Identification of Effective Sub-Word Unit for Speech Recognition in Tamil

S. Thirumal*

Department of Computer Science and Engineering, VISTAS, Pallavaram, Chennai.

A. Manikandan

Department of Computer Science and Engineering, VISTAS, Pallavaram, Chennai.

E-mail: mani.se@velsuniv.ac.in

Dr.R. Anandan

Department of Computer Science and Engineering, VISTAS, Pallavaram, Chennai.

E-mail: anandan.se@velsuniv.ac.in

**Corresponding author E-mail: thirumal.se@velsuniv.ac.in*

Abstract

In the speech recognition the unit selection plays a vital role, here it is decided to take a sub-word unit syllable with is the optimal level to have when compared to phone because training the system is more complex. It is proposed to develop a speech recognition system for Tamil which uses syllable a sub word unit. Initially we start with isolated word recognition for Tamil using syllables and expand the system to continuous speech recognition system.

Key Terms: Hidden Markov Model, Context Independent, Context Dependent, Word Error Rate, Indian Standard Code for Information Interchange.

1. Introduction

Speech recognition system is technology of recognizing and converting the speech signals into a series of units called symbols and then converting those symbols into text, corresponding to the speech. It can also be defined as the ability of the system to identify the sentences of the language which in turn converted to machine understandable format. In speech recognition there have been various researches, using models like phone model, triphone model, syllable model and word model. Each model has its own advantage, for example triphone can model a short – term contextual variation to certain extent. Similarly syllable has its own importance, being a natural unit of speech it takes the main role in Tamil Language, here nearly all the syllables where denoted by a letter. In speech recognition, bigger the unit like syllable and word are referred for training but at the same time the training the system becomes complex.

We propose an approach in which the processing deals with Unicode. We obtain a Tamil text in Unicode format and convert it into Indian Standard Code for Information Interchange (ISCII) and from ISCII to ITRANS code, which is Romanized text of Tamil. The transcribed sentences were parsed into syllable using the language linguistics and the text is converted back to Unicode.

2. Units of Speech

In this section we discuss about the various units of speech and their significance in detail.

Received: 25 Oct 2018/Accepted: 25 Nov 2018

Phones: The mot convincing sub-word units are phone and phone sets which can be trained reasonably. But the problems we face is that the phonetic sounds for different word varies because of this the phonemes gets affected. The variability in the acoustics of phonetic units based on the context not understandable in most of the language.

Triphone: It is another sub-word models which is also most successful unit. There are also some results shown that the triphone will reduce the error nearly for the 50%. But the problem in this units was it requires more memory space for training set. It is also noted that the Triphone system neglects the similarity between them but this problem was addressed by Lee (1990) at the same time another approach by Huang and Hwang based on clusters similar states of model, finds a way for making granularity more finer and betterment in the performance also it represents states similar to Markov states

Syllable: The syllable is one of the most popular which is some bigger than triphone. The significance of using the syllable is it minimizes the contextual variations that occur in the phone models and also based on the generation and receiving of speech syllable is considered as the compromising unit for segmentation of speech signals. Now more number of researchers are concentrating their work on speech, using syllable as unit of recognition by adapting different approaches.

3. The Tamil Language

Tamil is an ancient language which has the history of two thousand years it an Dravidian language spoken mostly in Tamilnadu a state in India and Sri Lanka. It is the official language of the state of Tamilnadu and also has official lang in Sri Lanka and Singapore. It identified as one of the classical languages in India, and funds are the allocated for development the language. There are more than 75 million speakers for Tamil and this is one of the widely spoken languages of the world.

3.1. Tamil alphabets

The language is composed of 12 vowels and 18 consonants. The vowels and consonants combine to form a 216 compound characters set. There is another character in called aytham ( ) which is rarely used in modern Tamil, but it is common in traditional Tamil. In addition to this there are some characters which are not native to Tamil but became integral part of modern Tamil language. Those characters came from Grantha script i.e. the word borrowed from Sanskrit and other languages.

4. Speech Corpus Development

The first step in speech recognition is to collect speech data for training the system, for English this data was provided by Carnegie Mellon University (CMU) called TIMIT corpus which are all phonetically rich sentences and captures all types of co- articulations. This type of speech data is not available for most of the languages in the world, in the countries like India there are more than 300 spoken languages with more than 20 official languages in it. Most of the Indian language lack speech data, at the same time more works have been done for the development of speech database for Indian languages Central Institute of Indian Language (CIIL) Mysore, is providing corpus for the research purpose. Also in ASR the reduction in the domain of the application will increase the performance. In our paper we have chosen a farmer application for which the sentences are collected by from the horticulture and news portals and also some sentences which are generated using the tool.

4.1 Font Converters

As mentioned earlier our methodology is being dealt with Unicode characters, to ease the analysis and work with the text corpus, it is required that the text is in a representation that can be easily processed.. So, font converters have to be made to convert the corpus into the desired electronic character code as well for the other text processing methods. The font will be reflective of the

Received: 25 Oct 2018/Accepted: 25 Nov 2018

language and will have as many unique codes for each of the characters in the language. It is necessary to ensure that there are no mapping and conversion errors.

5. Proposed Methodology

We propose a methodology that utilizes the constant vowel cluster method. This consonant vowels cluster method is successfully implemented for the Romanized Tamil text. Our proposed methodology is unique by the way the used text used for recognition as it all composed of the Unicode characters, here the dictionary contains the syllable transcription in the Tamil. In previous method the transcription will be in the Romanized text for Tamil.

This proposed methodology is achieved through the following way. Since we use Unicode in the experiment, it is possible to convert a Unicode in the experiment it is possible to convert a Unicode to ISCII and ISCII to ITRANS. We apply consonant vowel cluster methodology for syllabification using linguistic rules of the language. After syllabication the Romanized Tamil is brought back to Unicode.

The problems we face here is the tool we use will not support the Unicode instead it will generate the octal equivalent of the characters which are in Unicode and we have to apply a mapping mechanism for Unicode to octal via hexadecimal values of the corresponding characters.

5.1 Experimental setup

The following diagram will provide a brief explanation of the speech recognition. It gets the speech signal as input whose features are extracted then it is decoded using the acoustic model, dictionary for pronunciation and model for language. The feature extraction is process of conversion of waveform to some type of parametric representation. The acoustic model, is one which is created by taking audio recordings of speech and their transcriptions and 'compiling' them into a statistical representations of the sounds that make up each word (through a process called 'training'), which should be supported by language model. A language model is a input file contains the probabilities of possible sequences of words. A grammar is another file with contains combination of words which is predefined already. The models of language are used of dictating applications, but the grammar is used for desktop commands. The pronunciation dictionary is one which contains the required transcription i.e., it may contain phoneme or syllable transcription of all words that are expected to speak.

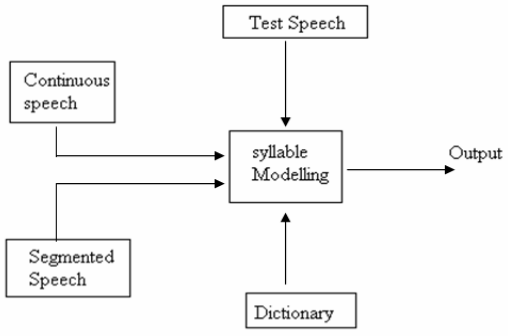

Fig. 1: Block diagram of syllable modeled recognition system

6. Comparison of Performance

It is discovered that word error count of this syllable model have is being impressively diminished and contrasted with word demonstrate. It is seen that in this syllable models there is more substitution blunders when compared to additions and erasures while on account of word models, there is a lion's

Received: 25 Oct 2018/Accepted: 25 Nov 2018

share of cancellation mistakes. Greater part of erasure mistakes in word model connote OOV rate because of structure emphases and represented in syllabic model. This demonstrates the way that syllables are successful as sub-word units. Notwithstanding, contrasted with triphonic models. The expansion in WER can be credited to the extensive number of syllables to be demonstrated with the accessible constrained preparing set. This likewise shows the nearness of a small logical impact in between the syllables.

7. Conclusion

The proposed framework demonstrates the working model of discourse recognizer with syllable unit. An enhancement to syllable model can be made to this displaying setting between syllables units, yet this may prompts a huge number of new inputs that should be demonstrated. In future the analysis and research can be taken towards the following path.

References

[1] L.R. Bahl, R. Bakis, P.S. Cohen, A.G. Cole, F. Jelinek, B.L. Lewis, R.L. Mercer, Further results on the recognition of a continuously read natural corpus, IEEE International Conference on acoustics, speech, signal processing, 1980.
[2] L.R. Bahl, P.F. Brown, P.V. De Souza, R.L. Mercer, Acoustic Markov models used in the Tangora speech recognition system, IEEE international conference on acoustics, speech, signal processing, 1988.
[3] CIIL, Central Institute of Indian Languages, Mysore, India.
[4] D. Jurafsky, Speech and language recognition, Pearson Education, 2005.
[5] O. Fujimura, Syllable as a unit of speech recognition, IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP, 23(1), 1975, 82–87.
[6] M.Y. Hwang, X.D. Huang, Shared distribution hidden Markov models for speech recognition, IEEE Transactions on Speech and Audio Processing, 1(4), 1993, 414–420.
[7] A.N. Khan, B. Yegnanarayana, Development of speech recognition system for Tamil for small restricted task, Proceedings of national conference on communication, 2001.
[8] A. Lakshmi, A.M. Hema, A syllable based continuous speech recognizer for Tamil, Interspeech, 2006, 1878–1881.
[9] K.F. Lee, Context dependent phonetic Markov models for speaker independent continuous speech recognition, IEEE Transactions on Acoustics, Speech and Signal Processing, 38(4), 1990, 599–609.
[10] T. Nagarajan, V. Kamakshi Prasad, A.M. Hema, The minimum phase signal derived from the magnitude spectrum and its applications to speech segmentation. In Sixth biennial conference of signal processing and communications, 2001.
[11] T. Nagarajan, A. Hema Murthy, R.M. Hegde, Segmentation of speech into syllable-like units, In Eighth European Conference on Speech Communication and Technology, 2003, 2893–2896.
[12] B. Paul Douglas, E.A. Martin, Speaker stress-resistant continuous speech recognition, In International Conference on Acoustics, Speech, and Signal Processing, 1988, 283-286.
[13] L. Rabiner, Fundamentals of speech recognition, Pearson Education, 2006.
[14] S. Saraswathi, T.V. Geetha, Lecture notes in computer science: Implementation of Tamil speech recognition system using neural networks, 2004.
[15] S. Rich, Y. Chow, S. Roucos, M. Krasner, J. Makhoul, Improved hidden Markov modeling of phonemes for continuous speech recognition, IEEE International Conference on Acoustics, Speech, and Signal Processing, 9, 1984, 21-24.

Received: 25 Oct 2018/Accepted: 25 Nov 2018