Speech to Speech Language Translator-Sinhala to English and English to Sinhala Translator

Shreya G S[1], Nidhi P[2], Oshine Chandra[3], Saad Ahmed[4], Dr. Kavitha C[5]

[1] *Student[BE], Department of Computer Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India*
[2] *Student[BE], Department of Computer Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India*
[3] *Student[BE], Department of Computer Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India*
[4] *Student[BE], Department of Computer Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India*
[5] *Professor, Department of Computer Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru , Karnataka, India*

ABSTRACT

Language translator is a speech to speech translation application for android mobile phone, which translates from English speech to Sinhala speech and vice-versa. The main objective of this translator is to enable smooth communication in medical sector between patients and Doctors/administration staff, therefore eliminate the language barrier. Language Translator consists of three modules- Automatic Speech Recognition (ASR), Machine Translation (MT) system, Text to Speech Synthesis. ASR module captures the source speech/voice of the user from the mobile through speaker, then converts the speech into text and then the text is sent to next module, Machine Translation system for further translation. Machine Translation (MT) module is the process of translation, this module already consists of library for both language and when text is received, it converts the text of source language to the target language as per user choice and sends the translated text to last module for further translation. Text to Speech Synthesis module is text to speech translator i.e. when it receives the translated text, this module further processes on converting it into speech and then makes it sends as user output in the desired target language demanded by user. Thus, Language Translation application works by combining all these three modules and gives the user best output.

Keyword:- *Sinhala to English Translator, English to Sinhala Translator, English, Singlish, Sinhala, machine translation, speech-to-speech translation, speech recognition and Speech Synthesis.*

1. INTRODUCTION

Sinhala, one of the official language of Sri Lanka is originated from one of the branches of the Indo-Aryan group of languages. The sound system of Sinhala consists of 14 vowels and 26 consonants. Sinhala is written from left to right. Any written Sinhalese word is pronounced the same way it is written, unlike English. English is a dominant West German language which originates from England. It has now become the most important language around the world because of its broad use. Translation of non-similar languages is much more difficult than the translation of similar languages. In this project, we are mainly focusing on medical sector. The conversation between a non–English-speaking immigrant of Sri Lanka and a Doctor or a administrative staff. Thus, this language translator was build with the intension of overcoming the language barrier up between the two parties. In this paper we show,

Sinhala to English and English to Sinhala language translator from speech to speech translation, which eliminates the limitations existing in the available translators. The Language Translator uses Speech to-Speech translation (S2ST) which removes the communication barrier between people speaking in different languages. S2ST systems are usually used for applications that supports conversations in non-native languages.

This work is done by three modules: Automatic Speech Recognition (ASR), Machine Translation (MT) and Text To Speech (Speech Synthesizer) modules. ASR captures the speech from devices and converts the dialect speech into text and the text is transcribed from English to Sinhala or Sinhala to English using Machine translators like Natural Language Processing, Hidden Markov Model, etc. and the transcribed text is synthesized into speech using text-to-speech synthesis (TTS) and generates speech in target language from the translated text.

Automatic speech recognition (ASR) is important information processing applications in the area of human language technology. ASR module captures the voice or source speech from the device through speaker and identifies the words spoken by the user and converts the speech into text and then the text is sent to next module. For example, a spoken language translation system takes the source speech signal as input and the output of ASR is as text which is then put into a machine translation system, which results in producing a translated text of another target language. The full SLT system can be viewed as ASR and MT subsystems in order. A spoken language understanding (SLU) system recognizes the input speech first, and then feeds the transcribed text to a natural language understanding (NLU) system. The NLU system will then identify the domain that the speech represents, and/or analyze the meanings enclosed in the source speech. In all the information mentioned above, ASR is a common component and plays a vital role.

Machine Translation (MT) Module converts the text of source language to target language without losing the meaning of source language and send the translated text to next module with least humans involvement. It matches source sentence with the source language parts of translation and if the exact match is obtained, then the target language sentence will be the desired result .This paper presents our approach to evaluate English to Sinhala and vice-a-versa machine translation system. There are numerous approaches to obtain the translation but we will be using Google API. The procedure to obtain the expected output in machine translation module begins with a input string of the source language in a text format, then the machine translator checks the database server to translate the input string to the expected output string and checks for error and translates the result in text format and sends the output to next module.

Speech Synthesis module is the last and final module which converts the text translated by Machine Translator into speech form in the target language and sends the output to the user. Speech synthesis module plays a key role in converting from text to speech which makes the Speech to Speech Translation system a better communication through device between patients and Doctors/Administration staff. The text obtained my Machine Translator(MT) in Sinhala language is converted to speech in Sinhala language.

2. LITERATURE REVIEW

In the modern time, its difficult to find and accompany a human translator everywhere we go. It's also difficult to learn all the languages of a place we would visit. Hence Language Translator comes into existence. It translates the native language to the desired language using Automatic Speech recognizer(ASR), Machine translator(MT), Text to Speech(TTS).Converting the words from native to desired language without changing its actual meaning is called Successful Translation of a language.

Machine translation decreases the human effort and time required for translation hence its has received lots of importance nowadays. Many translators have been proposed for different languages. Most of them have been focused on Indo-European, Indo-Aryan or Sino-Tibetan families Sinhala is an Indo-Aryan Language. The language structure of Sinhala and English are different from each other, hence only few translators have been proposed to translate Sinhala to English and vice versa. The English to Sinhala translator proposed in achieved 89% accuracy with a word error rate of 7.2% and a sentence rate error of 5.4% for 200 sample sentences. A translator from Singlish or English to Sinhala is proposed, some are Unicode Converter, Google Transliteration IME. English and Sinhala definitions are stored in dictionaries, one of such dictionary is Madura Dictionary. In any language there would be ambiguity for a word and the meaning of it totally depends on the situation it is being used. This is the main barrier for language translation.

Therefore when the machine is translating the sentence this ambiguity needs to be handled efficiently to output the correct interpretation of the sentence. Wrong interpretation of a word can lead to miscommunication. Human Interaction is required for the correct interpretation of the sentence. It can be improved by increasing the learning rate of the machine by providing lot of statistical data.

3. METHODOLOGY

In the proposed system, it has three main parts
• Recognition
• Translation
• Synthesis

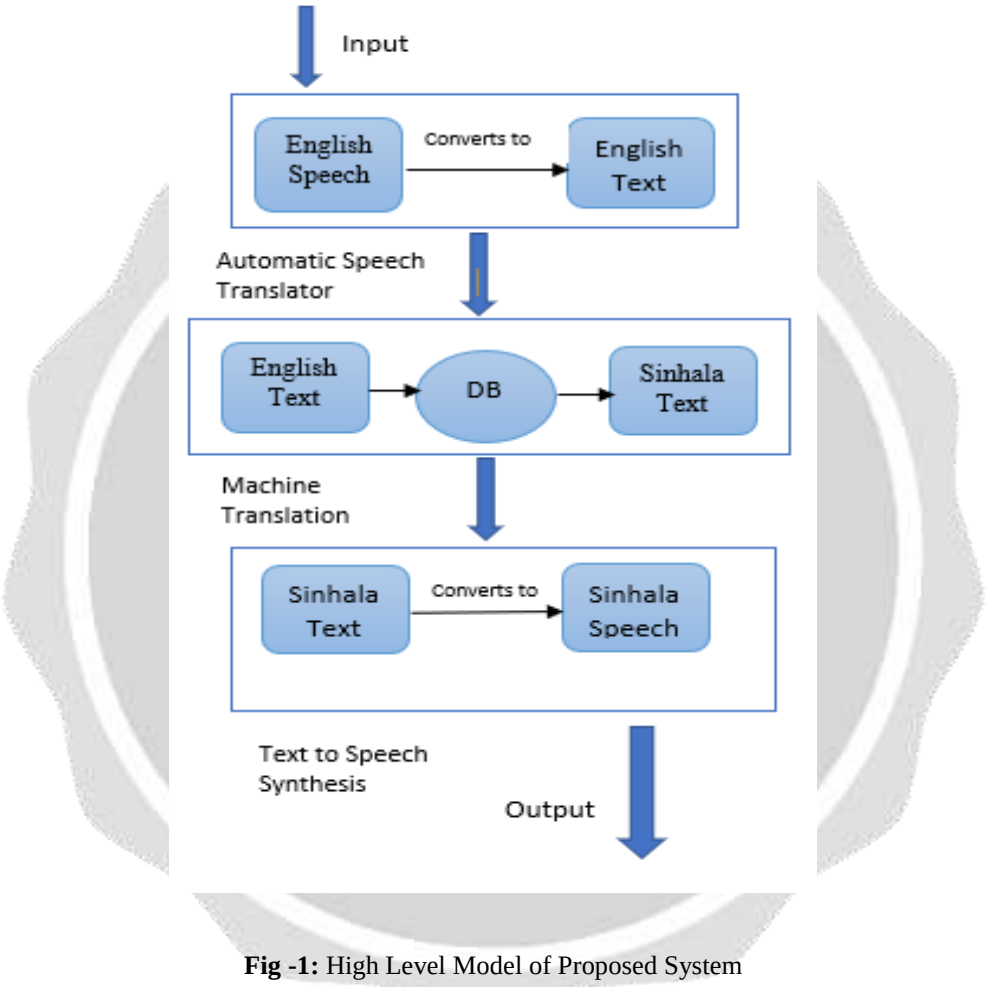

Fig -1: High Level Model of Proposed System

User's/Patient's voice will be recorded and transcribed from their dialect into text. In Translation part this text is then transcribed from one language to another language text using technologies like NLP, HMM. While Synthesizing the text is then translated into voice of same language.

3.1 Methodology used for Speech Recognition

The task of a Automatic Speech Recognition (ASR) motor is to require sound input and turn it into a composed representation. Typically now and then called "translating," since you'll think of the sound as an input channel and the composed frame, whether it could be a arrangement of words or discourse sounds, as the yield channel, with the decoder working between them. Once the discourse has been changed over into a few content representation, it can be utilized to send occasions to an application. In our project we are using CMU Sphinx tool for the purpose of speech recognition. It is a speaker independent large open- source vocabulary continuous speech recognizer that is released under the BSD style license.

3.2 Methodology used for Machine Translation

The process of translating the native language into the chosen language, without losing the meaning of source language is said to be Machine Translation. The SLT system, which comprises of ASR and MT subsystems in tandem, is further associated in tandem with an NLU subsystem, we create a cross-lingual SLU framework. The applications of SLT are different, extending from machine-aided human interpretation to proficient interpretation administrations for worldwide organizations.

An SLT system with the ASR component to supply the input to the MT component is more troublesome than text based MT since of the compounded challenges of ASR and MT. A specific issue in SLT is discourse disfluency, making the input to the MT component of the SLT framework, indeed with culminate ASR, go astray from lexical, syntactic, and semantic designs of typical composed writings that are typically used for preparing the MT system. Examples incorporate filled delays, paragraph and sentence delimiters, accentuation marks, and capitalized words. This deviation, in conjunction with ASR blunders, produces a genuine ‘‘mismatch’’ between preparing and testing conditions. One way to address this mismatch issue is to embrace the Bayesian approach where the vulnerability of ASR yields is taken into consideration. Whereas the beginning unrefined scientific detailing of this approach showed up within the early days of SLT investigate and afterward amplified to joint ASR and MT decoding through an ASR lattice or confusion network, only at the translating stage has the ASR vulnerability been considered until or maybe as of late when the same uncertainty was consolidated into the preparing prepare with a decision-feedback style. We are using Google API’s for implementing the above described architecture.

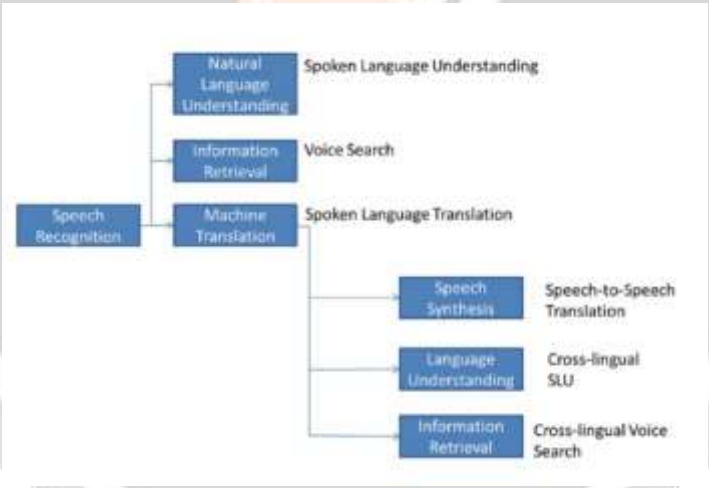

Fig -2: Illustration of the general architecture of common Language translator systems and their relations in terms of the shared subsystem components.

3.3 Methodology used for Speech Synthesis

The TTS framework changes over an subjective ASCII text to speech. The primary step includes extricating the phonetic components of the message, and we get a string of symbols representing sound-units, boundaries between words, phrases and sentences together with a set of prosody markers. The moment step consists of finding the coordinate between the sequence of symbols and suitable things put away within the phonetic stock and official them together to create the acoustic flag for the voice yield gadget. We are implementing Free TTS for speech synthesizing which is an open source software.

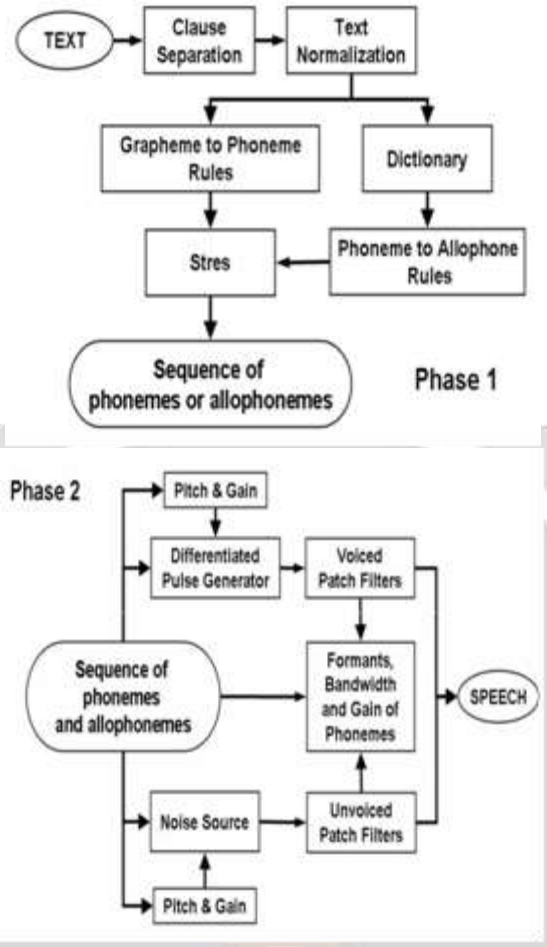

Fig 3: Phases of speech synthesizer

4. RESEARCH FINDINGS AND EVIDENCE

There is no such machine which can convert Sinhala to English or vice-versa with 100 percent accuracy. No matter how efficient our model is, we can't achieve 100% accuracy. Keeping this in mind, we are trying to build a model which can maximize this efficiency and give us as accurate results as possible. We would be aiming to implement this keeping in mind all the tenses as well, that is, present, past and future tense.

Therefore, sentences of different tenses would also be converted accordingly. Our primary goal would be to build a model which can convert Sinhalese language to English and English to Sinhalese language with minimum error rate. These are the following research findings while doing so.

Cross-Lingual Spoken Language Understanding: Porting an SLU service from one source language to another destination language is of huge importance and is thus of increasing interest to both the language understanding and machine translation research communities. One possible method could be: first translate the utterances in the second language to the initial language and then to use the initial languages SLU models to analyze them. Another method could be: first translate the annotated corpora to the second language, which is quite expensive, followed by training models from understanding examples in the second language.

Spoken Language Understanding: Majority of the SLU methods are based on the divide and conquer approach that separates the ASR front–end and the NLU back–end, there are a few exceptions. Better speech recognition accuracy is a positive indicator for better SLU accuracy. In fact, there is experimental proof that higher WERs may correlate with better slot filling accuracy as long as model training criteria match the optimization objective for understanding.

Spoken Language Translation: Spoken language translation has got a lot of attention in recent times. It allows translation of speech signals in a corresponding source language to text in the corresponding target language. This issue essentially deals with Machine translation, Machine Learning and Automatic Speech Recognition. The utterances from the spoken dialect are first recognized and converted to the corresponding text and later this source text is converted to the target language.
Further we would try and make our model adapt to the following grammatical patterns:

4.1 Ambiguous Words of Sinhala Language

One crucial problem in Sinhalese communication is the ambiguity that exists in the language. There may be certain situations where the meaning of a word can vary. The exact meaning of the word is based on the situation in which it was used. This requires more of the human brain to decode it and correlate the actual meaning. Now a machine cannot work exactly like the human brain. So there must be an alternative approach to this problem. The solution to this is based on making our machine output all the possible meanings of such words and henceforth, give the user an option to choose the word with the nearest meaning to the word being referred to.

For example:

Input	ඔහු **ඉරක්** අඳින්නේය.
Output 1	He draws **a line.**
Output 2	He draws **a sun.**

Fig -4: Example

4.2 Sinhala Words with Multiple English Words

Each English word itself can have many meanings. Now, one can imagine the variety of words that may arise while translating a word from Sinhala to English. It is not surprising that each Sinhala word could translate to many English words. This could become troublesome because we would have to choose the right word for the sentence to make sense and thus accurate translation of the sentence becomes cumbersome.

For example, take a look at this sentence conversion with multiple English meanings to choose from:

Input	**sudhu** Lamaya: pothak: kiyawan:ne:ya සුදු ළමයා පොතක් කියවන්නේය.
Output 1	**White** child reads a book.
Output 2	**Fair** child reads a book.

Input	Lamaya: pa:ra **pAn:ne:ya** ළමයා පාර **පැන්නේය.**
Output 1	Child **crossed** the road
Output 2	Child **jumped** the road

5. CONCLUSIONS

Conventional machine interpretation strategies are speedy and effective enough for processing. They have been appeared to be vital in giving incredible results with their limit capability of operation. The English to Sinhala machine interpretation framework assessment has been conduct through the three stages. As a to begin with step, assessment was conduct through the white box testing approach and tests each module within the MT system through the created testing instruments. At that point, assess the framework execution and calculate the mistake rate through the assessment test tools. At last, comprehensible and precision tests will be conducted through the human support.

At present system uses restricted number of lexical assets. Upgrading lexical assets are the further work of this project. One of the main barriers in further upgrading the translator is the word ambiguity. A few words cannot be interpreted without appropriate human interaction. As a future work framework can be created utilizing factual base investigation with the inclusion of a prepared corpus. The other remaining work with respects to the interpreter is to include more words to the information base. As of now deciphered output is generated with the use of a rule base where only the introduced rules are being utilized. Our focus is to automate the as of now utilized process with the use of artificial intelligence.

6. ACKNOWLEDGEMENT

We would like to express our utmost gratitude to our principal Dr. Ravishankar M sir and our vice principal and HOD of CSE dept Dr. C Nandini ma'am for giving us an opportunity to showcase our project and research work and our beloved guide Dr. Kavitha C ma'am for guiding us in innumerable ways towards the development and success of the project**.**

7. REFERENCES

[1]. Xiaodong He, Li Deng. Speech-Centric Information Processing: An Optimization-Oriented Approach. Proceedings of the IEEE (Volume: 101, Issue: 5, May 2013) , 07 March 2013.
[2]. Lara J. Martin, Andrew Wilkinson, Sai Sumanth Miryala, Vivian Robison, and Alan W Black. Utterance classification in speech-to-speech translation for zero-resource languages in the hospital administration domain. 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU),3-6 Dec 2018.
[3]. M.D. Faizullah Ansari, R.S. Shaji, T.J.SivaKarthick, S.Vivek, A.Aravind.Multilingual speech to speech translation system in Bluetooth environment. 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 22-Dec-2014.
[4]. L.Wijerathnl, W.L.S.L.Somaweera, S.L.Kaduruwana, Y.V.Wijesinghe, O.I.Oe Silva, K.Pulasinghe, S.Thellijjagoda. A Translator from Sinhala to English and English to Sinhala (SEES). International Conference on Advances in ICT for Emerging Regions (ICTer2012),01- Feb-2013.