International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-4, April 2019

A Simple Tamil Speech Recognition System Based on Cmusphinx

Arvind Madaboosi Mukund, Priyanka Balaji Ramanathan, T. Sujithra

Abstract*: This paper focuses on the research of phonemes and their usage for Tamil speech recognition using the CMUSphinx API [4]. Tamil did not have a solid speech recognition application especially the Tamil that is spoken on a daily basis. This paper is the outcome of the authors building a new dictionary for 418 words (as of April 27 2019). This paper is intended solely to show the result of mapping Tamil grapheme to phoneme in hope that it will help similar sounding languages and the development of the Tamil language itself. The model used recorded speech to map the graphemes in the built dictionary to the phonemes in the recorded words. The accuracy of the model is represented by determining the Word Error Rate generated by the language model.

***Index Terms*: Language Models, Word Error Rate, Speech Recognition, Phonemes, Graphemes.**

I. INTRODUCTION

Tamil is one of the oldest languages in the world. It is recognized as an official language in countries such as India, Sri Lanka, Singapore, South Africa, Malaysia, and Mauritius. As of 2018, around 1% of the world's population communicate in Tamil. Tamil grammar has 18 consonants and 12 vowels and one aayutha ezhuthu (ஃ). To help transcribe sounds that aren't originally part of the Tolkapiyyam classification, Tamil incorporates phonemes from the Grantha Script in Sanskrit. The Grantha consonants in Tamil are ஜ், ஷ், ஸ், ஹ், க்ஷ [1]. Every alphabet in Tamil falls under one of the three categories: Vallinam (hard), Mellinam (soft), Idhaiyinam(medium) based on the hardness of the sound produced when uttering the alphabet's sound.

Our research focuses on improving the technology for spoken Tamil by aiming to increase the accuracy of Tamil speech recognition. [2] concludes that the CMUSphinx Language model is determined to relatively be the most accurate.

CMUSphinx [4], a set of speech-recognition systems, was established by Carnegie Mellon University in the year 2015. Pocketsphinx, which is an application of CMUSphinx [4], helps with creating a language model for us.

This paper attempts to map the phonemes with their Tamil counterparts to create a spoken Tamil Speech Recognition tool.

Manuscript published on 30 April 2019.
***** Correspondence Author (s)
Arvind Madaboosi Mukund, SRM Institute of Science and Technology, Chennai, India.
Priyanka Balaji Ramanathan, SRM Institute of Science and Technology, Chennai, India.
Dr. T. Sujithra, SRM Institute of Science and Technology, Chennai, India.

II. ARCHITECTURE

The build system we use creates the files required for phinx-train in a process shown in Figure 1. These files are

A. Language Model

The Language Model file is based on a corpus containing number of sentences and words. The file, using n-grams, calculates the probability of a specific n-worded sequence is uttered. 1-grams, 2-grams, 3-grams are determined in this file. This file is modified with a specific n-gram modifier by testing to improve the word detection accuracy. This file also helps identify and thereby prevent contextual errors that could arise from homophones and other words that have a very similar pronunciation. Similar sounding words, like for example, எண்ணை (oil) and என்னை (me) could lead to erroneous translations by the system. Using bigrams and identifying the words linked to the word could assist the system into accurately associating the meaning of the word

n-gram is a measure that employs conditional probability to identify and predict the sequence of a phrase.

$$P(w_n | w_{n-1} w_{n-2} w_{n-3}) \approx count(w_n w_{n-1} w_{n-2} w_{n-3})$$

$$P(Enna | Unnoda\ Peyar) \approx \frac{count(UnnodaPeyarEnna)}{count(UnnodaPeyar)}$$

B. Phonetic File

The .phone file holds the list of all phonemes that is incorporated in the dictionary. This is used to verify and test if similar sounding words have same phonemes.

C. Transcription File

The training transcription is generated from the wav files of Codec: PCM S16 LE (s16l). The wav files are recorded at a specific frequency. The frequency is determined by the parameters in the *sphinx_train.cfg*. CMUSphinx supports only mono-channeled recordings. Once these files are read by the build system, it generates a training transcription for all the above words and sentences.

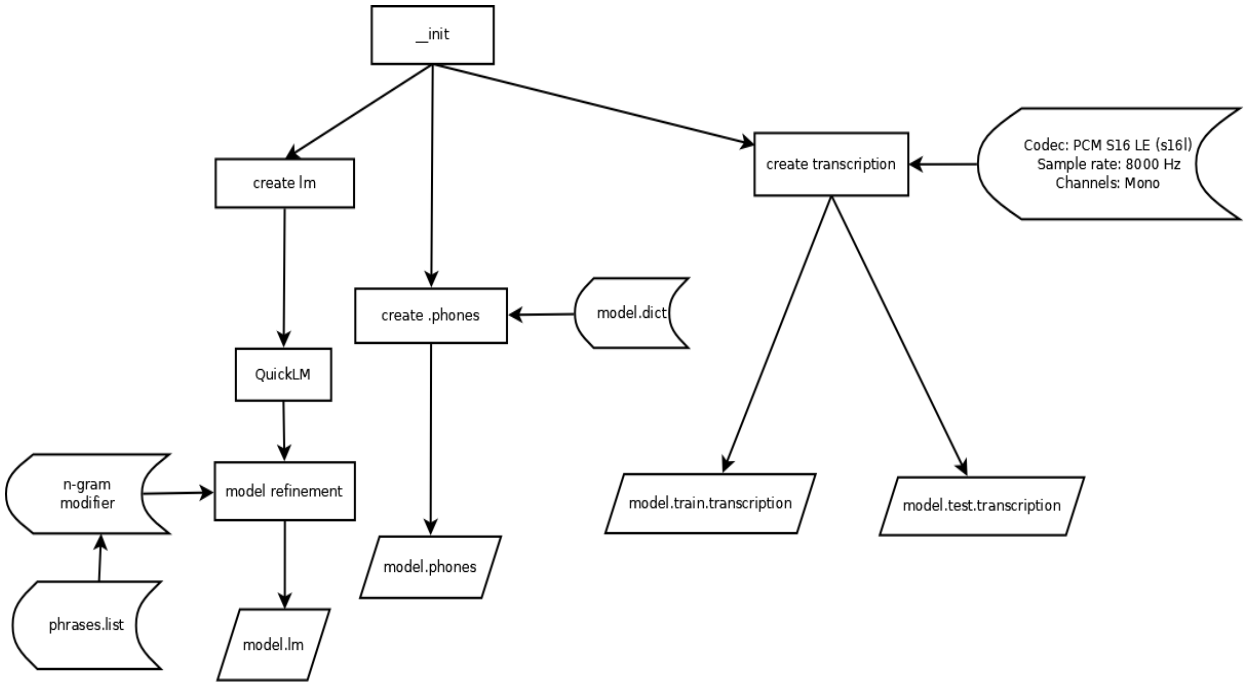

Figure 1 – Build System Architecture

helps to check for error rate under different noise environments and ambient sounds.

D. Other Files

1. Filler File
   In day to day conversation, people are likely to pause between phrases. These pauses help them think about they are about to say whether it is to find the appropriate word or to give an appropriate reply. These pauses are called fillers. The filler file contains the list of all the conversation fillers that one might come across in day to day conversation including "silences" and "um".
2. Dictionary
   The dictionary (DICT) file contains the grapheme to its phoneme conversion. A grapheme is a smallest fragment of a word that cannot be further divided without changing the phonetic sound of the word. In English, graphemes can consist of more than one letter. However, in Tamil, all the graphemes typically consist of a single letter. The phonemes are transcribed for every grapheme in that word manually and stored in the DICT file. The accuracy of translated phonemes is tested on a trial and error basis to be able to accurately match the grapheme to the phoneme. Refer Table [1] for a representation from Tamil speech sounds to its phoneme.

III. IMPLEMENTATION

From a list of 1000 most popular words in Tamil [3], the list was shortened the list to exclude words that aren't used in common, informal conversation. Words that are phonetically similar despite having difference in meanings were merged.

Other archaic words were also removed thereby bringing the word vocabulary list to 418 words.
These words were then recorded at 8000Hz frequency. The phonemes for each of these words are manually inputted into the DICT file.

Implementation of this project depends on the CMUSphinx [4] project or more specifically the pocketsphinx [4] Application Program Interface (API) which contains a lightweight speech recognition library with a python wrapper. The dictionary is created by hand initially containing the essential Tamil sounds and its corresponding phonemes. Sequence to sequence conversion tools like the g2g-seq2seq could be used to expand the amount of words in this dictionary.

A. Training Phase

The feature vectors from the configured feature parameters for the model is extracted from the wav files and saved as its feature vector file (MFC) counterparts.

B. Testing Phase

Initially we introduce some noise to all of our training data

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Tamil Grapheme	Phoneme
அ	AH/A
ஆ	AA
இ	IH
ஈ	IY
உ	UH
ஊ	UW
எ	EH
ஏ	EY
ஐ	AY
ஒ	OH
ஓ	OH
ஔ	OW
க	K
ங	NN
ச	S/CH
ஞ	NN J
ட	D
ண	NN
த	DH
ந	NN
ப	P/B
ம	M
ய	Y/YY
ர	R
வ	V
ல	L
ழ	Y H
ள	LL
ற	TR
ன	N
ஸ	S
ஷ	SH
ஜ	J
ஹ்	H
க்ஷ	K SH

Table 1 – Phoneme to Grapheme conversion

The testing transcription is generated pseudo-randomly from input wav files with distortion and noise. This file helps to check for error rate under different noise environments and ambient sounds.

IV. FINDINGS

What is observed from the Table 1, is that, although the Tamil graphemes consists of one unit, the phonemes do not have the same rule. Graphemes such as ங, ஞ, and க்ஷ have NN G, NN J, and K SH as their phonemes respectively. Words such as மஞ்சள் (Manjal), வாங்க (Vaanga), and நட்சத்திர (Nakshatira) could offer more insight on the phonetic identities.

Vowels in Tamil generally are of two types: the short vowel, and their elongated counterpart. It is observed that the short vowels are phonemes that are typically ending with H such as EH, AH, UH, et cetera whereas the longer vowels are matched to phonemes that end with other characters that produces a harder sound to the phonemes. Examples of this case are UW which is the elongated counterpart of UH. In this case, W is used to emphasize the phoneme. Another example is EY which is elongated from EH by replacing H to Y. Similarly, phoneme IH is mapped to இ and ஈ maps to IY.

Table 2 – Example Tamil phonetic pronunciations for phonemes

Phoneme	Example Tamil Word	Transliteration
A	நட்சத்திர	Nakshathira
AA	பால்	Paal
AH	அவர்	Avar
AY	கை	Kai
B	நம்பிக்கை	Nambikkai
CH	முயற்சி	Muyarchi
D	வேண்டும்	Vendum
EH	தெரு	Theru
EY	மேகம்	Megam
G	கடிகாரம்	Gadigaaram
IH	இசை	Isai
IY	தீ	Thee
J	ராஜா	Raja
K	கவிதை	Kavidhai
L	தலைவர்	Thalaivar
M	மீன்	Meen
N	நடனம்	Nadanam
OH	சொல்	Sol
P	பந்து	Pandhu
R	மரம்	Maram
S	சண்டை	Sandhai
SH	விஷயம்	Vishayam
T	வெட்டு	Vettu
TH	தோட்டம்	Thottam
TR	வெற்றி	Vetri
UH	பத்து	Pudhu
UW	பூனை	Poonai
V	சுவர்	Sevar
YY	பையன்	Paiyan

Some words with the அ did not map to AH but mapped to A. This is because of the emphasis that is placed on the அ when uttering the words.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

For example, the word நட்சத்திர (Nakshatira) has a softer emphasis on the அ sound. Therefore, it maps to phoneme A. Whereas, the word அவர் has a harder emphasis on the அ sound thereby mapping to AH.

In the Tamil language, there are consonants that hold two different phonetic sounds. For example, the ச can read as CH **and** S. ப which can be pronounced as B **and** P. In these cases, the word is transcribed in English to identify the phonetic sound and mapped accordingly.

Another unique aspect of the Tamil language is the ழ consonant. This is a special consonant called the retroflex approximant. This consonant, earlier in all Dravidian languages, is presently only in Tamil and Malayalam. When transcribing Tamil words in English, ழ is written as 'zh'. Using trial and error method, we were able to map ழ to YH using words like பழம் and வாழ.

As mentioned eariler, Tamil letters fall under three categories based on the way they are pronounced: Vallinam (hard group), Mellinam (soft group), and Idaiyinam (medium group). It has been observed that some vallinam characters have a phoneme that are a combination of two individual phonetic sounds. Examples are ற which has the phoneme TR mapped to it.

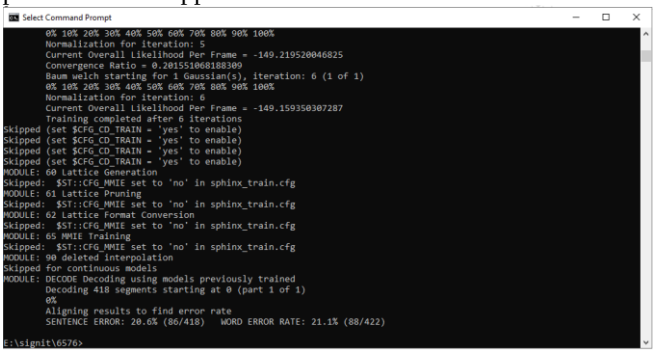

Figure 2 – Word error rate of implemented model

V. Conclusion

The speech recognition tool is created almost completely based on python with underlying C implementations. The speech recognition accuracy has been measured to have a WER of 21.1%. and a sentence error rate of 20.6%. The paper although for Tamil can be expanded to a lot of similar scripts with similar phonetic base as Tamil.

VI. Use Cases

This paper is being used to develop a speech to sign language tool for hearing impaired people. This paper intends to increase the diversity of languages in the speech recognition realm and hopes to bring uses in **spoken** Tamil.

REFERENCES

1. A. G. Ramakrishnan and L. Narayana, "GRAPHEME TO PHONEME CONVERSION FOR TAMIL SPEECH SYNTHESIS," in Workshop in Image and Signal Processing (WISP-2007), At IIT Guwahati, Guwahati, 2007 .
2. A. Mukund and P. Ramanathan, "Tamil Speech to Indian Sign Language using CMUSphinx language Models," International Research Journal of Engineering and Technology (IRJET), vol. 06, no. 02, pp. 1812-1814, 2019.
3. Common Tamil Words, [Online]. Available: https://1000mostcommonwords.com/tag/tamil-words/.
4. "CMUSphinx," [Online]. Available: https://cmusphinx.github.io/.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)