Computational Phonology and the Development of Text-to-Speech Application for Tamil

Vasu Renganathan

University of Pennsylvania (vasur@sas.upenn.edu)

Phonology is a study of sound systems specific to any particular language, whereas Phonetics is a study of linguistic sounds in general from a universal point of view without having any specific reference to the sounds of any particular language. In this sense the standard system of phonetic notations as devised in International Phonetic Alphabet (IPA) is an attempt to identify almost all of the universally valid linguistic sounds using a set of standard symbols of Latin alphabet[1]. Not all of the linguistic sounds that are accounted for within IPA system can be found to occur in any single language, but almost all of the linguistic sounds that one can find in any given language can be related to one of the sounds as described within IPA system of linguistic sounds. In this sense, *Computational Phonology* is a study of speech sounds of any given language from the point of view of how they are mapped using computational algorithms. For some, the field of Phonology itself is considered deeply to be computational as it involves a complex algorithm of determining the environmental factors in any speech (see Bird, 1995). This paper, in this sense, is an attempt to summarize the relationtionship between graphemes, phonems and allophones of Tamil in the context of making a Tamil text-to-speech application[2].

Many-to-one relatinship between Linguistic sounds and Orthographic symbols

Not all linguistic sounds that one can hear and produce in any particular language can be visually represented by unique orthographic symbols and there can be instances where more than one sound is represented by single orthographic symbol. Learning to read and pronounce a wide variety of sounds correctly using any such limited orthographic symbols commonly becomes part of the native speakers' linguistic intuition and language skill in the respective language. This is true with Tamil as well in the sense that the sounds and orthography of this language

[1] Cf. http://www.langsci.ucl.ac.uk/ipa/

[2] This Tamil text-to-speech application can be tested online at: http://www.thetamillanguage.com/tamilnlp/speak/speak.html.

have a many-to-one relationship and accounting for those sounds with missing symbols needs to be either represented in diacritics or special ASCII characters in the context of computer processing of the language. To cite one example, the orthographic symbol க 'ka' is used in Tamil to represent the sounds ஹ 'ha', ங 'nga' and க 'ka'. It is to be noted that the sound 'k' under this circumstance is called a phoneme in Tamil and the corresponding contextually realized sounds that do not exhibit any minimal pairs, such as 'h' and 'ng', are called allophones or positional variants. In general, one can conclude that all of the phonemes can have a grapheme, but not all the allophones. Identifying such a many-to-one relationship as well as mapping them to a one-to-one correspondence of symbols and sounds using a common transliteration scheme, as well as the corresponding audio files is pretty much the main task of the field of *Computational Phonology* in the context of making text-to-speech application. This paper is an attempt to describe one of the working version of text-to-speech application for Tamil written in the programming language PHP and is made available on the web at the address: http://www.thetamillanguage.com/tamilnlp/speak/speak.html.

Syllabification of Tamil words

The foremost step one might want to take in the context of making text-to-speech application is to identify the number of syllables that can be distinguished in an unique manner. Tamil syllables, in general, can be conveniently grouped under the basic categories of V (ஆ 'cow', ஈ 'fly' etc.), CV (தா 'give', கா 'protect' etc.), and CVC (கல் 'stone', பல் 'tooth' etc.) types, along with a host of other complex forms constituting one or more of the combinations of these basic structures. Some of those examples include, CVC.CVC (கப்பல் 'ship'), VC.VC (அவன் 'he'), CVC.CV (தம்பி 'younger brother') and so on. Syllabification of Tamil words is especially essential for both morphological as well as phonological applications for the fact that identification of words and affixes constitute the significant part of any part-of-speech taggers as well as phonological applications. For a detailed account of how the process of syllabification is significant in the context of morphological taggers, see Renganathan (2001) and Rajendran et al. (2001).

Tamil text-to-speech Application and the Phonological Rules of Tamil

Although there are many phonological rules that can be used to account for the difference between written literary and spoken varieties of Tamil for the purposes of computer processing of Tamil language in the context of text-to-speech application and voice recognition, only a specific set of rules that are normally called automatic rules, such as inter vocalic devoicing rule, nasal assimilation, enunciative rule etc., are required to be accounted for, and others such as palatalization rule, v-deletion rule etc., become irrelevant[3]. To cite a few examples, words such as மகன் 'son' (mahan), தங்கம் 'gold' (tangam), பூனை 'cat' (pūnɛ), எடு 'take' (eḍʉ) etc., to be pronounced correctly, one needs to account for the phonological rules namely devoicing/fricative rule, nasal assimilation/devoicing rule, and vowel reduction rules respectively. What is significant in this context is to identify all of the possible distinct syllables and mark them with special symbols on a one-to-one fashion based on the relevant phonological rule in order to match them appropriately with words. As Tamil alphabet exhibits many-to-one correspondence between Tamil letters and speech sounds, one needs to carefully identify all of the positional variants of all the phonemically explainable linguistic sounds and subsequently map them with that of the appropriate allophones with distinct symbols. What follows is a description as to how Tamil alphabet is realized with multiple allophonic sounds in various linguistic environments. As one can see, only the instances of Tamil stop sounds undergo this type of variations and all the other consonants are realized in surface with single speech sound. As for vowels, only two vowels in Tamil namely உ (u) and எ (e) are realized with multiple tokens of sounds. The vowels உ and எ are realized as enunciative vowels ʉ and ɛ respectively mostly in word final positions. Thus, what seems to be significant in the context of writing algorithms for Tamil text-to-speech application is that one needs to account for all of such phonological environments and correspondingly transliterate them with special symbols to accomplish a one-to-one structure of symboles of Tamil text. So, each of these syllables can then be related to corresponding sound files.

Voiceless and Voiced Sounds of Tamil Stops and their Surface Realizations:

[3] For a detailed description of Tamil phonological rules and their variations in the context of Tamil dialects, see Christdas (1988), Renganathan (2010) and Schiffman (1999).

க -> k, h, ng: k is realized as க in word initial position and in clusters. க is realized as ஹ intervocalically and after *l̲, r* and *l*. க is realized as *g* after nasals. ச -> c, s, j: ச is realized as *c* in word initial position and in clusters. ச is realized as *s* intervocalically and optionally in word initial positions. ச is realized as *j* after nasals. ட -> ṭ, ḍ: ட is realized as ṭ in word initial position and in clusters. ட is realized as ḍ intervocalically and after nasals. த -> t, d: த is realized as *t* in word initial position and in clusters. த is realized as *d* intervocallically and after nasals. ப -> p, b: ப is realized as *p* and in word initial position and in clusters. ப is realized as *b* intervocalically and after nasals.

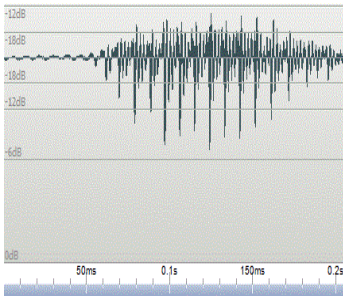
One of the challenging tasks in any text-to-speech system is making speech files which can sound as close to human's speech as possible. Although Tamil does not exhibit any stress as part of any word and it is not phonemic in this language, the varying expressions and multiple shades of meaning can be captured by way of varying intonations of Tamil sentences. Such intonations are usually distinguished based on the length of long vowels that might occur within in any word. To cite one example, the question வந்தானா? *vantāṉā*? 'did he come?' can be a simple question of asking the factual meaning whether a person came, or it can also be an expression of one's surprise or discontent about someone's arrival. The latter types of expressions are usually distinguished based on how the long vowels of the last two syllables are alternatively pronounced in varying degrees of pitch. Accounting for long vowels, thus, should constitute the significant portion of reasonable Tamil text-to-speech application.

Making Audio files for Text-to-Speech Application

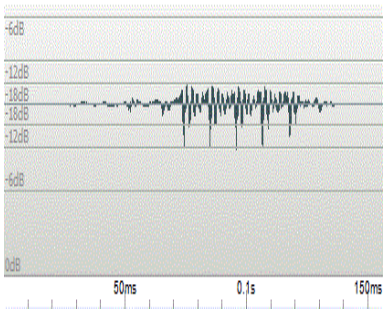
Special care needs to be taken as to how the sound units corresponding to each syllable in Tamil is to be constructed. Importantly, the sounds that are heard or uttered in isolation vary considerably when they occur with other sounds as part of a word. In this respect, what is the most efficient way of making audio files for each syllable is to eliminate much of the contents of coda, such as with clusters of multiple types, syllables with short and long vowels and so on. Consider below the reduction of duration of sound between individual vowels and their occurrences in the corresponding syllables. The vowel *e,* for example, takes 200 milli seconds where

as the syllable *be* takes only 150 milli seconds. What it implies is that it may not be advisable for one to concatenate sound files of pure consonants along with vowels to make corresponding syllables. Instead, one needs to create sound files of all possible distinct syllables of various types in isolation and concatenate them to form words.

e

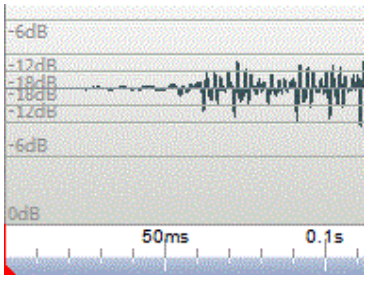


be

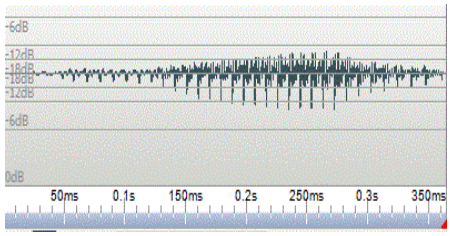


Further, consider below how the long vowel *ō* and the corresponding syllables *nō* and hō are distinguished in terms of their distinct duration. As far as the syllables with long vowels are concerned, one can construct different durations of length and it can not be restricted to any single duration. In fact, as already mentioned, this varying length of duration in syllables with long vowels is one of the significant factors that determine the nature of speech overall, as far as Tamil speech is concerned.

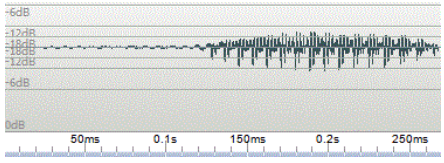
ō



nō



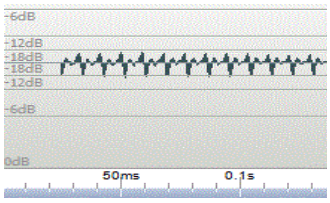
hō



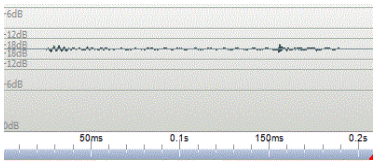
Making sound files for pure consonants:

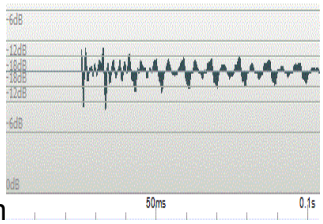
When pure consonants such as k, ng, c etc., are produced, one would normally accompany a vowel, namely 'i' as in 'ik', 'ing', 'ic' respectively, and it would be hard for anyone to produce them without this epithetic vowel. What matters in the context of making pure consonants is that such epithetic vowel part needs to be carefully removed in order to identify the acceptable pure consonantal sound. As a result, one would find pure consonants in their corresponding wave formats to be with very subtle pitch in them as shown below for the consonants 'n', 'k' and 'm'.

n

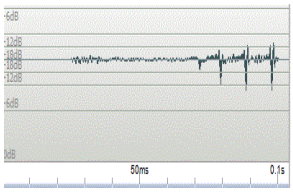


k





m



ka

In conclusion, for any naturally sounding Tamil text-to-speech application, one needs to take into consideration as how many distinct syllables would need to be made based on the facts of vowel duration of long vowels and use of pause in the beginning and at the end of each syllable. Thus, the total number of Tamil syllables to be constructed for the purpose of making any Tamil text-to-speech application is determined not only based on the basic linguistic phonological environments, it also needs to be extended further to the nature of vowel duration of long vowels as well as the nature of pause that is to be applied for each syllable. Especially, the length of long vowels for questions is to be distinguished from other intonational factors such as exclamation, expressing one's discontent and so on. Similarly, appropriate length of pause is needed for making words with a sequence of single syllable words, but not in any combination of consonantal clusters involving both identical

consonants as well as those with liquids. Further the duration of pause between words also determines the nature of speech in terms of how it is produced, either with continuous or discontinuous stream of speech.

References:

Bird, Steven (1995). Computational Phonology: A Constraint-Based Approach. Studies in Natural Language Processing. Cambridge University Press.

Christdas, Prathima (1988). *The Phonology and Morphology of Tamil.* Unpublished Ph.D. dissertation, Cornell University.

Rajendran S, Viswanathan S and Ramesh Kumar (2001). "Computational morphology of verbal complex", *Language in India*, www.languageinindia.com, Vol. 3:4.

Renganathan, Vasu (2001). "Development of Part-of-Speech Tagger for Tamil", Tamil Internet Conference, 2001 (http://infitt.org/drupal7a/TIconferencepapers/TI2001/vasur.pdf).

Renganathan, Vasu (2002). "An interactive approach to development of English to Tamil Translation machine Translation system on the web", Tamil Internet Conference, 2002. (http://infitt.org/drupal7a/TIconferencepapers/TI2002/15VASUR.PDF)

Renganathan, Vasu. (2010). *Tamil Language in Context: A Comprehensive Approach to Learning Tamil,* Department of South Asia Studies, University of Pennsylvania, http://www.thetamillangauge.com.

Renganathan, Vasu. (forthcoming). *Computational Approaches in Tamil Linguistics.* Unpublished monograph.

Schiffman, Harold (2006). *A Reference Grammar of Spoken Tamil.* Cambridge University Press, Cambridge.