GRAPHEME TO PHONEME CONVERSION FOR TAMIL SPEECH SYNTHESIS

A. G. Ramakrishnan and Laxmi Narayana. M

Department of Electrical Engineering, Indian Institute of Science, Bangalore 560 012.

ABSTRACT

The input to a TTS system may contain acronyms, abbreviations and non-standard words, which must first be converted to the corresponding Tamil graphemic form. A text normalization module is developed for accomplishing this task. We have developed a quality grapheme to phoneme (G2P) conversion module, which converts the normalized orthographic text input into its phonetic form. The phonetic transcription helps in identifying and analysing the basic units such as mono-phones, diphones and syllables and also for segmenting the speech corpus. A character to phoneme mapping interface is developed to map the Tamil graphemic text to the corresponding phonetic representation in Roman script. A Rule base is created which contains the inter and intra word rules for changing the default character phone mapping wherever necessary. A proper noun lexicon as well as foreign word lexicon is also incorporated for dealing cases where G2P fails. The G2P module designed is to be used for MILE Tamil TTS synthesizer in both Windows platform and Festival (Linux) environment.

Index Terms— G2P, Tamil TTS, normalization, lexicon

1. MOTIVATION

Over 65 million people worldwide speak Tamil, the official language of the south Indian state of Tamil Nadu, and also of Singapore, Sri Lanka and Mauritius. In addition to the above countries, it is spoken in Bahrain, Malaysia, Qatar, Thailand, United Arab Emirates and United Kingdom. Tamil is a syllabic language which contains 12 vowels and 18 consonants. There are 5 other phones introduced for representing some consonants of Sanskrit. The language has well defined rules, which introduce seven other phones based upon the relative positions of consonants with respect to vowels or other consonants. Hence there are 42 phones in the language.

Text to Speech (TTS) synthesis is an automated encoding process, which converts a sequence of symbols (text) conveying linguistic information, into an acoustic waveform (speech). The two major components of a TTS synthesizer are - Natural Language Processing Module (NLP), which is capable of producing a phonetic transcription of the given text and Digital Signal Processing module (DSP), which transforms this phonetic transcription into speech [2]. One of the characteristics, based on which a TTS system is evaluated, is its intelligibility i.e., its ability to accurately synthesize the input text before naturalness and expression. The G2P module is responsible for the determination of the phonetic transcription of the incoming text. This involves normalizing the input text and mapping the graphemic representation to a corresponding phonetic representation. Since the orthographic representation and pronunciation do not match in some cases, the default mapping needs to be changed wherever necessary. Section 2 describes the need for Text Normalization and how the non-standard words like acronyms and abbreviations in the input text are dealt with in NLP. Section 3 describes the process of converting a pure word sequence into its phonetic equivalent, the inter and intra word rules which are made use of, for G2P conversion, and the creation of Foreign word lexicon. The results are presented in section 4.

2. NORMALIZATION OF NON-STANDARD WORDS

The text input to the TTS system may not be pure Tamil text. It may contain some non-standard words like acronyms, abbreviations, proper names derived from other languages or clutters, phone numbers, decimal numbers, fractions, ordinary numbers, sequence of numbers, money, dates, measures, titles, times and symbols. The Natural Language Processing module should be able to handle such non-standard words. Standard words are those, whose pronunciation can be obtained from the G2P rules. A G2P converter maps a word to a sequence of phones. All the non-standard words must be expanded into the corresponding Tamil graphemic form, before sending to the G2P module for phonetic expansion. This module should also take a decision of how a non-standard word is being pronounced. For example, a phone number must be read with each digit treated as a single number and read in isolation.

The corresponding Tamil graphemic representations of possible non-standard words, English words and Tamil short forms are written in ‘iLEAP’ format. iLEAP is a software, where one could type in many Indian languages. The ISCII (Indian Script Code for Information Interchange) file is exported as an ASCII (text) file and this file is used by the Text normalization module. The format of the text normalization file is shown in Figure 2.

The input text file is searched for abbreviations or acronyms (can be in Tamil or English). They are replaced by the corresponding expansion (graphemic form) in Tamil in the output (ISCII) file. This is illustrated in Figure 1.

Ex: *aug* is replaced as ஆகஸ்ட் (normalization)
august is also replaced as ஆகஸ்ட்
(Tamil transcription)

Further, there are a number of words, used regularly in Tamil, which are originally Sanskrit or English words. The G2P fails to give the accurate phonetic transcription in case of such words. Hence, we have created a lexicon of foreign words.

Expansion of the numbers is a 'special' case in normalization, because a *decision* needs to be taken whether the encountered number is an ordinary number, a phone number, date, time or currency. If it is currency, it decides whether it is rupee or dollar or pound or yen. The input number is considered as a string. An *ordinary number* is expanded according to the length of the string. A module is written to expand a 3 digit number. The control chooses different paths according to the length of the string. If the number is 4 or 5 digits long, then it must be in thousands or ten thousands. In this case, this module is called twice, first to convert the number of thousands to words (1 or 2 digits) and next to convert the remaining 3 digit number to words. For example, if the number is 12345, in the first call, the number 12 is processed and in the next call, 345 is processed. If it is a 6 or 7 digit number, it must be in lakhs or 10 lakhs and if the number has 8 or 9 digits, then it must be in crores or 10 crores; then this module is called thrice or four times, respectively and so on. If the number of digits is less than or equal to 3, then the module is called only once.

If the number string is not an ordinary number, a parameter (a number corresponding to the decision taken) is set according to the type of the number string. If the number string is a decimal number (Ex: 23.8756) the number before the dot (.) is treated as one number and the digits after the dot are spoken in isolation. If the number string is a *date*, the delimiters can be '/' or '-' (Ex: 25-10-1999 or 25/10/1999). All the three values (date, month, and year) are extracted from the input string and processed separately. Similarly the different types of number strings namely, currency, range of numbers, arithmetic, phone numbers and time are identified by the delimiters present and expanded accordingly.

Example:

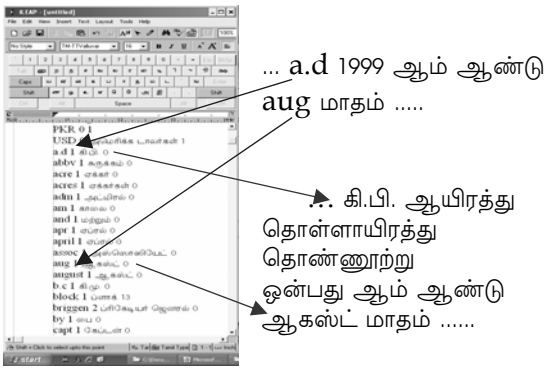


Figure 1: Text Normalization using look up table

ph.no	தொலைபேசி எண்
jan	ஜனவரி
rs	ரூபாய்கள்
மி.மீ	மில்லி மீட்டர்

Figure 2: Format of Normalization file

3. DESING OF THE G2P MODULE

In a TTS system, the G2P module converts the normalized orthographic text input into the underlying linguistic and phonetic representation. G2P conversion therefore is the fundamental step in a TTS system [1]. The text normalization module inputs a word sequence to the G2P module. The Grapheme to Phoneme conversion of the word sequence can be done using the Letter to Sound Rules. The rules are based on a pronunciation dictionary, in which a mapping of the spelling of a word into a sequence of phones can be found. For example, consider an English word – "speech". The pronunciation dictionary converts this word to the phone sequence – S P IY CH[1]. Traditional orthography in some languages, particularly French and English, often does not coincide with pronunciation. However, in other languages such as Spanish and Italian, there is a consistent relationship between orthography and pronunciation (Link).

If there is no pronunciation dictionary, a simple set of rules to convert the graphemic form of a word into the corresponding phonemic form is used. Using such rules is more relevant for Indian languages. In many cases, there is a direct correspondence between what is written and what is spoken. For example, consider a Tamil word - "asiriyar", the corresponding phonetic transcription is - /A/ /s/ /i/ /r/ /i/ /y/ /a/ /r/ which is very similar to the word.

3.1. Mapping from Tamil Graphemes to Phonemes

The grapheme to phoneme (G2P) conversion module receives the sequence of Tamil words, from the Text normalization module, which then are converted to a phonetic transcription represented in Roman script. This is obtained by using a *character to phone mapping,* which gives the corresponding phonemic representation of Tamil graphemes, in Roman script. The Roman character (or sometimes, combination of letters in case of diphthongs like /ae/ or /au/ or genitives like /kk/ and /tt/) which represents a Tamil phonemic unit may not 'sound' exactly like the Tamil phoneme and is used only as an unambiguous representation. The *.wav* (speech) files in the inventory are labeled according to this Roman representation. The DSP module in the TTS engine picks up the corresponding speech units (which are labeled according to the mapping), given by the phonemic representation, for concatenation.

Words are converted one by one. Two kinds of rules – inter-word rules and intra-word rules are applied to the input text during conversion. If the last character in a word is *halanth*, and the last but one character is /k/ or /ch/ or /th/ or /p/ and the next word starts with /k/ or /ch/ or /th/ or /p/, respectively, the two words are concatenated. This inter word rule is very relevant here because, while speaking such words, a speaker does not pronounce the phoneme /k/ or /ch/ or /th/ or /p/ two times. Such junctions of words will combine into a single word in which those two phonemes at the junction would manifest as a genitive, for example /p/ /p/ becomes /pp/. The double

letters are labeled with a suffix 'l' to the basic phoneme. This is illustrated in Figure 3(a) and 3(b). It was found that the duration of a double letter occurring in the middle of a word and that which was formed at the junction of two words is comparable.

Tamil uses a syllabic script that is largely phonetic in nature. Thus, in most cases, there is a one-to-one mapping between graphemes and the corresponding phones. The architecture of the Natural language Processing module is shown in Figure 4. Language specific information is fed into the system in the form of mapping and rules. The default character to phone mapping is defined in the mapping file. The format of the mapping is shown below and explained subsequently.

Format: *Character Type Class Phoneme*

Example: அ V VOW a ⟶ ¤ V VOW a

Character: The orthographic representation of the character.

Type: Three types of characters are identified, C-Consonant, V-Vowel and H-Halanth

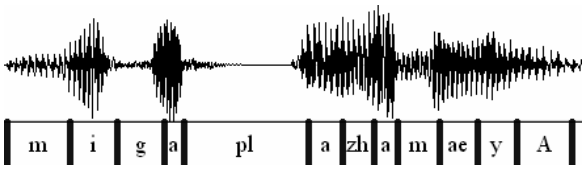
Class: The class to which the character belongs. These class labels can be effectively used to write a rule representing a broad set of characters.

Phoneme: The default phonetic representation of the character.

Type gives a broad classification of the characters while *Class* mainly classifies the C type characters (consonants) into different clusters like KA, CA, TA, tA, PA and YA. Some examples of the default mapping are shown in Figure 5.

3.2. Rule Base for Tamil G2P

The Rule Base contains a set of *rules* that modify the default mapping of the characters based on the context in which a particular phone occurs. Specific contexts are matched using rules. The system triggers the rule that best fits the current context. The rule format is given below.



(b)

Figure 3(a) Examples of Inter-word rules in Tamil
(b) /p/ /p/ becoming /pp/ (pl)

Format: $\alpha_1\ \alpha_2\ \ldots.\ \alpha_m\ \{\ \beta_1\ \beta_2\ \ldots\ \beta_n\ \}$

Example: `VOW KA VOW { K:1:X R:2:g K:3:X }`

α_i Class label of the i[th] character as defined in the character phone mapping. Together, these α_is represent the context that is being matched.

β_j j[th] action specification node. Each such node has the form: *Action_Type:Pos:Phoneme_Str*

Action_Type This field specifies the type of this action performed at this node. Possible values are *K* (Keep), *R* (Replace), *I* (Insert) and *A* (Append).

Pos The index of the character being covered by the context of the rule ($1 \le Pos \le m$)

Phoneme_Str represents phoneme string output by this action node.

The example rule given above says that if the grapheme /k/ (/k/ belongs to class KA) appears between two vowels (`VOW KA VOW`), keep the first character (vowel) as it is (`K:1:X`), replace the second character(/k/) with /g/ (`R:2:g`) and keep the third character (vowel) as it is(`K:3:X`).

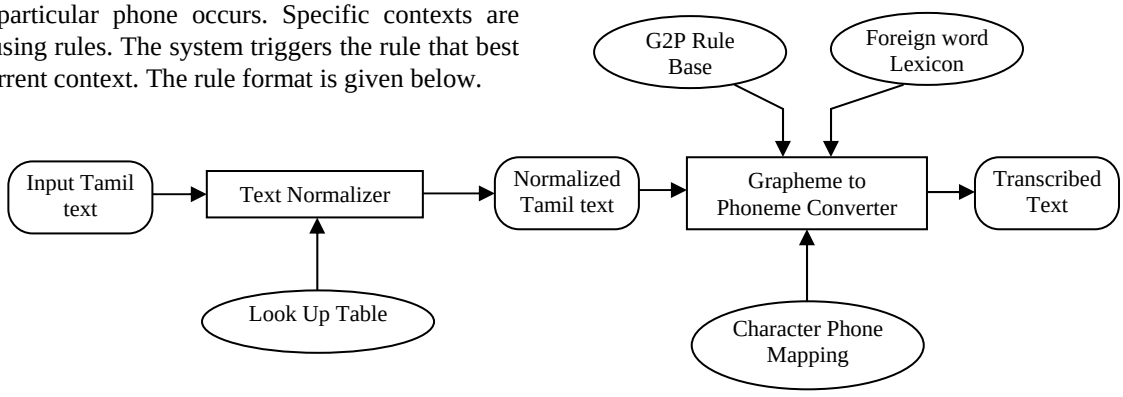


Figure 4: Natural Language Processing (NLP) module

மிகப் பழமையானது ⟶ மிகப்பழமையானது
மிகச் சிறந்த ⟶ மிகச்சிறந்த
மட்டக் களப்பு ⟶ மட்டக்களப்பு
விழித் திரையில் ⟶ விழித்திரையில்

(a)

அ	a	க	k
ஒ	o	ஆ	A
த	t	ஓ	O
ட	T	எ	e

Figure 5: Mapping examples

If the same consonant occurs twice consecutively, the genitive is represented by replacing the second grapheme by 'l'. Ex: `TA HAL TA { A:1:l } (T T -> Tl)` The reason for using this kind of representation for double letters is that the speech files in the inventory are labeled accordingly and if 'T T' is kept as it is, then the DSP module selects two 'T' segments instead of selecting a single 'Tl' segment. Also, linguistically, TT is a single phone (genitive), not two.

Graphemes exist in Tamil script only for the phonemes /k/, /ch/, /T/, /th/, /p/. But these five graphmes would manifest as /g/, /j/, /D/, /dh/, /b/, respectively, if they occur between two vowels or prefixed by nasals.

Ex: `NAS1 HAL KA { K:1:X K:2:X R:3:g }`
`VOW PA VOW { K:1:X R:2:b K:3:X }`

However, there is an exception for /ch/. If it occurs between two vowels, it becomes /s/. Also, /ch/ becomes /s/ when it comes in the beginning of a word. There are some more rules, which are not listed here. Figure 6 gives some sample input Tamil sentences, the normalized Text and the phonetised text.

3.3. Lexicon for handling Foreign Words

The process of corpora phonetization or the development of phonetic lexicons for the western languages is traditionally done by linguists. These lexicons are subject to constant refinement and modification. But the phonetic nature of Indian scripts reduces the effort to building mere mapping tables and rules for the phonetic representation. These rules and the mapping tables together comprise the phonetizers or the Grapheme to Phoneme converters [1].

The G2P rule base cannot be generalized to handle all the words in the input text, especially for the proper nouns derived from other languages like Sanskrit and Urdu. For example, the word 'Buddha' is written in Tamil as 'புத்தா' (putla); the grapheme /p/ in the beginning of the word should be pronounced as /b/ and the /tl/ should be pronounced as /dl/. But there are no such rules in the rule base, since it is basically a Sanskrit word used as it is in Tamil. If a rule is introduced for this purpose, that will affect Tamil words. For example, the Tamil word 'புத்தகம்' (putlakam) is pronounced as 'putlagam' only. The initial /p/ doesn't change to /b/ or the genitive /tl/ doesn't change to /dl/.

To cater to such exception words and proper nouns, a lexicon has been created. The lexicon dictates the phonetic composition, or the pronunciation of each entry in the list. The G2P first looks in the lexicon file for each input word. If the word is present in the lexicon file, its phonetic transcription is taken from the lexicon itself. Otherwise, the G2P applies mapping and the rules to produce the phonetic transcription.

The phonetic transcription generated by the G2P converter can be used for segmentation of the speech corpus. The phonetic transcription can be aligned with the speech waveform and the phone boundaries can be adjusted manually or by automatic speech segmentation algorithms.

4. CONCLUSION

The Tamil graphme to phoneme conversion module has been developed effectively. Efficient rules have been designed in, which cover most of the contexts in which the default mapping needs to be changed. A foreign word lexicon handles cases where the general G2P rules do not give the exact phonetic transcription. This has been used in the Tamil TTS developed around Festival, as well as an independent, stand-alone TTS. The developed C code for G2P module is designed to be used for Tamil TTS in both Windows and Linux platforms.

திரிவேணியின் பிறந்த தேதி september 1, 1928. பல pages கொண்ட 21 நாவல்கள், 41 சிறுகதைகள் அடங்கிய 3 தொகுப்புகள், கன்னட இலக்கியத்துக்கு அவருடைய பங்களிப்புகள்.

(a)

திரிவேணியின் பிறந்த தேதி ஸெப்டெம்பர் ஒன்று / ஆயிரத்து தொள்ளாயிரத்து இருபத்து எட்டு பல பக்கங்கள் கொண்ட இருபத்து ஒன்று நாவல்கள், நாற்பத்து ஒன்று சிறுகதைகள் அடங்கிய மூன்று தொகுப்புகள், கன்னட இலக்கியத்துக்கு அவருடைய பங்களிப்புகள்.

(b)

```
> t i r i w E N i y i n # p i R a n d a # t E d
i # s e p T e m b a r # o n R u # # A y i r a tl
u # t o Ll A y i r a tl u # i r u b a tl u # e
Tl u # # p a l a # p a kl a ng g a L # k o N D a
# i r u b a tl u # o n R u # n A w a l g a L $ #
n A R p a tl u # o n R u # s i R u g a d ae g a
L # a D a ng g i y a # m U n R u # t o g u pl u
g a L $ # k a nl a D a # i l a kl i y a tl u kl
u # a w a r u D ae y a # p a ng g a L i pl u g a
L <
```

(c)

Figure 6: (a) Input Tamil text in ISCII format (b) Tamil text after normalization (c) Transcribed Output of G2P Converter.

5. REFERENCES

[1] A. Gopalakrishna, Rahul C, Sachin J, Rohit K, Satinder Singh, R.N.V Sitaram and S.P. Kishore, "Development of Indian Language Speech Databases for Large Vocabulary Speech Recognition Systems," *Proc International Conf Speech and Computer* (SPECOM), Patras, Greece, Oct. 2005.
Link (http://en.wikipedia.org/wiki/Phonetic_transcription)

[2] Thierry Dutoit, "High-quality Text-To-Speech synthesis: an overview," *Journal of Electrical and Electronics Engineering*, Australia: Special Issue on Speech Recognition and Synthesis, vol. 17 no 1, pp. 25-37, 1997.