International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 08 | Aug 2019 www.irjet.net p-ISSN: 2395-0072

SPEECH TRANSLATION SYSTEM FOR LANGUAGE BARRIER REDUCTION

Swetha V Patil[1], Pradeep N[2]

[1]Post graduate in Computer Science and Engineering, BIET College, Davangere-577004, India.
[2]Associate Professor, Department of Computer Science and Engineering, BIET College, Davangere-577004, India.

Abstract - *Speech is the most natural form of human communication. Communication is needed to delivery information. The language which is used as a communication to interact between others. But not everyone has the same language to communicate; there are many languages from different countries of the world. Speech-to-speech translation system able to converts source language into the target language.*

In this paper, an attempt is made to develop a real time speech-to-speech translation system to overcome the language barrier. The speech-to-speech translation system which recognizes speech in one form of language and translates it into target language. The translation is done only for the local languages they are Kannada, English, Hindi and Telugu.

Key Words: *Input speech, Speech Recognition, Speech Translation, Translator.*

1. INTRODUCTION

Speech-to-speech translation is a technology that automatically translates one language into another in order to enable communication between two persons whose mother tongues are different. In other words, speech translation is one of the process in which spoken words or sentences are immediately translated and spoken aloud in second language. From speech-to-speech technology it is possible to build the system that translates continuously spoken words from one language to another. Speech translation technology is selected as one of the ten technologies that will change the world. Recent technological advances and breakthroughs leading to the constant increase in computational power with an unprecedented tendency of building smaller and smaller devices that yield higher and higher performance, has inevitably lead to a strong demand for information retrieving [1] and communication enabling technologies that are multilingual aware. Speech-to-speech (S2S) translation is a complex process designed to assist communication between individuals that speak different languages.

Speech systems find wide range applications in education, medicine, military, marketing and cross-border operations etc. Language of speech is an important factor to be dealt with in order to complete an effective communication link. Over the past few decades, the need to overcome this language barrier between people belonging to different linguistic backgrounds has been an area of interest in research. Language translation systems have served as a major breakthrough for this issue. [2]

Language difference has been a serious barrier for globalization, international travel, and language divides. Breaking the language barrier will bring larger markets for international business, international tourism, and mobile speech translator may contribute to equalize communication ability. Our goal is to realize an automatic speech-to-speech translation system for four language pairs. [3]

1.1 COMPONENTS OF SPEECH-TO-SPEECH TRANSLATION SYSTEM

a) Speech recognition technology which recognizes the user's speech input and converts into source language text or the technology to recognize speech.
b) Machine translation which translates source language text into the target language text or the technology to translate the recognized words.
c) Speech synthesis or Text-to-speech synthesis which converts translated text into speech or the technology to synthesize speech in the other person's language.
In addition, the technology understands natural language and UI-related technology also plays an important role in this speech-to-speech translation system. [4]

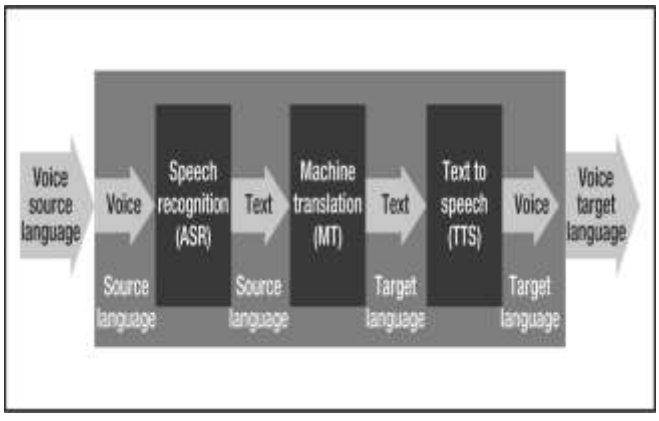

Fig 1: Components of speech-to-speech translation system

1.2 APPLICATIONS OF SPEECH-TO-SPEECH TRANSLATION

As machine translation applications are reaching significantly high accuracy levels, they are being increasingly employed in more areas of business, introducing new applications and improved machine-learning models. [5]

a) Machine Translation in Industry for Business Use

Although big players like Google Translate and Microsoft Translator offer near-accurate, real-time translations, some "domains" or industries call for highly-specific training data related to the particular domain in order to improve accuracy and relevancy. Here, generic translators would not be of much help as their machine-learning models are trained on generic data. These applications are used by small, medium and large enterprises. Some organizations offer multi-domain translation services–that is, customizable solutions across multiple domains–and other organizations offer translation solutions only for a specific domain. These solutions, although automated for the most part, still depend on human translators for pre- and post-editing processes.

Some fields that warrant domain-specific machine translation solutions are:

- Government
- Software & technology
- Military & defense
- Healthcare
- Finance
- Legal
- E-discovery
- Ecommerce

b) Benefits of translation technology for healthcare [6]

- Reduced costs

Healthcare translation technology can significantly reduce costs for hospitals and providers in their interpretation needs, while also boosting productivity. "This sort of technology is a low-hanging fruit CFOs and senior admins hardly recognize," "That cost saving can be leveraged to be used for more critical, clinical applications that are much more sensitive to cost cutting."

- Reduced administrative and staff burden

Some healthcare organizations have a limited number of people available for interpretation, especially on an immediate basis. "The wait time for interpreters can sometimes be upwards to 25 minutes," Speech-to-speech technology can solve this the wait time problem, which places less of a burden on staff while increasing patient throughput. "That level of healthcare productivity is really a goal for most hospital administrations. It's something that's highly sought after: reduce costs of translation and save money, that hits a benefit twice."

- Increased quality and accuracy

Accuracy can be a major problem when it comes to translation and interpretation. "Usually the interpreters aren't the ones also filling out the paperwork. It's important to make sure that what was said in one language is accurately being reflected in another language, Inaccuracy can have significant implications on insurance reimbursement billing, and healthcare record management. "As a patient and physician are sitting side-by-side having a conversation that's being recorded in both languages on a screen, they can see if there's a mistake or if something needs clarification. There's no wait time: immediate corrections can be made with people involved."

- Mobility

Another benefit of healthcare translation technology is its mobility. Consider the back-up hospitals can face at an emergency department admissions desk due to lack of available interpreters. Having a translation product brought to the ER when needed can reduce wait times for patients. Instead of having to wait for an interpreter to be found, the technology is already available in the hospital. "It would increase the quality of patient care, throughput, and overall healthcare experience, which means patient satisfaction goes way up,".

- Clinical applications

Technology like this isn't just beneficial for the administrative side of hospital life; it also makes a big difference on the clinical side, too. "You have three people in a conversation - the person speaking the foreign language, an interpreter and the physician. It seems like communication should flow freely and evenly in a situation like this, but great nuances can be missed," Having two records in front of a physician, one in the patient's language and a translated version, allows medical practitioners to see whether the questions they're asking are truly being understood by both the patient and the interpreter. "This technology allows us to capture the spirit and intent of a physician's effort to communicate with a patient while avoiding the game of 'Telephone, "The message isn't going to get lost along the way.

2. LITERATURE SURVEY

Quoc Truong Do, Sakriani Sakti, et al., [7] in 2018 proposed a work on emphasis speech translation .Emphasis is used to distinguish between focused and unfocused part of an utterance and it is useful in misheard situations in which speakers must repeat the most important words or phrases. The proposed approach to handle continuous emphasis levels which is based on sequence models and also they have combined the machine and emphasis translation into the single model. They have proposed hard attention emphasis speech translation and joint model. In the hard attention emphasis translation that can translate continuous emphasis weights without quantization. And in the joint model which

simplifies the translation and jointly translates words and emphasis with one-word delay. They have used LSTM-based encoder-decoder to solve the problems in single model of hard attention sequence-to-sequence model.

Hariz Zakka Muhammad, Muhammad Nasrun, et al., [8] in 2018 proposed a work on speech recognition by using hidden markov model. They consider the translation language from English to Indonesian. The classification method used is the Mel frequency Cepstral coefficients (MFCC) and Hidden markov model (HMM). The proposed system converts speech-to-text and uses the existing Google translation or Microsoft translation for translation. The speech signal will be processed by using MFCC. The algorithm used here is k-means algorithm.

J Poornakala, A Maheshwari., [9] in 2016 proposed a work speech-speech translation from English to Tamil language. Speech recognition system recognizing English speech via a voice recognition device. English speech translate the text into English using the speech synthesis, after converting it to text, compare with the words stored in the database if it matches the Tamil text stored in the database. The English text is converted to Tamil text is made by the machine translation system, and then the text will be displayed on the screen. It is designed to translate up 60 words in the Android application. The algorithm used here is HMM algorithm.

Sangmi Shin, Eric T.Matson, Jinok Park, et al.,[10] in 2015 proposed a work on speech-to-speech translation humanoid robot. The proposed system is helpful for English speaking patients they can explain their problems to Korean doctors. It can solve the problems of many patients who don't know the Korean language although they can describe their problems to doctor which replace the role of human workers. It uses CMU Sphinx-4 tool for speech recognition. English-Korean translation is based on rule based translation.

Seung Yun, Young-Jik Lee, et al., [11] in 2014 proposed a work on Multilingual Speech-to-Speech Translation System for Mobile Consumer Devices. In this paper they established a massive language speech database closest to the environment where the speech-to-speech translation device is actually used after mobilize many people based on the user survey requests. It was possible to secure excellent basic performance in the environment similar to speech-to-speech translation environment rather than just under the experimental environment. Moreover, with the speech-to-speech translation interface, a user-friendly interface has been designed and at the same time the errors were reduced during the translation process so many steps to improve the user satisfaction were employed.

3. SYSTEM ARCHITECTURE

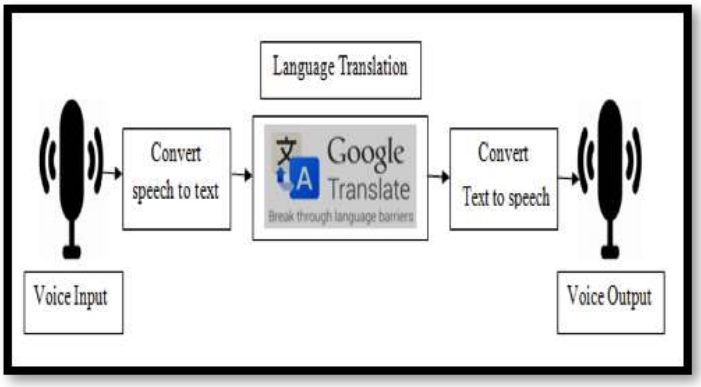

Fig 2: Architectural diagram for speech to speech Translation

The figure 2 depicts the architectural diagram for speech to speech system. First the user has to give the voice input in any of the four languages that are English, Kannada, Hindi and Telugu which then converts speech to text using the Google API. Then it gives into language translation where the preferred language is translated and the meaning of the sentence remains same which is done by computational intelligence. After the language translation it converts text to speech and finally at the last step we will get the converted voice output. Following steps involves in Speech-to Speech translation.

1)Voice Input: The voice input is taken as input to the system. The input voice may be any of the local languages here we taken for the four languages Kannada, English, Hindi and Telugu.

2)Speech to Text Conversion: Speech is a particularly attractive modality for human-machine interaction: it is "hands free"; the acquisition requires only modest computer equipment and it arrives at a very modest bit rate. Recognizing human speech, especially continuous speech (connected), without heavy training (independent of the speaker), for a vocabulary of sufficient complexity (60,000 words) is very difficult. However, with modern methods, we can easily process speech signals and convert them into text. The output of this phase is the text that was spoken by the caller.

3)Language Translation: Language Translation is done by Google Translator.

4)Text to speech Conversion: The output of the previous phase is Text. The existing speech synthesizer will be adopted which will convert text- to- speech. This speech will be easily understandable in the preferred language selected.

5)Voice output: After the language translation it converts text to speech and finally at the last step we will get the converted voice output.

1)Input

The voice input is taken as input to the system. User can give the voice input in any of the four languages they are Kannada, English, Hindi, and Telugu.

Classification of Speech Recognition

Speech depends on various factors that are speaker, bandwidth and size of the vocabulary. Speech recognition is classified into various types based on utterance, speaker mode and vocabulary size.

Classification of Speech Recognition based on the Utterances

There are four different types of speech available -Isolated Word, Connected Word, Continuous Speech, Spontaneous Speech.

Classification of Speech Recognition based on the Speaker Mode

Speech recognition system is broadly classified into main categories based on speaker models, namely, speaker dependent and speaker independent.

Classification of Speech Recognition based on the Vocabulary Size

The vocabulary size is divided into five categories namely, Small vocabulary, Medium vocabulary, Large vocabulary, Very-large vocabulary, Out-of-Vocabulary. [12]

2)Speech to Text Conversion

Input from microphone: Speech word samples are going to be extracted from no. of samples & unbroken in separate word. The microphone is employed to store the signal into the system.
The input is human voice that is sampled at rate of 16,000 per second. It should to run in live mode.

Speech generation: Speech generation devices, also known as voice output communication aids, are augmented and alternative electronic communication systems used to replace the speech or writing of people with severe speech to communicate verbally. Speech generation devices are important for people who have limited means of verbal interaction, as they allow individuals to become active participants in communication interactions. They are particularly useful for patients with amyotrophic lateral sclerosis (ALS), but have recently been used in children with predicted speech deficiency.

Speech Processing: Speech processing has been defined as the study of speech signals and their processing methods, as well as the intersection of digital signal processing and natural language processing. Speech processing technologies are used for digital speech coding, spoken language dialogue systems, speech-to-text synthesis, and automatic speech recognition. Information (such as speaker identification, gender or language identification, or speech recognition) can also be extracted from speech. Speech can be a more intuitive way to access information, control things and communicate, but there may be viable alternatives: speech is not necessarily the "natural" way to interact with a computer. Speech is hands free, without eyes, fast and intuitive.

Text output: Text output is generated after the speech processing.

3) Language Translation

Google translate

Google's translation system works primarily on text, but has a built-in feature that allows you to pick up a microphone input and then play back the sound from the speakers. Google Translate uses the classic speech-to-speech translation style with the use of a speech identifier for text speech, text translation, and speech synthesis to generate the audio associated with the text. It should be noted that Google's translation service is a difficult candidate to beat in terms of correct translations due to its well-designed implementation with huge amounts of input data from many different sources. [13]

A basic machine translation system makes a simple word substitution to convert text from one language to another. The translation process involves decoding the meaning of the source text and re-encoding the meaning in the target language. Word meaning refers to the concept or sense of the source text.

Machine Translation Systems can be built using several approaches like Rule-based, Statistical-based, and Example-based systems. In Rule-based translation to translate from English-to-Telugu and Example-based translation to translate from Telugu-to-English. In Rule-based translation, we perform source language text reordering, to match target language syntax, and word substitution. On the other hand, in Example-based translation, we map examples of text in source language to the corresponding representations in destination language, to achieve translation.

4)Text to speech Conversion: The output of the previous phase is Text. The existing speech synthesizer will be adopted which will convert text-to-speech. This speech will be easily understandable in the preferred language selected.

5) Voice output: After the language translation it converts text to speech and finally at the last step voice output is generated.

4. RESULTS

The speech translation system translates voice input given in Kannada, English, Hindi and Telugu language in real time.

Fig 3: Input speech format

The figure 3 shows the input speech given by the user in any of four languages given in the list.

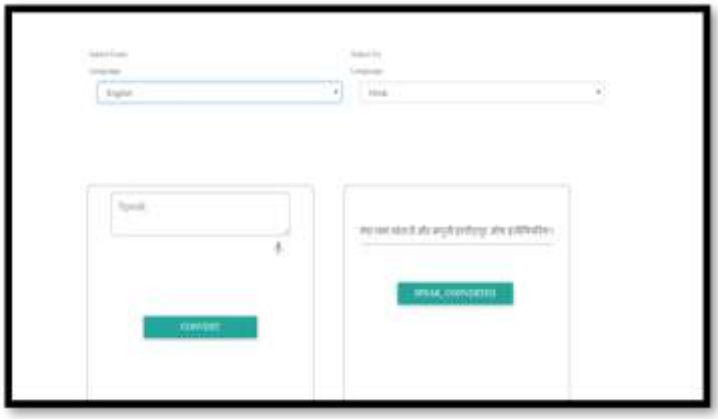

Fig 4: After language translation

The figure 4 shows the language converted by the Google API, after the language conversion user can listen it.

5. CONCLUSIONS

This section summarizes and concludes the contributions made by our project. A system is developed for the real time speech to speech translation based on the input speech given by user in any of the four languages they are Kannada, English, Hindi and Telugu. After the input speech is given, language translation is done via the Google API. Language barrier reduction is done via the speech to speech translation system. The system can defeat the constant challenges of unskilled individuals and improve their way of life

The outcomes show sensibly great achievement in perceiving Continuous speech from different speakers, for an enormous vocabulary. The various modules were examined in their separate areas and were effectively checked for various speech input.

6. FUTURE SCOPE

- **Retain the voice characteristics:** This will be achieved by fine tuning the system with larger training databases. The amplitude and frequency of the input and output voice should be same. That is to produce the same user.
- **Identify the human age based on voice:** There are various techniques available to identify the age of the person, one among them is to identify the age of any person by observing the voice characteristics. Finding age using voice recognition can be efficient for defensive purpose. The range of age of a person can be found by the speech recognition by programming millions of inputs with different pitches.

REFERENCES

- NEY, H., Speech translation: Coupling of recognition and translation, Acoustics, Speech, and Signal Processing, IEEE International Conference on., Vol. 1, 1999, pp. 517–520.
- Mrinalini Ket al: Hindi-English Speech-to-Speech Translation System for Travel Expressions, 2015 International Conference On Computation Of Power, Energy, Information And Communication.
- Development and Application of Multilingual Speech Translation Satoshi Nakamura', Spoken Language Communication Research Group Project, National Institute of Information and Communications Technology, Japan.
- Speech-to-Speech Translation: A Review, Mahak Dureja Department of CSE The NorthCap University, Gurgaon Sumanlata Gautam Department of CSE The NorthCap University, Gurgaon. International Journal of Computer Applications (0975 – 8887) Volume 129 – No.13, November2015.
- https://emerj.com/ai-sector-overviews/machine trans lation -14-current-applications-and- services/
- https://www.healthcarefinancenews.com/news/5-bene fits-healthcare-translation-technology.
- Sequence-to-Sequence Models for Emphasis Speech Translation. Quoc Truong Do,Skriani Sakti; Sakriani Sakti; Satoshi Nakamura, 2018 IEEE/ACM

Transaction on Audio, Speech and Language Processing.

- Speech recognition for English to Indonesian translator using Hidden Markov Model. Hariz Zakka Muhammad, Muhammad Nasrun, Casi Setianingh, Muhammad Ary Murti, 2018 International Conference on Signals and Systems (ICSigSys).

- Automatic Speech-Speech translation form of English language and translate into Tamil language. J Poornakala, A Maheshwari. International
Journal of InnovativeResearch in Science.Vol. 5, Issu e3, March
2016 ... DOI:10.15680/IJIRSET.2016.0503110. 3926.

- Speech-to-speech translation humanoid robot in doctor's office, Sangmi Shin, Eric T.Matson, Jinok Park, Bowon Yang, Juhee Lee, Jin-Woo Jung.2015 6th International Conference on Automation, Robotics and Application(ICARA).

- Multilingual speech-to-speech translation system for mobile consumer devices.Seung Yun, Young-Jik Lee, Sang-Hun Kim.2014 IEEE Transaction on Consumer Electronics.

- Sanjib Das, "Speech Recognition Technique: A Review",International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 Vol. 2, Issue 3, May-Jun 2012.

- https://static.googleusercontent.com/media/resear ch.google.com/en//pubs/archive/36340.pdf.