Complexity of Letter to Sound Conversion (LTS) in Khmer Language: under the context of Khmer Text-to-Speech (TTS)

(Annanda th. Rath, Long S. Meng, Heng. Samedi, Long Nipaul, Sok K. heng) [1]

[1] NLP lab, Department of Computer and Communication Engineering, Institute of Technology of Cambodia, Cambodia, PAN10 and IDRC Canada

rannanda@itc.edu.kh, seangmeng@itc.edu.kh, sheng@itc.edu.kh, nipaul@itc.edu.kh, sok.kimheng@itc.edu.kh

Abstract

Text to sound conversion is an important step in the process of creating the Text-to-Speech application for any languages. The complexity is generally based on the way the script is written. Some languages such as Khmer, Thai or Lao, one word may be pronounced to more than one ways depending on the syllable boundary in the word. This leads to the great difficulty in conversing correctly the text. The correctness of conversing text-to-sound in Khmer language is hard to archive 100%; however, it is not impossible.

This paper presents the steps in Khmer LTS, the difficulties and our research result.

Index Terms: LTS, TTS

1. Introduction

Speech synthesis systems use two basic approaches to determine the pronunciation of a word based on its spelling, a process which is often called text-to-phoneme, text-to-sound or grapheme-to-phoneme conversion. The simplest approach to text-to-phoneme conversion is the dictionary-based approach, where a large dictionary, also called lexicon, containing all the words of a language and their correct pronunciations is stored by the program. Determining the correct pronunciation of each word is a matter of looking up each word in the dictionary and replacing the spelling with the pronunciation specified in the dictionary. The other approach is rule-based, in which pronunciation rules are applied to words to determine their pronunciations based on their spellings.

Each approach has advantages and drawbacks. The dictionary-based approach is quick and accurate, but completely fails if it is given a word which is not in dictionary. As dictionary size grows, so too does the memory space requirements of the synthesis system, another drawback is that we have to update our dictionary whenever the new word comes. On the other hand, the rule-based approach works on any input, but the complexity of the rules grows substantially as the system takes into account irregular spellings or pronunciations. As a result, nearly all speech synthesis systems use a combination of these approaches.

Some languages, like Spanish, have a very regular writing system, and the prediction of the pronunciation of words based on their spellings is quite successful. Speech synthesis systems for such languages often use the rule-based method extensively, resorting to dictionaries only for those few words, like foreign names and borrowings, whose pronunciations are not obvious from their spellings. On the other hand, speech synthesis systems for languages like English, which have extremely irregular spelling systems, are more likely to rely on dictionaries and to use rule-based methods only for unusual words, or words that aren't in their dictionaries. [1]

Like English, our future Khmer Text To Speech System will rely on lexicon and use rule-based methods only for unusual words, or words that are not present in the lexicon.

2. Background of Khmer Language

The Cambodian scripts (called Khmer letters) are derived from various forms of the ancient Brahmi script of South India, and the ancient Khmer script itself. The Cambodian script has symbols for thirty-three consonants, twenty-four dependent vowels, twelve independent vowels, and several diacritic symbols. Most consonants have reduced or modified forms, called sub-consonants or subscript, when they occur as the second member of a consonant cluster. Vowels may be written before, after, over, or under a consonant symbol. [2]

The consonants are divided into 2 groups or series: ɑ-series and ɔ-series (see fig. 1 for the group of consonants along with their pronunciation). Most of consonants (in ɑ-series resp. ɔ-series) have their counterparts (in ɔ-series resp. ɑ-series). For consonants in one group that don't have their counterpart in another group, diacritic mark ◌̈ (MUUSIKATOAN) and ◌̎ (TRIISAP) are used to produce their counterpart. ◌̈ changes the consonant to ɑ-series while ◌̎ changes the consonant to ɔ-series. Among the 24 vowels there is 1 abstract (inherent) vowel which is embedded in a consonant. The pronunciation of a vowel, including the inherent vowel, is determined by the series of the initial consonant or consonant cluster that it accompanies. Khmer syllable in written form can contain 1 or 2 vowels. In the later case, last vowel is a vowel that has embedded consonant (see fig. 2 for the list of vowels and sequences of vowels that come together along with their pronunciation for each group). Independent vowels incorporate with both an initial consonant and a vowel (see fig. 3 for list of independent vowels along with their pronunciations). Khmer words are predominantly of one or two syllables. Words with three or more syllables exist, particularly those pertaining to science, the arts, and religion. These words are loanwords, usually derived from Pali, Sanskrit, or more recently, French. [3] Among the loanwords, about 90% are from Pali and Sanskrit. [4] Here, we classify words in Khmer language into 3 groups: original Khmer words (group 1), words borrowed from Pali and Sanskrit (group 2) and words borrowed from other languages like English, French and Chinese (group 3). The majority of words in Khmer language are in the first 2 groups. The last group contains only a small number of words.

Consonants and subscripts				
Letter				Sound
ɑ-series		ɔ-series		
Cons.	Sub.	Cons.	Sub.	
ក	◌្ក	គ	◌្គ	k
ខ	◌្ខ	ឃ	◌្ឃ	k^h
ង	◌្ង	ង	◌្ង	ŋ
ច	◌្ច	ជ	◌្ជ	c
ឆ	◌្ឆ	ឈ	◌្ឈ	c^h
ញ	◌្ញ	ញ	◌្ញ	ɲ
ដ	◌្ដ	ឌ	◌្ឌ	d
ឋ, ថ	◌្ឋ, ◌្ថ	ឍ, ធ	◌្ឍ, ◌្ធ	t^h
ណ	◌្ណ	ន	◌្ន	n
ត	◌្ត	ទ	◌្ទ	t
ប	◌្ប	ផ	◌្ផ	b
ផ	◌្ផ	ភ	◌្ភ	p^h
ព	◌្ព	ព	◌្ព	p
ម	◌្ម	ម	◌្ម	m
យ	◌្យ	យ	◌្យ	j
រ	◌្រ	រ	◌្រ	r
ឡ	◌្ល	ល	◌្ល	l
វ	◌្វ	វ	◌្វ	w
ស	◌្ស	ស	◌្ស	s
ហ	◌្ហ	ហ	◌្ហ	h
អ	◌្អ	អ	◌្អ	ʔ

figure 1. Text to sound mapping for consonants and subscripts

Vowels		
Letter	Sound	
	ɑ-series	ɔ-series
Inherent vowel	ɑ	ɔ
◌ា	a:	i:ə
◌ិ	e	i
◌ី	ej	i:
◌ឹ	ə	ɨ
◌ឺ	œ:	ɨ:
◌ុ	o	u
◌ូ	ɔ:o	u:
◌ួ	u:ə	u:ə
◌ើ	a:ə	ə:
◌ឿ	ɨ:ə	ɨ:ə
◌ៀ	i:ɜ	i:ɜ
◌េ	e:	ɐ:
◌ែ	a:ɛ	ɛ:
◌ៃ	aj	ej
◌ោ	a:o	o:
◌ៅ	aw	əw
◌ុំ	om	um
◌ំ	ɑm	ɯm
◌ាំ	am	oam
◌ះ	ah	ɛah
◌ិះ	eh	ih
◌ឹះ	əh	ɨh
◌ុះ	oh	uh
◌ុះ	eh	ih
◌ោះ	ɑh	uəh

Figure 2. Text to sound mapping for vowels and sequence of vowels

Independent vowels	
Letter	Sound
អា	ʔa:
ឥ	ʔe
ឦ	ʔej
ឧ	ʔu
ឩ	ʔu:
ឪ	ʔo:w
ឫ	ʔɨ
ឬ	ʔɨ:
ឭ	lɨ
ឮ	lɨ:
ឯ	ʔa:ɛ
ឰ	ʔaj
ឱ	ʔa:o
ឲ	ʔa:o
ឳ	ʔaw

Figure 3. Text to sound mapping for independent vowels

3. Pronunciation of Khmer Word

As we can see in fig. 4, the third consonant ក can be mapped to /kɑ:/ or /k/ depending on whether it is in the onset or the coda of a syllable. Note that the consonant រ is deleted if it is in the coda. So without knowing syllable boundary, pronouncing Khmer word is relatively difficult if not impossible. But once we have identified syllable boundary, the pronunciation is quite easy. So our approach to find pronunciation is in three steps. The first step is to do preprocessing to find syllable boundary and to replace diacritic marks with appropriate text. The second step is to find the pronunciation of each syllable. And the last step is to apply sound change rules. Note that for our approach, we don't need syllabification at phoneme level.

We will firstly present the rule to pronounce each syllable and after that we will show how the preprocessing is done.

ក	.	ក	ែ◌	.	ក	.	ក	រ
kɑ:	.	k	aɛ	.	kɑ:	.	kɑ:	

ក	.	ក	ែ◌	ក	.	ក	រ
kɑ:	.	k	aɛ	k	.	kɑ:	

Figure 4. Pronunciation of the word កែកកករ

3.1 Rule and algorithms for syllable pronunciation

In written form, a syllable in Khmer word can be one of this: C, CC(◌̍), CS, CSC(◌̍), I, IC, CV, CVC(◌̍), CSV, CSVC(◌̍), CSSV where C is a consonant, S is a subscript, I is an independent vowel, V is a vowel, ◌̍ (called Bantoc) is diacritic mark to make vowel a short vowel and parenthesis means optional. **Remark:** C, S, I and V are all in script form not in phoneme form.

Algorithm:
For each character in the syllable:
  If the character is a consonant or subscript
    Take corresponding sound in the text to sound mapping for the consonant and subscript
    If it is not followed by a vowel or a subscript
      Add inherent vowel (see below)
    If the character is a vowel (including inherent vowel which may have been added in previous step)
      Take the corresponding sound in text to sound mapping for the vowel where the group is the group of the preceding consonant or consonant cluster. To find group the consonant is in, just look in text to sound mapping table for the consonant. For consonant cluster see section 4 below.
    If the character is an independent vowel
    Take the corresponding sound in text to sound mapping table for independent vowel.
    If there is ◌̍ at the end then make the vowel a short vowel

3.2 Find the group of the consonant cluster

Consonant cluster in Khmer word is one of the forms CS or CSS. For the second form CSS there is only 2 possibilities: ស្ត្រ /str/ and ស្ក្រ /skr/ which are in ɑ-group.

For CS form, the algorithm to determine the group is as follow:
  If the consonant and subscript are in the same group
    The group of the cluster is the same as the group of consonant or subscript
  Else If the subscript is not in [◌្ង ◌្ញ ◌្ណ ◌្ម ◌្ប ◌្រ ◌្ល ◌្វ]
The group of the cluster is the same as the group of the subscript with two exceptions: ព្រឹះ (solidly packed) and ភ្ស៊ាំ (to tame)
  Else the group of the cluster is the same as the group of the consonant

3.3 Preprocessing

For words with more than one syllable, the group (ɑ or ɔ) of the consonant or consonant cluster of the second or higher

syllable may change. The change is unpredictable. Sometimes it depends on group of the preceding consonant or consonant cluster; sometimes it depends on the word itself. As examples, consider the following words: សាលា (school), ចំទាល (loudly), វិហារ(monastery), រហែក(torn), កិរិយា(manner). We use the following conventions to describe the pronunciation of the above words: Cons(C, group, ipa), Vowel(V, ipa for ɑ-group, ipa for ɔ-group), Independent vowel(I, ipa), i.e. ស(C, ɑ, /s/) means that ស is a consonant in ɑ-group with ipa /s/, ◌ា(V, /a:/, /iə/) means that ◌ា is a vowel with ipa for ɑ-group /a:/ and ipa for ɔ-group /iə/, and ឯ(I, /ʔaɛ/) means that ឯ is an independent vowel with ipa /ʔaɛ/.

សា.លា→ ស(C, ɑ, /s/) + ◌ា(V, /a:/, /iə/) + ល(C, ɔ, /l/) + ◌ា(V, /a:/, /iə/) → sa:.la: (1)

ចំ.ទាល→ ច(C, ɑ, /c/) + ◌ំ(V, /ɑm/, /um/) + ទ(C, ɔ, /t/) + ◌ា(V, /a:/, /iə/) + ល(C, ɔ, /l/) → cɑm.tiəl (2)

វិ.ហារ→ វ(C, ɔ, /v/) + ◌ិ(V, /e/, /i/) + ហ(C, ɑ, /h/) + ◌ា(V, /a:/, /iə/) → vi.hiə (3)

រ.ហែក→ រ(C, ɔ, /r/) + Inherent(V, ɑ:, ɔ:) + ហ(C, ɑ, /h/) + ែ◌(V, /aɛ/, /ɛ:/) + ក(C, ɑ, /k/) → rɔ:.haɛk (4)

កិ.រិ.យា→ ក(C, ɑ, /k/) + ◌ិ(V, /e/, /i/) + រ(C, ɔ, /r/) + ◌ិ(V, /e/, /i/) + យ(C, ɔ, /j/) + ◌ា(V, /a:/, /iə/) → ke.ri.ja: (5)

For (1), the group of consonant ល in the second syllable changes from ɔ to ɑ because the group of consonant ស in the first syllable is ɑ. The vowel following ល is pronounced using ɑ-group not ɔ-group. For (2), the group of the consonant ទ in the second syllable doesn't change to ɑ-group even if the group of the consonant in the first syllable is ɑ. The vowel following the consonant ទ is pronounced using ɔ-group. For (3), the consonant ហ in the second syllable changes from ɑ to ɔ because the consonant វ in the first syllable is in ɔ group. The vowel following ហ is pronounced using ɔ-group. For (4), the consonant ហ in the second syllable doesn't change from ɑ to ɔ even the consonant រ in the first syllable is in ɔ-group. The vowel following ហ is pronounced using ɑ-group. For (5), the consonant រ in the second syllable doesn't change from ɔ to ɑ even if the the consonant ក in the first syllable is in ɑ-group. Hence the vowel ◌ិ following រ is pronounced using ɔ-group. On the other hand, the consonant យ in the third syllable changes from ɔ to ɑ because the consonant ក in the first syllable is in ɑ-group. The vowel ◌ា following យ is pronounced using ɑ-group.

From the above example, we can see that it is not easy to find rules to solve this issue because the change seems irregular. A statistical approach would be needed to tackle this problem. So for now our preprocessing consists of breaking the word into syllables and replace diacritic mark with appropriate text.

As we have mentioned in Background, words in Khmer language is divided into 3 groups. While words in group 1 and group 3 have a similar rule for pronunciation, words in group 2 have a very different rule. We usually say that words in group 1 and group 3 have Khmer style pronunciation while words in group 2 have Pali style pronunciation. So we have 2 preprocessing rules: one for the words in group 1 and 3 and another for the words in group 2. To know which processing rule to apply for a given word, we need to know the group in which the word belongs. Hence we also need to detect word group. Our method is to detect whether the word is in group 2 or not. If it is not in group 2 then it is in group 1 or 3. We will call this module "Pali / Sanskrit word detection".

So in this section we present the following points: Pali / Sanskrit word detection, preprocessing for words in group 1 and 3, preprocessing for words in group 2 and the complete algorithm for our preprocessing.

3.3.1 Pali/Sanskrit word detection

Detecting whether word is Pali / Sanskrit can be done with and without syllable boundary knowledge.

Pali / Sanskrit word detection without knowing syllable boundary:

- one of these diacritic marks [◌៌ ◌៍ ◌៑ ◌ះ] appear in the word
- one of these vowels [◌ិ ◌ី ◌ុ] is at the end of the word
- there is CS form which is not a possible consonant cluster providing that C is not one of [ង ញ ន ណ ម]

If any of the above conditions is true then the word is Pali / Sanskrit word.

Pali / Sanskrit word detection knowing syllable boundary:

- one of these vowels [◌ិ ◌ី ◌ុ] is at the end of a syllable
- more than 4 syllables in a word

If any of the above conditions is true then the word is Pali / Sanskrit word.

3.3.2 Preprocessing for words group 1 and 3 (Detect syllable boundary)

1. Insert syllable boundary after vowels ◌ះ ើ◌ ៅ◌ and diacritic marks ◌់

2. Insert syllable boundary before and after the form C◌៍ and remove diacritic mark ◌៍

3. for vowel ◌ំ

If it is preceded by vowel ◌ា and followed by consonant ង

Insert syllable boundary after ង

Else Insert syllable boundary after ◌ំ

4. Insert syllable boundary before independent vowel

5. For the form CV, insert syllable boundary before C

6. Insert syllable boundary before the form CC◌់

7. For consonant cluster CS which is not at the beginning of the word and no syllable boundary has been added before C

If C is in [ង ណ ន ម] Insert syllable boundary after C

If C is preceded by a consonant Insert ◌់ after C (note that ◌់ will be before the syllable boundary)

Replace the subscript S with the corresponding consonant

Else if c is in [ញ] Insert syllable boundary after C

If C is preceded by a consonant Insert ◌ា before C and ◌់ after C

Replace the subscript S with the corresponding consonant

Else Insert syllable boundary before C

8. for the form Cរ if រ is at the end of the word or there is syllable boundary after it Insert syllable boundary before C and Remove រ

9. Insert syllable boundary after the form CC, CVC, CSVC provided that no syllable boundary has been inserted inside these forms. (This is maximum template matching from left to right)

These rules are applied in order from top to bottom.

3.3.3 Preprocessing for words in group 2

1. Remove all string of the forms C◌៍ and SV◌៍

2. Insert syllable boundary after vowels ◌ះ ◌ៃ ◌ោ and diacritic marks ◌៉

3. Insert syllable boundary before and after the form C◌៏ and remove diacritic mark ◌៏

4. for vowel ◌ុ

If it is preceded by vowel ◌ា and followed by consonant ង

Insert syllable boundary after ង

Else Insert syllable boundary after ◌ុ and change it to ◌ាម៉

5. Insert syllable boundary after diacritic mark ◌ៈ and replace it by ◌ាក់

6. Insert syllable boundary before independent vowels

7. For the form CV, insert syllable boundary before C

8. Insert syllable boundary before the form CC◌៉

9. If the form C_1C_2◌ិ is at the end of the word, insert syllable boundary before it and change C_2◌ិ to ◌ិះ

10. If the form C_1C_2◌ិ is not at the end of the word, insert syllable boundary before it and change it to C_1◌ិះ.C_2

11. If the form C_1◌ិC_2 is at the end of the word, insert syllable boundary before it and change it to ◌ាក់. C1C2◌៉ ('.' denotes syllable boundary)

12. If the form VC◌ិ is at the end of the word, change it to V◌ះ

13. If the form ◌ិCS appears at the end of the word, remove S

14. If the form ◌ិCV appears at the end of the word, remove V

15. If the form V ◌ិ appears at the end of the word, remove vowel ◌ិ

16. For consonant cluster CS which is not at the beginning of the word and no syllable boundary has been added before C
If CS is possible consonant cluster (see fig. 6 for the list of possible consonant clusters)
Insert syllable boundary before C
Else Insert syllable boundary after C
If C is preceded by a consonant
Insert ◌ិ before C Replace subscript S with the corresponding consonant

17. For the form ◌ិC, insert syllable boundary after C

18. If the form C◌ឺ appears at the end of the word, insert syllable boundary before C and change ◌ឺ to ◌ិត

19. If the form C◌ិ appears at the end of the word, insert syllable boundary after C

20. If the form C◌េ◌ុ appears at the end of the word, insert syllable boundary before C and change ◌េ◌ុ to ◌ែត

21. If the form C◌ា◌ុ appears at the end of the word, insert syllable boundary before C and change ◌ា◌ុ to ◌ាត

22. If the form ស◌ុ appears at the end of the word, change it to ស◌ំ

23. If the form ◌ឹCS appears at the end of the word, remove S

24. For the form ◌ឹស◌ំ S, insert syllable boundary after ស◌ំ and insert ស◌ំ before S

25. For the form ◌ឹC, insert syllable boundary after C

26. If the form C_1C_2 is not at the end of the word and it is not followed by diacritic mark ◌៉, insert syllable boundary after C_1

27. If the form VC is not at the end of the word, insert syllable boundary after V

4. Evaluation and discussion

We implemented this algorithm using Python. To test the accuracy, we have built a lexicon containing 9445 words. The accuracy is about 72%. Most of the incorrect results are due to the group change we mentioned at the beginning of section 2. There are also some incorrect results due to incorrect Pali / Sanskrit word detection. Some of the errors are Pali / Sanskrit words. We haven't tested the accuracy of our Pali / Sanskrit word detection yet because we don't have information on word group in the lexicon we have built for the test.
The result obtained after the evaluation is acceptable, but not well. Further works have to be done to improve our Text to Sound Conversion algorithms. In our method, the improvement can be done by adding the module to predict group change of consonant or consonant cluster of the second or higher syllable in the word with more than one syllable. The statistical approach as described in [5] may be also be used.

5. Complexities of LTS in Khmer language

The difficulties that we found are like others languages such as Thai or Lao, where one word may be pronounced to more than one ways depending on the syllable boundary. This leads to the ambiguity of the meaning while conversing to the sound. For example, in Khmer language, the word កែកករ (to mumble) can have 2 different pronunciations depending on syllable boundary. first one is /kɑː.kaɛ.kɑː.kɑː/ corresponding to ក.កែ.ក.ករ and another is /kɑː.kaɛk.kɑː/ corresponding to ក.កែក.ករ. Some methods such content analysis technique is needed to improve the performance of the TTS application creating. In Khmer language, there are many most frequently used words that fall into the category mentioned above. As the result of our experimentation, we found that most of the error came from the mis-identifying of the meaning of the words as those words have the ambiguity form.

6. Conclusions

Our method to do text to sound conversion is by doing some preprocessing at script level to determine syllable boundary and replace diacritic marks with appropriate text. After the preprocessing, we convert each syllable to sound using some simples' rules. The last step is the application of sound change rules. The output of our text to sound conversion also includes syllable boundary. To measure the accuracy of our method, we implemented our algorithm using Python and built a lexicon containing about 10000 words. The test yields about 72% accuracy. This result is the first outcome of our research on Khmer language. The further research needs to be done in order to improve the accuracy of the algorithms that we have developed. This work will be carried on in next phrase.

References

[1] http://en.wikipedia.org/wiki/Speech_synthesis, 09/10/2008

[3] http://en.wikipedia.org/wiki/Khmer_language, 10/10/2008

[2]http://www.seasite.niu.edu/khmer/writingsystem/writing_introduction/ intro_set.htm, 10/10/2008

[4] Khmer grammar

[5] Issues in Building General Letter to Sound Rules, Alan W Black – Kevin Lenzo – Vincent Pagel