Interspeech 2018
2-6 September 2018, Hyderabad

Online speech translation system for Tamil

Madhavaraj A, Shiva Kumar H R and A G Ramakrishnan

MILE Lab, Electrical Engineering, Indian Institute of Science, Bangalore 560012, India

madhavaraja@iisc.ac.in, shivahr@gmail.com, agr@iisc.ac.in

Abstract

In this paper, we present an application, which recognizes spoken Tamil utterances and speaks out the recognized text in Tamil through our Tamil text-to-speech (TTS) system. Further, we translate the recognized Tamil text to English using google translate and play it through our English TTS. Our Tamil speech recognition system, which can recognize about 75,000 words, has been trained on a 150-hour transcribed speech corpus. We have trained a deep neural network for the acoustic model and employed tri-gram language models to build our recognition system. Our Thirukkural TTS system performs unit-selection based, concatenative speech synthesis, using 2.5 hours of Tamil spoken utterances transcribed at the phone-level. Our English TTS uses 2.7 hours of phone-transcribed utterances. This is a technology demonstration of a complete web application, which, when perfected, could be used to assist Tamil users in learning English, by speaking in Tamil into the system. The playback of the recognized text from Tamil TTS serves to demonstrate the effectiveness of the Tamil ASR to the majority of the conference registrants (who cannot read the recognized Tamil text.

Index Terms: Speech recognition, text-to-speech, Tamil, English, translation, deep neural networks, acoustic model, language model, web application.

1. Introduction

Recent decade has seen steady progress in the areas of speech recognition and translation. Most of the research is carried out for English and other European languages. Substantial progress has not been made for Indian languages, especially Dravidian languages. Tamil is one of the Dravidian languages and is the oldest living language in the world. It is spoken by 80 million people around the world and is the official language of the states of Tamil Nadu and Pondicherry, and one of the official languages of Srilanka and Singapore. Our research group has developed technologies for Tamil and Kannada, like recognition of online handwriting, optical character recognition, text-to-speech and transliteration systems.

In this paper, we present a translation system which converts spoken Tamil utterances to English speech. It also plays back the spoken Tamil utterance in a different voice, as a technology demonstration. The heart of any speech translation system is large vocabulary, continuous speech recognition (LVCSR). Developing such a system for Tamil involves many challenges: (i) unavailability of standard, transcribed speech corpus, (ii) infinite vocabulary problem, and (iii) no standardized lexicon. In spite of a rich heritage and literature, Tamil is considered as a low-resourced language in this respect. The translation system presented comprises: (i) a Tamil LVCSR system, built by us using 150 hours of transcribed speech and built with a vocabulary of 75,000 words, (ii) a Tamil and English TTS engine built using 2.5 and 2.7 hours of phone transcribed speech respectively, and (iii) a text-translation system which basically uses Google translation service.

The rest of the paper is organized as follows. Section 2 describes the workflow of our speech translation system. In section 3, we elaborate on the design of our LVCSR and TTS engines, which are the main building blocks of our system. We conclude in Section 4 and provide future research and development directions.

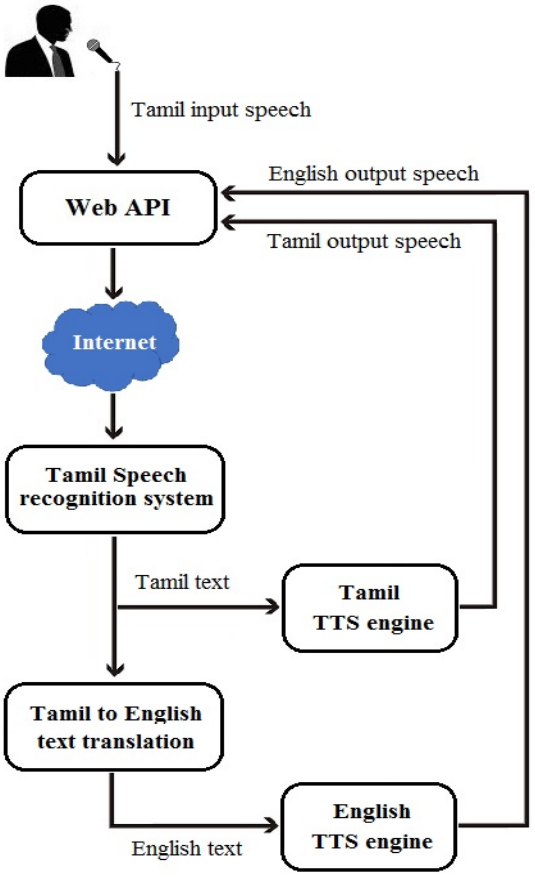

Figure 1: *Block diagram of the Tamil to English speech translation system.*

2. Workflow of the system

A web application program interface (API) using javascript and php scripting languages has been developed (see Fig. 2), where the user can click the record button and speak a Tamil sentence. The utterance is recorded and the audio file is sent to our server, which performs the speech recognition task and passes the rec-

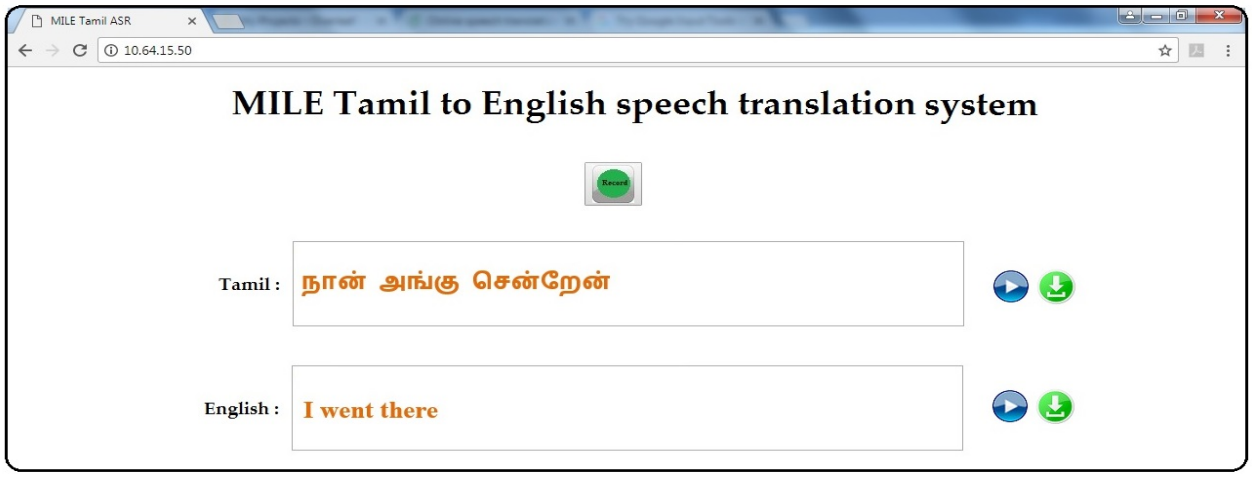

Figure 2: *Screenshot of the web API, which demonstrates our Tamil ASR, TTS and English TTS.*

ognized Tamil Unicode text to Google translate API and gets the English-translated text which is then passed to our English TTS engine to synthesize the speech. The recognized Tamil text is also sent to our Tamil TTS engine, which synthesizes Tamil speech. The recognized Tamil text, translated English text and their synthesized speech waveforms are then transferred to the client-side and are made available for viewing, playback and download. The Tamil TTS playback is also meant to demonstrate to the majority non-Tamil audience in this International Conference, so that they can appreciate the performance of our Tamil ASR.

3. Building blocks of the system

Three main blocks of this translation system are LVCSR, text translation and TTS, which are explained below.

3.1. MILE Tamil LVCSR

MILE Tamil LVCSR is an extension of our previous work in [1]. This system has been trained on a 150-hour transcribed speech corpus, spoken by 450 speakers. We have used time-delay neural network (TDNN) and hidden Markov model (HMM) based, context-dependent triphone acoustic model and tri-gram word-level language model. We have also written our own application, which converts Tamil Unicode words to phones, which serves as the pronunciation dictionary for the LVCSR. We have used the wsj recipe of Kaldi [2] to train a sequence of models, namely monophone, triphone, linear discriminant analysis and speaker adaptive training (SAT) based acoustic models and the alignments obtained after SAT has been used for TDNN training. Our TDNN-HMM based LVCSR system can recognize 75,000 words present in the vocabulary, with an accuracy of about 83%.

3.2. Tamil to English Text translation

Text translation from Tamil to English is performed through Google's cloud translation API [3]. This API is called from the server-side, since the translated text needs to be further processed by the English TTS engine.

3.3. MILE Tamil and English TTS

From the initial sentence-level transcribed corpus, we perform forced-alignment using SAT model on short utterances using the procedure explained in section 3.1. MILE Tamil and English TTS [4] uses this phone-level transcribed utterances for unit selection. For an input text from a given language (i.e., outputs from LVCSR and translation systems), we use lexicon for that language to convert graphemes to a sequence of phones and their corresponding speech units are selected based on a Viterbi search. The objective function minimized for selecting the best path is based on the difference between mean pitch values of the subsequent concatenation units. Once the best sequence of speech units is selected, POS tag information is used to insert pauses of varying durations at the required locations in the sequence. Post-processing such as low-pass filtering is performed at the locations, where the speech units are concatenated.

4. Conclusion and future directions

We have presented a web API, which can be used to convert spoken Tamil utterances to Tamil and English texts, which can then be played back as speech utterances. This tool will be helpful for users who wish to learn English through Tamil, and to learn proper pronunciation for Tamil words. We have built ASR, TTS and used Google text-translation tool to build this application. Our future work will be to extend this application to Kannada language.

5. References

[1] A. Madhavaraj and A. G. Ramakrishnan, "Design and development of a large vocabulary, continuous speech recognition system for Tamil," *Proc. 14th IEEE India Council International Conference (INDICON)*, Dec 2017.

[2] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The kaldi speech recognition toolkit," *IEEE Workshop on Automatic Speech Recognition and Understanding*, 2011.

[3] *Google translate*. https://translate.google.com.

[4] B. S. R. Rajaram, H. R. S. Kumar, and A. G. Ramakrishnan, "Mile tts for Tamil for blizzard challenge 2014," *Proc. Blizzard Challenge Workshop*, pp. 201–202, 2014.