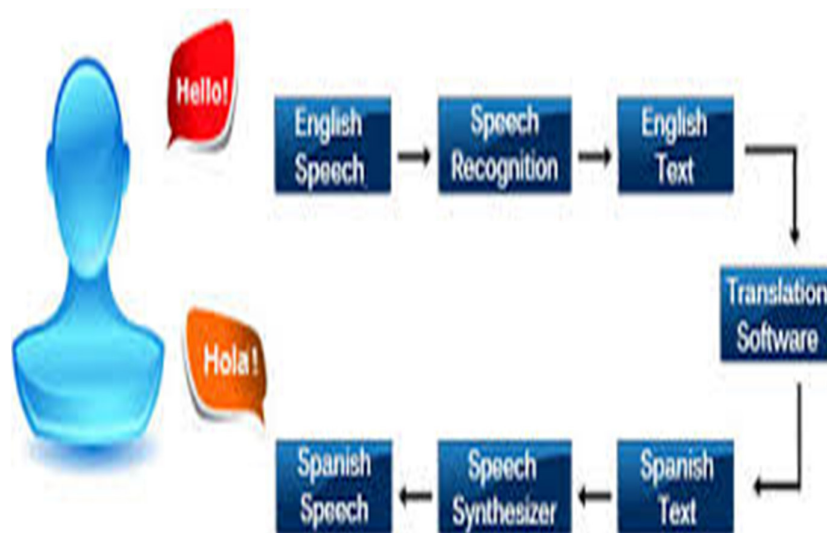
**Cite this article as:** Buela Kutti, S Hari, G Anbuselvi. “An Application for Performing Real Time Speech Translation in Mobile Environment”. *International Conference on Computer Applications 2016*: 89-93. Print.

translated to every member in their own choice of language in real time.

## 2. Proposed Work

Speech Translation is a process by which conversational spoken sentences are instantly translated and spoken aloud in a second language. This differs from phrase translation, which is where the system only translates a fixed and finite set of phrases that have been manually entered into the system. Speech translation technology enables speakers of different languages to communicate.

Speech Translation system would typically integrate the following three procedures. Such as: Automatic Speech Recognition (ASR), Text-to-Text Translation (TTT) module and Voice Synthesis(TTS) module.



1. Speech Translation

The speaker of language A speaks into a microphone and the speech recognition module recognizes the utterance. The Text-to-Text translation module then translates this string to the desired language B. Instead of translating the input speech, word by word, this method translates the entire speech by identifying the “Parts of Speech” and translating them into the text format of the output language. The text instances are then processed in speech synthesis module, which identifies the text input and produces an equivalent speech of the same.

### 2.1 Automatic Speech Recognition (ASR) Module

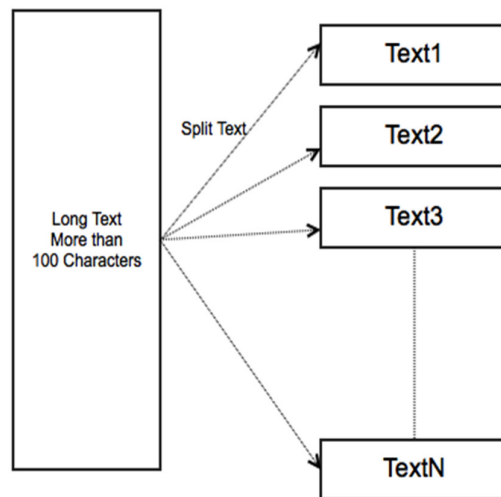
The speech-to-text part of the input language involves implementation of Speech recognition for which a java script is written with the inbuilt function audio recorder and the recognized speech is passed to a PHP file where the “Google Speech To Text API” is called by accepting the POST requests with voice file encoded in FLAC format which is forwarded by the java script, and query parameters for control. The FLAC (Free Loss-less Audio Compression) format is an audio coding format for lossless compression of digital audio, and is also the name of the reference codec implementation. Digital audio compressed by FLAC's algorithm can typically be reduced to 50–60% of its original size and decompressed to an identical copy of the original audio data.

The Request URL should look like this: <https://www.google.com/speech-api/v1/recognize>.

The ASR module consists of the following Query parameters:

- **Client:** The client's name you're connecting from. For spoofing purposes, let's use chromium
- **Lang :** Speech language, for example, ar-QA for Qatari Arabic, or en-US for U.S. English
- **Maxresults:** Maximum results to return for utterance
- **Post:** body should contain FLAC formatted voice binary.

**Cite this article as:** Buela Kutti, S Hari, G Anbuselvi. “An Application for Performing Real Time Speech Translation in Mobile Environment”. *International Conference on Computer Applications 2016*: 89-93. Print.



## 2. Speech to Text conversion

And the final output will be a text equivalent to the input speech which is forwarded to the next module Text-To-Text Translation (TTT).

### 2.2 Text-to-Text Transition (TTT) Module

The output text from the previous module is the input for this TTT module which is present in the PHP file following the previous module and it is called automatically. This input text is then converted into equivalent text of the desired speech language.

This module also uses an API called "BING TRANSLATOR" developed by MICROSOFT. Recently Microsoft announced the latest version of Bing Translator, this new version (v2) added some cool and new features, including collaborative translations, customizable widgets, powerful API, and Translate-to-Speak. In this post I will describe the simplest way to use those new APIs, provide simple example and demonstrate new features.

Bing translator APIs could be easily targeted through the various available APIs: AJAX, HTTP and SOAP. But first, you should obtain a valid Bing AppID. Sign-in using your live ID then get it from here. The AppID will be used as a validation parameter when calling any API such as Detect, Translate, and Speak.

The Translate method requires the following parameters:

**AppID:** Which is a string of a valid Bing AppID?

**From:** A code represents the language of the translated text. (You could get the available translate language using Get Languages Names Method)

**To:** Another code that represents the language to translate text into.

**Text:** And of course the text that's to be translated.

**On Complete:** The Call back function that will be called on the completion of the request.

The Request URL should look like this:

```
http://api.microsofttranslator.com/V2/Ajax.svc/Translate?appId=MyAppID&from=en&to=ar&text=hello&oncomplete=doneCall
back
```

All you have to do is just replace MyAppID with your own valid Bing AppID and define the doneCallback function. The following is a full simple example that calls the Translate method passing the parameters described above. Notice that the doneCallback function receives a response parameter and simply displays it inside a div.

The output of this module will be a text, equivalent to the input text in the desired output speech language.

**Cite this article as:** Buela Kutti, S Hari, G Anbuselvi. "An Application for Performing Real Time Speech Translation in Mobile Environment". *International Conference on Computer Applications 2016*: 89-93. Print.

Now this output text is forwarded to the Voice Synthesis (TTS) module to obtain the equivalent speech of this text in desired language.

<p>Original text:</p> <div style="border: 1px solid gray; padding: 5px; width: fit-content;"> <p>भारत में महँगाई की दर 42 महीने के उच्चतम स्तर पर पहुँच गई है। 19 अप्रैल को खत्म हफ़्ते में मुद्रास्फीति की दर 7.57 फ़ीसदी आंकी गई.</p> </div> <p>Hindi to English ▾ Translate</p>	<p><u>Automatically translated text:</u></p> <p>India's rate of dearness 42 months at the highest level of access to one. Frying on April 19 weeks, the inflation rate was 7.57 percentage assessed.</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### 3. Text-To-Text translation

## 2.3 Voice Synthesis (TTS) Module

The output text of the previous TTT module is now the input for this TTS module which is present in that PHP file following the previous module and it is called automatically where this module is used to process the input text into the equivalent speech of the desired language.

The Request URL should look like this: <http://tts-api.com>



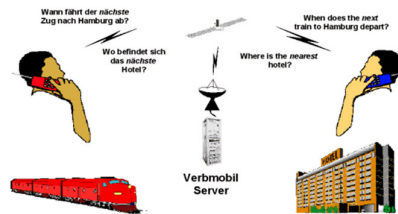
### 4. Text-To-Speech Synthesis

The final output of this module is nothing but the equivalent speech of the original input speech of different language and this speech is passed back to the java script written at first which automatically outputs the speech in the desired language.

## 3. Assumptions

The implementation of speech translation over cellular devices will include a talk with the top telecommunication companies such as Airtel, Vodafone, and Idea. The implementation will also include a privacy closure which would include protecting the privacy of users by not providing the recorded voice during speech translation to any unauthorized person. Speech translation systems should translate an input language to a desired output language only after identifying the completion of a sentence else the conversion would make no sense. At the same time a language conversion should also understand the noun, verb, adjective and translate the same accordingly.

**Cite this article as:** Buela Kutti, S Hari, G Anbuselvi. "An Application for Performing Real Time Speech Translation in Mobile Environment". *International Conference on Computer Applications 2016*: 89-93. Print.



#### 5. Working Model of Speech translation System

Speech translation can be initiated using voice dialing and speech commands. The speech API's can be implemented using an application over the phone or as an integrated service from the telephone providers. The above discussed method is only applicable for one-way communication which allows only one person to speak at a time. The usage of such translation techniques should overcome the distortion caused by noises during the conversation and provide an error free translation.

#### 4. New Working Implementation

In the current scenario, the machine does not identify whether the given word is in past tense or present tense. By using the algorithm, we search for a word to check if it is past or present by searching for the substrings, as “ed”, “had”, “Done”, etc., In the same manner, we search for a number to determine if a given specified number is a key or a date or a normal number. If the strings before the point of interest consist of dates, then we read the number as ‘date’, if the number has no reference to the speech then it is considered as a ‘normal number’. If authentications are available in the part of speech, then the number is evaluated as a ‘key’.

By increasing the pitch of the voice we can rectify the clock errors and incorrect recognitions.

#### 5. Conclusion & Discussion

Such speech translation techniques can help the people of different languages to communicate with each other without any language barriers. By doing so it brings the people of various religions, languages together. The telecommunication companies can also find profit in this technique by providing this service at a nominal rate along with the normal call rates. There are speech translation techniques such as Skype Translate, Verbmobil which do provide language translation for a few languages but neither of them provides translation for Indian languages. Thus developing a technology for speech translation in Indian languages will not only be a great invention in the field of natural language processing but also will provide as a great tool of communication for the people of India.

#### 6. References

1. “Multilingual speech-to speech translation system for mobile consumer devices” by Seung Yun; Young-Jik Lee; Sang-Hun Kim.
2. “An analysis of machine translation and speech synthesis in speech-to-speech translation system” by Hashimoto, K.; Yamagishi, J.; Byrne, W.; King, S.; Tokuda, K.
3. “Development and application of multilingual speech translation” by Nakamura.S.
4. Mobile Speech-to-Speech Translation of Spontaneous Dialogs: An Overview of the Final Verbmobil System .Wolfgang Wahlster
5. Jacob Benesty, M. Mohan Sondhi, and Yiteng Huang, Handbook of Speech Processing, Springer, 2008. Wiqas Ghai and Navdeep Singh, “Literature Review on Automatic Speech Recognition”, International Journal of Computer Applications vol. 41– no.8, pp. 42-50, March 2012.
6. R K Aggarwal and M. Dave, “Markov Modeling in Hindi Speech Recognition System: A Review”, CSI Journal of Computing, vol. 1, no.1, pp. 38-47, 2012.
7. Dev, A. (2009) ‘Effect of retroflex sounds on the recognition of hindi voiced and unvoiced stops’, Journal.
8. Rajesh Kumar Aggarwal and M. Dave, “Acoustic modeling problem for automatic speech recognition system: advances and refinements Part (Part II)”, Int J Speech Technol, pp. 309–320, 2011.
9. Kuldeep Kumar, Ankita Jain and R.K. Aggarwal, “A Hindi speech recognition system for connected words using HTK”, Int. J. Computational Systems Engineering, vol. 1, no. 1, pp.25-32, 2012.
10. Kuldeep Kumar R. K. Aggarwal, “Hindi speech recognition system using HTK”, International Journal of Computing and Business Research, vol. 2, issue 2, May 2011.
11. R.K. Aggarwal and M. Dave, “Performance evaluation of sequentially combined heterogeneous feature streams for Hindi speech recognition system”, 01 September 2011.