A Min-Max Syllable Compaction Method for Tamil Text-To-Speech

Elanchezhiyan Karuppasamy | **Tamil Selvi Elango** | **Suriya Murugesan** | **Karthikeyan Sethuraman** | **Madhan Karky Vairamuthu**

Karky Research Foundation,
Chennai.

chezhiyank@gmail.com | etamilselvi_cs@yahoo.co.in | suriyah.cse@gmail.com | sethuramankarthikeyan@gmail.com | madhankarky@gmail.com

Abstract

Tamil language is rich in syllables. Tamil language has 12 vowels, 18 consonants and 216 consonant vowels and 6 borrowed consonants. We identified over 1.9 lacks syllables in the language that could be used to form words. We compare the possible set of syllables against the usage statistics of these syllables. This paper aims at proposing a min-max method to compact the number of syllables for a text-to-speech system. The method aims at minimizing the number of syllables and maximizing the coverage of words. The paper also proposes compacting the syllable set by applying a transformation rule. The combined approach identifies 1997 syllables to cover 98.2% of words in Tamil language.

1 Introduction

The Text-to-speech synthesizers open better avenues of communication for the differently abled. Along with speech recognition, they are supposed to make Human Computer Interaction livelier. They can be employed in automatic announcement systems, entertainment, animation industries, etc.

Present text-to-speech systems rely on breaking the text to phonemes, or syllables, or morphemes, applying prosodic rules to the same and picking the corresponding sounds from a sound database and concatenating them to synthesize speech [8][9]. The text to speech system in this work is also approached in the same manner. Deciding the number of phonemes is seen as a min-max problem – minimizing the number of phonemes in the system and maximizing the number of words covered.

There are 18 consonants, 12 vowels and 216 consonant vowels. Phonetising Tamil words is a challenge since the pronunciation of words does not only depend on the constituent graphemes but also on factors such as place of occurrence of the grapheme, the preceding and succeeding graphemes, etc.

This paper is organized as follows. The first section talks about the background of some of the text-to-speech conversion systems introduced so far. The next section gives a view on Tamil phonology and syllables. The third section discusses the four methodologies proposed in this paper to decide the number of phonemes to be included in the system. Results of implementation are discussed in the section which follows. The next section concludes the paper with scope for future work. The references are included in the last section.

2 Background Study

The objective of text-to-speech system is to produce speech waveform from the input text. In an efficient TTS system, the output speech waveform should be indistinguishable from human speech.

"Festival Based Maiden TTS System for Tamil Language" (Sreekanth.M & A.G.Ramakrishnan, 2007) uses labeled database where the text in speech corpus is manually segmented and labeled to make a system capable of producing better speech quality.

"Text Processing for Text-to-Speech Systems in Indian Languages" (Arokia Raj, et al., 2007) proposed Akshara mapping and discussed the issues in font Akshara mapping. Indian languages are syllabic in nature. Akshara mapping defines the structure of syllables in a word which is a composition of vowel and consonants, sequence of consonants.

"Natural Sounding TTS Based On Syllable-like unit" (Samuel Thomas, et al., 2006) employs group delay segmentation algorithm for automatic generation and segmentation of syllable-like units in word.

"Experiments with Unit Selection Speech Databases for Indian Languages" (Kishore et al., 2003) concludes that in Indian languages the syllable units perform better than the other speech units which are di-phone, phone and half-phone units for text to speech synthesis.

"Syllable Based Indian Language Text to Speech System" (Sanghamitra Mohanty, 2011). proposed utterance algorithm. The algorithm helps to distinguish the different type of utterance of same character based on its position in the word.

The problem encountered in processing Tamil language for speech synthesis is discussed in "Morpheme Based Language Model for Tamil Speech Recognition System" (Selvarajan Saraswathi & Thekkumpurath Geetha, 2007). This paper illustrates that the Morphological pattern of Tamil is more complex than English. In addition there are results that shows word formation and word ordering are the major differences between these two languages.

Concerns in developing Tamil TTS include understanding Tamil phonetics, database creation of Tamil language, syllable level concatenation, complexity of the language etc are discussed in "Dravidian – Tamil TTS For Interactive Voice Response System" (S.Jothilakshmi, et al., 2013).

3 Tamil Phonology

Tamil is a very prominent member of the South Indian family of languages. Tamil language is composed of 12 vowels called as Uyir and 18 consonants as Mey. There are set of 216 graphemes termed as Uyirmey that forms as the combination of vowels and consonants. It also supports a special character called Aayutha Ezhutthu. In addition to these, we have six letters called as Grantha letters which are derived from Sanskrit.

A single character in tamil script can produce multiple sounds based on its context. For example, the letter ka(க) is pronounced as ka in karupu(கருப்பு), ga in magan(மகன்) and ha in pahal(பகல்).

Syllables play an important role in the study of speech and in phonetics. A syllable is made up of either single vowel sound or combination of vowels and consonants. The vowels generally accompany a consonant but they do appear independently when used at the beginning of a word.

4 Methodologies

In this section we propose two simple methods to identify possible syllables in Tamil language. Syllables can be extracted from words using simple v, vc, c_v, c_vc, vcc and c_vcc patterns, where v denotes a vowel, c denotes a consonant, c_v denotes a consonant vowel.

4.1 Brute Force Method

We find all possible syllables in Tamil Language using a brute force approach by combining the characters based on the patterns defined above. By brute force, we have 12 Vowels, 336 Consonant Vowels, 276 Vowel Consonants, 8004 cvc, 6348 vcc and 1,77,744 cvcc. Totally, we have 1,92,720 possible syllables in Tamil language.

4.2 Statistical Method

In this method, we extract syllables from the words used in Tamil language. 2.5 lacks Tamil root words from a dictionary and unique words with morphological inflections from blogs and news articles were used to extract the syllables. The count of number of times a syllable occurs across the words is also recorded. As most of the syllables identified using the brute force method are not used in the language. Also, most syllables occurred just once in a word. We identified 6058 syllables in the statistical method, in which 12 out of 12 Vowels, 295 out of 336 cv, 201 out of 276 vc, 3743 out of 8004 cvc and 2242 out of 184092 vcc + cvcc were identified.

4.3 Min-Max Method

In this method we aim to minimize the number of syllables and to maximize the coverage. Let $W = \{w1, w2 \ldots wm\}$ denote the set of all words in the language. Let $S = \{<s1, c1>, <s2, c2>, \ldots <sn, cn>\}$ denote the set of syllables identified using the statistical method described in section 4.2.

Let $Sc \subseteq S$ denote the set of compacted syllables. Sc is initially set to an empty set. Let $Wc \subseteq W$ denote a set of words that can be formed using the syllables in Sc

Our objective is to take elements from S and add to Sc to satisfy

$$\mathrm{Min}|Sc| \rightarrow \mathrm{Max}\ |Wc|$$

Which translates to finding optimal ∂, a constant such that

$$\forall i = 1 \ldots n\ ei<si,ci> \in S\ ,\ ei \in Sc \text{ if } ci > \partial$$

There by finding the optimal ∂ to $\mathrm{Min}|Sc| \rightarrow \mathrm{Max}\ |Wc|$.

4.4 Compaction and Transformations

Further to identifying the ∂ value, Sc can further be compacted by reducing redundant sounds. In Tamil, multiple characters carry the same sound associated with them. For instance, நான்(naan) and ஓணான்(oa-naan) both have the naan syllable. மாற்றம்(maat-Ram) and மாட்டு (maat-tu) both have maat has the common syllable even though the letters vary. In the second example, the similarity happened owing to the sound transformation process rather than the similarity of sounds of individual characters.

By removing such redundancies we would be able to achieve the 96.76% coverage with $|Sc| = 1997$. Further there is also scope to cover the remaining 3.23% words by applying some simple transformations such as replacing closest syllable. For instance, சர்ச்சில் (church+chil) has a cvcc pattern in church along with a vc pattern il. To cover this word, we would need சர்ச்(church) syllable in Sc. Instead, by applying a simple transformation such as removing the end ch in church, we get chur+chil, which will still give a sounding close to the original. The rule for transformation is as follows:

Transformation Rule: If a cvcc pattern ends with a hard consonant (k,ch,t,th,p,R) and the pattern following begins with a hard consonant, then the ending hard consonant in the cvcc pattern can be removed.

5 Results

In this section, we analyze the distribution of syllables across the patterns mentioned in section 4. Figure 1 and 2, compares the distribution of syllables in brute-force and statistical methods. We find that 44.72% of syllables, which were identified in brute-force, are not present in the statistical analysis. The $c_v c$ pattern takes the bigger share of the statistical method distribution.

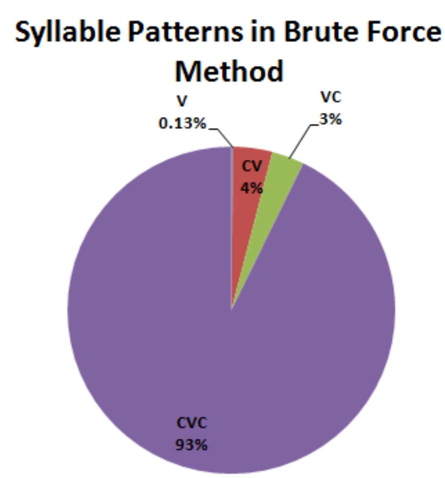

Fig 1: Syllable distribution across different syllable patterns in bruteforce.

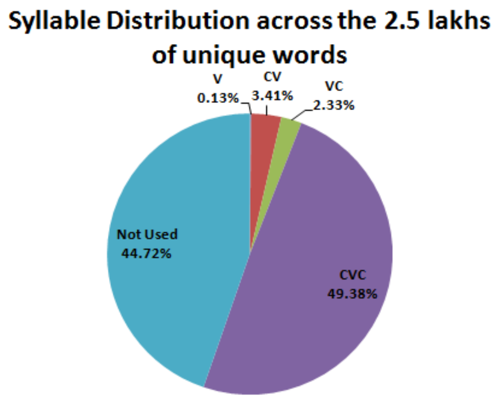

Fig 2: Syllable distribution across different syllable patterns in statistical method.

With a $\partial = 9$ we would be able to cover 96.76% words in W with $|S^c| = 2111$.

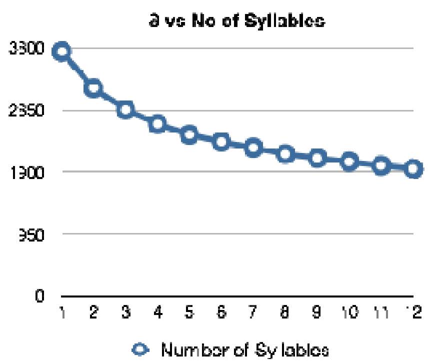

Fig 3: Syllable count.

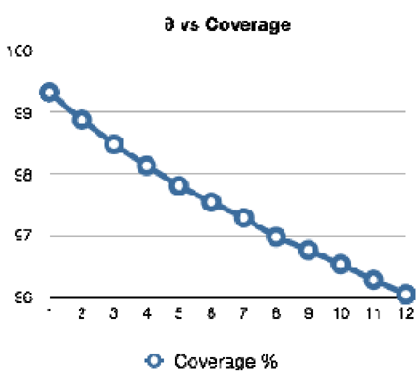

Fig 4: coverage against ∂

Additionally, with the transformation rule in effect, we can maximize $|W^c|$ without affecting $|S^c|$. Applying the rule we get 98.23% coverage.

6 Conclusion

In this paper, we presented two methods to analyze the syllables in Tamil language. With brute-force method, we listed all possible syllables. 1.92 lakhs syllables were identified via this method. We analyzed 2.5 lakhs Tamil root words in a Tamil dictionary and also words with morphological inflections from various Tamil news portals. Further, we extracted syllables and their frequency of occurrence. We were able to reduce the syllables necessary for an effective text-to-speech system to 6058 syllables. Further we proposed a min-max method to minimize the number of syllables needed and to maximize the coverage. With the min-max method and a simple transformation rule, we were able to reduce the number of syllables to 1997 and get coverage of 98.23%. We believe that studying the popularity of words would also help us to get an even better coverage in the future.

References

Sreekanth.M and A.G.Ramakrishnan. 2007. *Festival based maiden TTS system for Tamil language*, 3rd Language and Technology Conference, pp. 187-191.

Anand Arokia Raj, Tanuja Sarkar, Satish Chandra Pammi, Santhosh Yuvaraj , Mohit Bansal , Kishore Prahallad, Alan W Black. 2007. *Text Processing for Text-to-Speech Systems in Indian Languages*, 6th ISCA Speech Synthesis Workshop.

Samuel Thomas, M. Nageshwara Rao, Hema A. Murthy, C. S. Ramalingam. 2006. *Natural Sounding TTS based on Syllable-like Units*, 14th European Signal Processing Conference.

Kishore S. Prahallad, Alan W Black, Rohit Kumar and Rajeev Sangal. 2003. *Experiments with Unit Selection Speech Databases for Indian Languages*, National Seminar on Language Technology Tools: Implementations of Telugu, Hyderabad, India.

Sanghamitra Mohanty. 2011. *Syllable Based Indian Language Text to Speech System*, International Journal of Advances in Engineering & Technology.

Selvarajan Saraswathi and Thekkumpurath Geetha. 2007. *Morpheme Based Language Model for Tamil Speech Recognition System*, International Arab Journal of Information Technology, Vol. 4, No.3.

S.Jothilakshmi, S. Sindhuja,V. Ramilingam. 2013. *Dravidian – Tamil TTS for Interactive Voice Response System*. Internation Journal of Interactive Voice Response System, Vol 2.