Research report on phonetic and phonological analysis of Khmer

Abstract

This report presents result of research on phonetic and phonological analysis of Khmer as part of Khmer Text-To-Speech (TTS) project. The study includes phonetic analysis of Khmer sounds, syllabification in Khmer, stress analysis, intonation analysis and sound change rules.

In phonetic analysis of Khmer sounds, we have identified Khmer phonetic inventory with their associative features. The study of syllabification in Khmer leads to the identification of a set of rules for syllabifying a given word. The research on stress analysis in Khmer revels that Khmer words have stress and at least one stress on the last syllable. The location of stress depends on number of syllables in the words. Many rules are identified for sound change in Khmer.

I. Introduction

The stream of speech is composed of two kinds of phonological units: Segmental sounds and Suprasegmental sounds. Segmental sounds are those which can be segmented into distinct, discrete units, such as vowels and consonant [9]. The features of speech like variations of pitch, stress, and accents are called suprasegmental sounds. The naturalness of the synthesized speech is largely characterized by the fact that how successfully the above mentioned phonological units i.e. both segmental and suprasegmental features exists in the language are modeled in the speech synthesizer.

The research is focused on identifying unique characteristics of Khmer language in terms of phonetic and phonology for developing our Khmer TTS system.

Few researches have been carried out in the linguistic domain regarding Khmer phonetic and phonology. The themes elaborated in these researches are Khmer Stress and Intonation. In the domain of Khmer speech processing, a very little number of researches were reported. To our knowledge, there are two lecturers at Institute of Technology of Cambodia (ITC) who are doing their doctoral on this field: one on Khmer Automatic Speech Recognition (ASR) and another on Multilingual ASR.

II. Background

The Khmer language is the official language of Cambodia. It is one of the main Autroasiatic languages. It is spoken in Cambodia, Vietnam, Thailand, USA, France and Australia. The total number of Khmer speakers is about 15.7 to 21.6 millions (2004) in which there are about 14.7 to 20.6 millions of native speakers and about 1 million of second language speakers.

The Khmer script, called aksaa khmae (“Khmer letters”), is used to write the Khmer language. The Khmer alphabet is descended from the Brahmi script of ancient India by way of the Pallava script, which was used in southern India and South East Asia during the 5th and 6th Centuries AD. The oldest dated inscription in Khmer, found at Angkor Borei in Takev Province south of Phnom Penh, dates from 611 AD.

Structurally, the Khmer script has many features in common with other Brahmi-derived scripts, such as Devanagari and Myanmar. Consonant characters bear an inherent vowel sound, with additional sign placed before, above, below, and/or after the consonants to indicate a vowel other than the inherent one. The overall writing direction is from left to right.

III. Phonetic Analysis of Khmer Sounds

III.1. Phonemic Inventory

There are 21 phonemes of vowels (12 long phonemes and 9 short phonemes), 10 diphthongs and 21 phonemes of consonants. Khmer doesn't use tonal contrast. The air stream mechanism used is pulmonic egressive. We use voicing contrast (voiceless, voiced and aspirated). Hence [p], [b] and [ph] are different phonemes. Table below shows different places and manner of articulation used by Khmer language.

Place / Manner	Bilabial	Labio-Dental	Alveolar	Palatal	Velar	Glottal
Stops	p, b, p^h		t, d, t^h	c, c^h	k, k^h	ʔ
Nasals	m		n	ɲ	ŋ	
Flaps			r			
Fricatives			s			h
Lateral Approximants			l			
Approximants		w		j		

Table 3-1 Different places and manners of articulation used by Khmer language

Consonants minimal pair:

No.	Phoneme	Letter(s) Representing the Phoneme	Feature Description	Example word in the language	IPA	English Translation
1	p	ប៉ ព	Voiceless Bilabial Stop Aspirated	ប៉ា	pa:	father
2	b	ប ប៊	Voiced Bilabial Stop	បា	ba:	male (cow)

			Unaspirated			
3	p^h	ផ ភ	Voiceless Bilabial Stop Unaspirated	ផា	p^ha	cloth
4	t	ត ទ	Voiceless Alveolar Stop Unaspirated	តា	ta:	old man
5	d	ដ ឌ	Voiced Alveolar Stop Unaspirated	ដា	da:	stone
6	t^h	ថ ធ ឍ ឋ	Voiceless Alveolar Stop Aspirated	ថា	t^ha:	say
7	c	ច ជ	Voiceless Palatal Stop Unaspirated	ចារ	ca:	yes (female)
8	c^h	ឆ ឈ	Voiceless Palatal Stop Aspirated	ឆា	c^ha:	to fry
9	k	ក គ	Voiceless Velar Stop Unaspirated	ការ	ka:	to marry
10	k^h	ខ ឃ	Voiceless Velar Stop Aspirated	ខា	k^ha:	bad oder
11	ʔ	អ អ៊	Voiceless Glottal stop Unaspirated	អារ	ʔa:	to saw
12	m	ម ម៉	Bilabial Nasal Unaspirated	ម៉ា	ma:	grand mother
13	n	ណ ន	Alveolar Nasal Unaspirated	ផ្ទះណា?	p^htɛah na:? house which?	Which house? (question word)
14	ɲ	ញ ញ៉	Palatal Nasal Unaspirated	កញ្ញា	kaɲ.ɲa:	miss
15	ŋ	ង ង៉	Velar Nasal Unaspirated	ង៉ា	ŋa:	word used to call person
16	s	ស ស៊	Voiceless Alveolar fricative Unaspirated	សា	sa:	to roll
17	h	ហ ហ៊	Voiceless Glottal Fricative	ហា	ha:	to open (mouth)

18	r	រ រ៉	Voiced Alveolar-Flaps Unaspirated	រ៉ា	ra:	train station
19	l	ល ឡ	Voiced Alveolar Literal Approximant Unaspirated	កីឡា	keij.la:	sport
20	w	វ វ៉	Voiced Labiodental Approximants Unaspirated	វ៉ា	wa:	to overtake (driving)
21	j	យ យ៉	Voiced Palatal Approximants Unaspirated	កញ្ចាស់យ៉ា	kaɲ.cah ja: old very	very old (used with old to mean very)

Table 3-2 Minimal pair of Khmer consonants

Vowels minimal pair:

In writing, Khmer consonants are divided into two groups or series: ɑ-series and ɔ-series. When a vowel is attached to a consonant, the sound of the vowel depends on the group the consonant is in. For instance, consonant ក /k/ is in ɑ-groups and គ/k/ is in ɔ-groups. When we attach vowel ◌ា to the two consonants the sound of the two words is កា /ka:/ and គា /ki:ə/.

No	Phoneme	Letter(s) Representing the Phoneme		Minimal pair		
		ɑ-series	ɔ-series	Example word in the language	IPA	English Translation
1	i		◌ិ	មីន / មិន	min / mɨn	mine / not
2	i:		◌ី	ជី / ជា	ci: / ci:ə	fertilizer / to be
3	ə	◌ឹ		យឹត / យឺត	jət / jɨ:t	to pull down / slow
4	ə:		◌ើ	ទើរ / តើ	tə: / ta:ə	get stuck or caught / question word
5	u		◌ុ	ទ្រុង / ទ្រូង	truŋ / tru:ŋ	cage / chest
6	u:		◌ូ	ទ្រូង / ទ្រុង	tru:ŋ / truŋ	chest / cage

7	e	◌ិ		ច្រិច / ច្រេច	crec / cre:c	sound of water squirting out of a small hole / sound of cricket crying
8	e:	◌េ		ច្រេច / ច្រិច	cre:c / crec	sound of cricket crying / sound of water quirting out of a small hole
9	œ:	◌ឹ		ប៊ឹង / ប៊ឹង	pœ:ŋ / pəŋ	sound of bell / breechcloth
10	o	◌ុ		ចុក / ចូក	cok / cɔ:ok	to close with a cork / to shovel
11	o:		◌ោ	គោ ក្លូរ	ko: / kɔ:o	cow / to stir (soup)
12	ɔ		អ◌ំ	លង់ / លង	lɔŋ / lɔ:ŋ	to drown / to haunt
13	ɔ:		អ◌	លង / លង់	lɔ:ŋ / lɔŋ	to haunt / to drown
14	a	◌ា◌ំ		ចាក់ / ចាក	caʔ / ca:ʔ	to stab / to leave
15	a:	◌ា		ចាក / ចាក់	ca:ʔ / caʔ	to leave / to stab
16	ɑ	អ◌ំ		បក់ / បក	bɑʔ / bɑ:ʔ	to blow(wind) / to peel (fruit)
17	ɑ:	អ		បក / បក់	bɑ:ʔ / bɑʔ	to peel (fruit) / to blow(wind)
18	ɨ		◌ឹ	រឹង / រីង	rɨŋ / ri:ŋ	hard solid / to dry
19	ɨ:		◌ឺ	យឺត / យឺត	jɨ:t / jət	slow / to pull down
20	ɐ:		◌ែ	គែ / កេរ៏	kɐ: / ke:	he, she, they, one, someone / reputation
21	ɛ:		◌ៃ	ខែ / គែ	kɛ: / kɐ:	crop(of a bird) / he, she, they, one, someone

Table 3-3 Minimal pair of Khmer monothong vowels

Diphthongs minimal pair:

No	Phoneme	Letter(s) Representing the Phoneme		Minimal pair		
		ɑ-series	ɔ-series	Example word in the language	IPA	English Translation
1	a:ɛ	◌ៃ		ប៉ែន / ប៊ិន	paɛn / pen	draw-plate / skillful

2	a:ə	◌ើ		ទើរ / តើ	tə: / ta:ə	get stuck or caught / question word
3	ɔ:o	◌ូ		ចុក / ចូក	cok / cɔ:ok	to close with a cork / to shovel
4	u:ə		◌ួ	គួរ / កួរ	ku:ə / kɔ:o	suitable / to stir (soup)
5	a:o	◌ោ		ចោរ / ជោរ	ca:o / co:	thief / to rise (water, tide)
6	i:ɜ	◌ៀ	◌ៀ	កៀង / កឿង	ki:ɜŋ / kɨ:əŋ	to drive (animal) / to get stuck (ship)
7	ɨ:ə	◌ឿ	◌ឿ	កៀង / កឿង	ki:ɜŋ / kɨ:əŋ	to drive (animal) / to get stuck (ship)
8	i:ə		◌ា	មាន / មៀន	mi:ən / mi:ɜn	to have / longan
9	oa		◌ាត់	វាត់ / វាត	woat / wi:ət	to beat (with a stick) / to expand (the area)
10	ɛa		◌ាក់	ពាក់ / ពាក្យ	pɛaʔ / pi:əʔ	to wear / word

Table 3-4 Minimal pair of Khmer diphthong vowels

III.2. Acoustics Phonetics

III.2.1 Acoustics Cues of Vowels

IPA	F1	F2	F3	Duration
i	347.72	2622.6	4205.8	0.092
i:	420.58	2931.3	4114.8	0.102
e	449.36	2543.2	4525.8	0.087
e:	438.71	2386	4023.8	0.104
ɐ	347.72	2313.2	4060.8	0.112
a:ɛ	597.6	2079.9	3432.6	0.103
ɛ:	356.72	2524.6	4025.5	0.130
a	690.24	2172.6	3784.6	0.056
a:	504.95	1727.9	2487.6	0.120

ɨ	430.84	1839.1	2487.6	0.068
ɨː	347.72	2258.6	3623.6	0.146
ə	931.12	2432	3710.5	0.083
œː	912.59	2487.6	3932.9	0.119
əː	393.78	1987.3	40.81.1	0.196
ɑ	875.54	2561.7	3284.3	0.047
ɑː	727.3	2617.3	3377	0.182
u	493.31	2477	3769	0.028
uː	727.3	2394.9	3692	0.154
o	542.01	1968.8	3526.2	0.066
oː	384.12	2003.8	3659.9	0.161
ɔ	393.7	1394.4	2561.7	0.30
ɔː	584.3	2622.6	2932.09	0.347
ɔo	393.8	941.5	2542.7	0.108
iɜ	319.13	713.23	2483.4	0.22
ɛa	438.71	1546.43	2221.45	0.09
iə	347.72	1438.54	2158.03	0.127
ao	634.66	2138.23	2421.01	0.163
aə	616.13	1911.87	2174.34	0.367
ɨə	438.7	1751.76	2241.76	0.302
uə	430.84	1113.34	2462.23	0.203
oa	529.71	2345.98	3021.7	0.219

Table 3-5 Acoustic cues of Khmer vowels

Other distinctive acoustic cues

- Lip Rounding: We don't use lip rounding to distinguish vowels.
- Tone: Khmer language is not a tonal language.

III.2.2. Acoustics Cues of Consonants

III.2.2.1. Stops

- The average closure duration for stops is about 0.119258 seconds.
- Specific place cues
  - Burst patterns
    1. Bilabial: the burst is thin. The burst for other places are thicker. (Figure 3-1)
    2. Alveolar: the burst is thicker than for the bilabial. It is thinner than for palatal and velar. (Figure 3-2)
    3. Palatal: the burst for this place is thicker than the one for bilabial and alveolar. It appears only for high frequencies (Figure 3-3).
    4. Velar: the burst for this place is thicker than the one for bilabial and alveolar. It is not like palatal because it appears for almost all frequencies (Figure 3-4).
  - Formant transitions for F1 and F2
    1. Bilabial: F1 and F2 go down. (Figure 3-1)
    2. Alveolar: F2 converges to 2000Hz. (Figure 3-2)
    3. Palatal: F1 go down, F2 and F3 go up. (Figure 3-3)
    4. Velar: F1 go down, F2 go up and F3 doesn't change or slightly go down. (Figure 3-4)
- Voicing contrasts: we have consonants that are different in voiceless, voiced and aspirated. Let see the case of /p/, /b/ and /p^h/. The pre-voicing for /b/ is 0.084508 seconds; the aspiration for /p^h/ is 0.066355 seconds.
- All stops pulmonic egressive.
- Nasal stops: during the stop we can see noise (Figure 3-5).

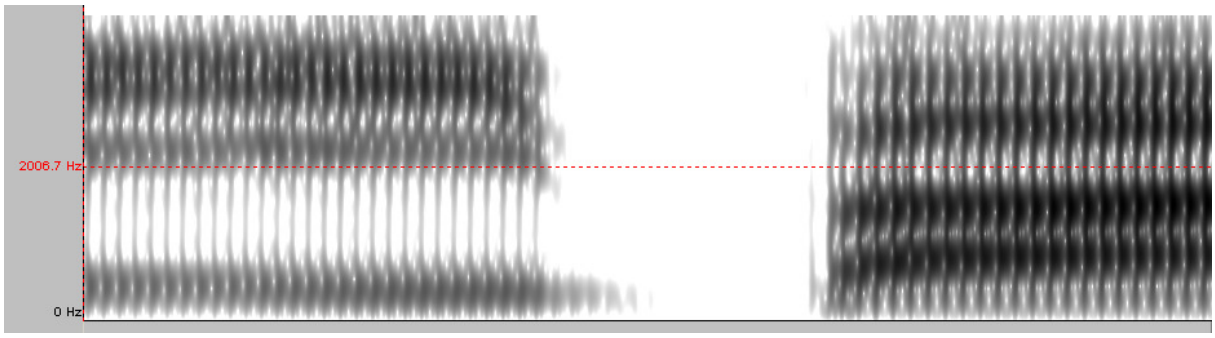


Figure 3-1 Spectrogram of [ipa] (Bilabial)

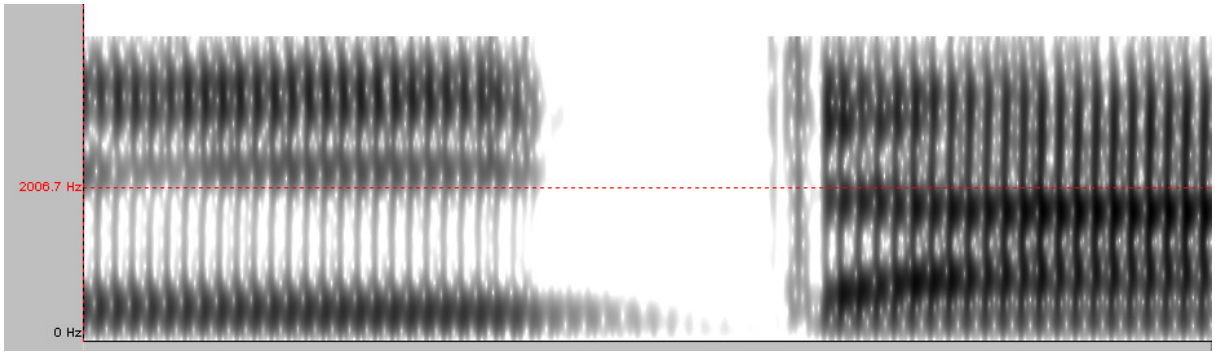


Figure 3-2 Spectrogram of [ita] (Alveolar)

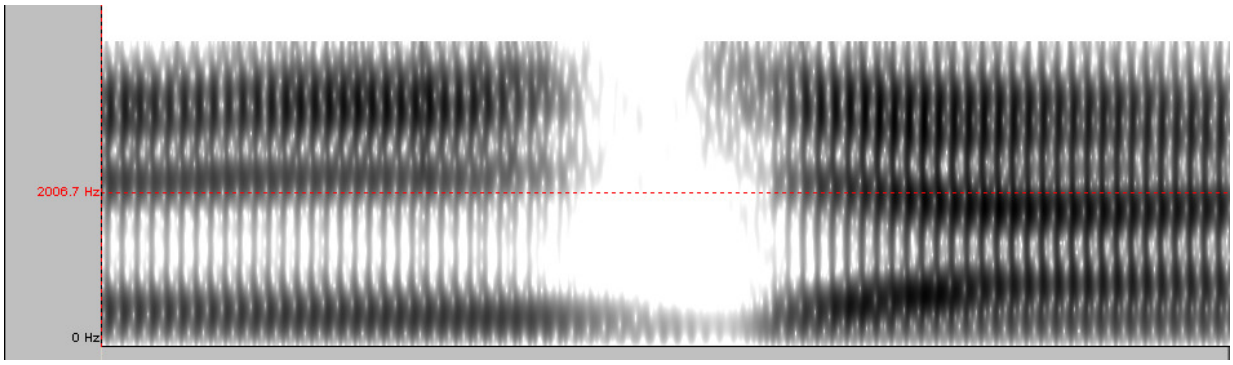


Figure 3-3 Spectrogram of [ica] (Palatal)

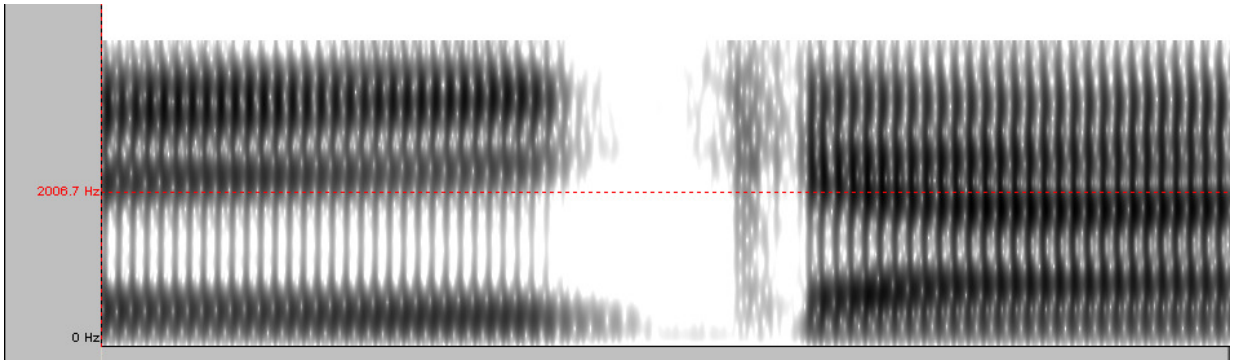


Figure 3-4 Spectrogram of [ika] (Velar)

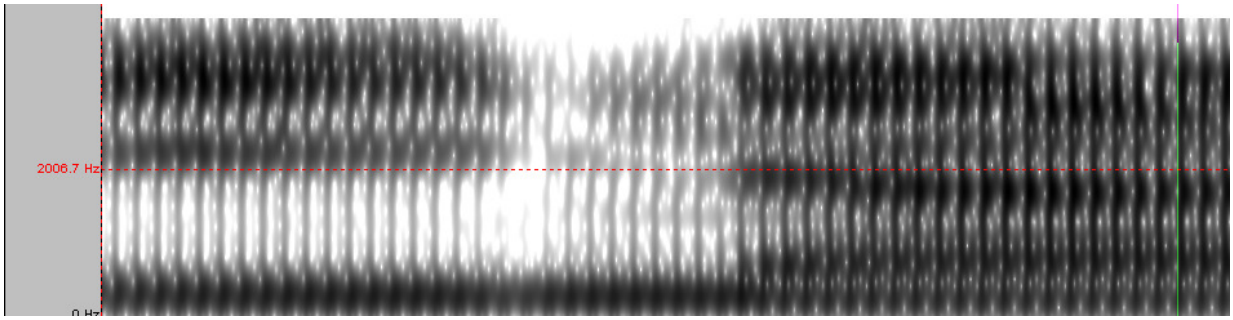


Figure 3-5 Spectrogram of [ina] (Alveolar nasal)

III.2.2.2. Fricatives

We have only 2 fricatives: /s/ and /h/. For the glottal fricative /h/, there are noises everywhere during the consonant. (Figure 3-7)
For the alveolar fricative /s/ the cut-off frequency is about 3000Hz. (Figure 3-8)

- Fricative duration intervocalicaly (between vowels) : 0.118614 seconds
- Khmer language doesn't make voicing contrast.

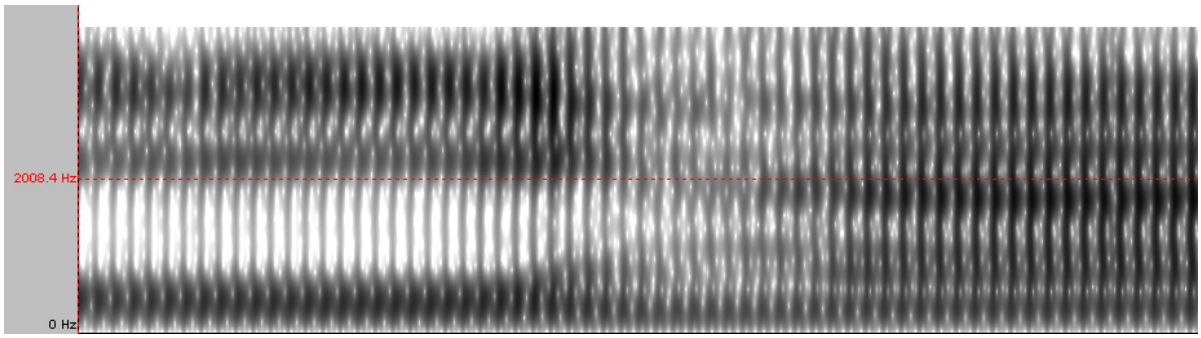


Figure 3-7 Spectrogram of [iha] (Glottal fricative)

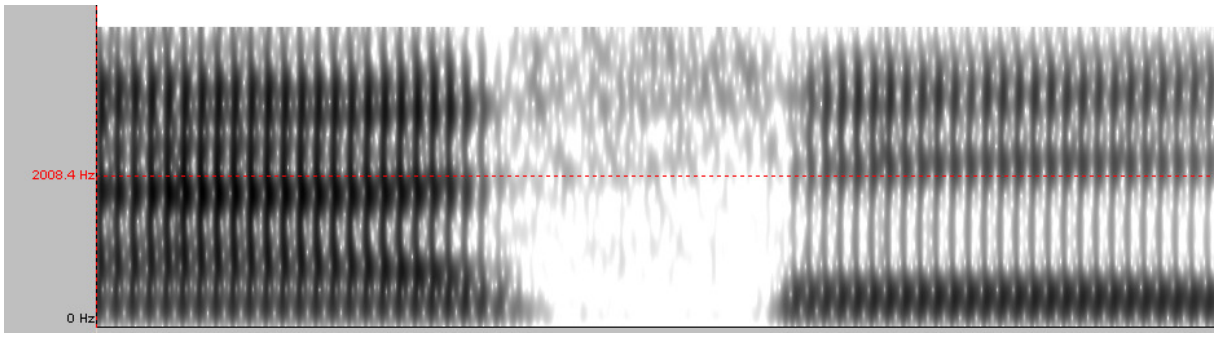


Figure 3-8 Spectrogram of [asi] (Alveolar fricative)

III.2.2.3. Approximants

We have 2 approximants /w/ and /j/. They are similar to those of English. /w/ is like /u/ and /j/ is like /i/.

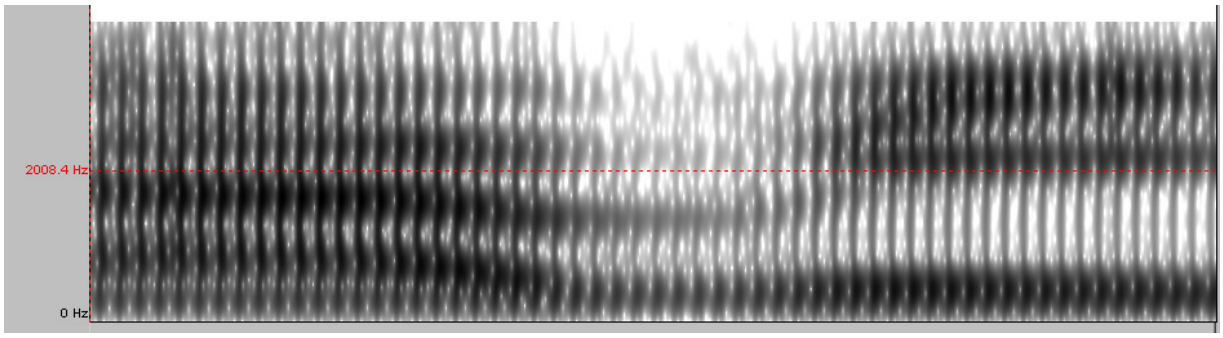


Figure 3-9 Spectrogram of [awi]

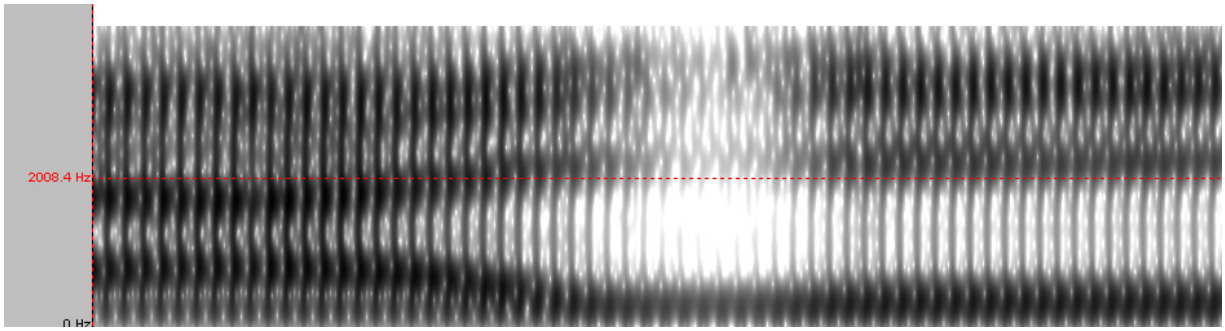


Figure 3-10 Spectrogram of [aji]

III.2.2.4. Laterals

We have only one lateral /l/. To see its feature let compare with the flap below.

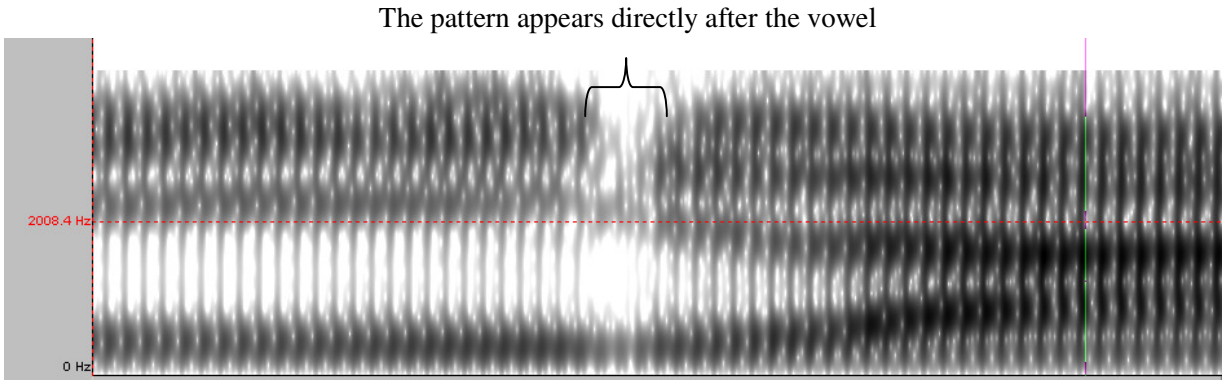


Figure 3-11 Spectrogram of [ila]

III.2.2.5. Flaps/Taps

We have only one flap /r/.

The pattern doesn't appear directly after the vowel, it takes a little time after the vowel.

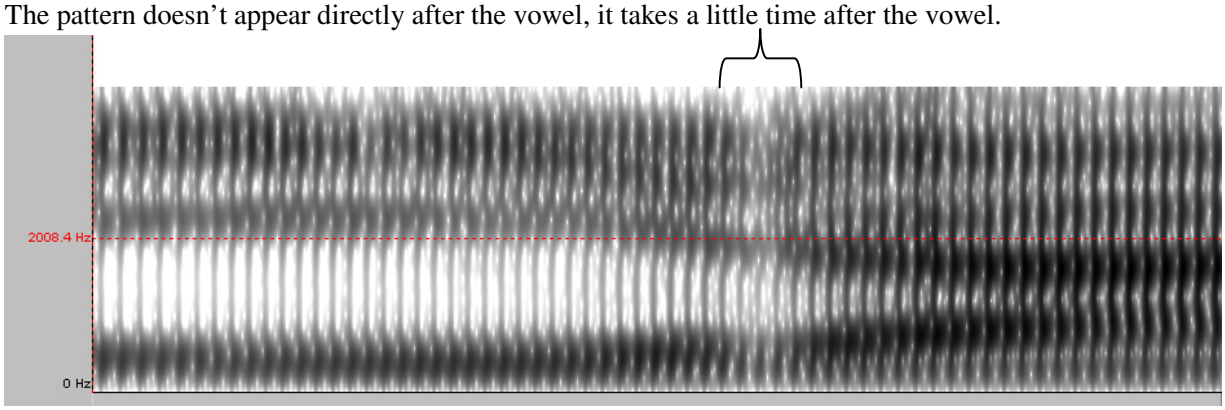


Figure 3-12 Spectrogram of [ira]

III.3. Conclusion

We have identified phonemic inventory of Khmer. The list of diphones, which is one of the important parts of diphone concatenation TTS system, can be created using this inventory.

The identification of acoustic cues of vowels and consonants is very helpful for diphone splitting and sound corpus tagging.

IV. Syllabification in Khmer

IV.1. Introduction

In one sense, a syllable is the smallest possible unit of speech. Every utterance must contain at least one syllable. It is convenient to talk of speech as being composed of segments such as vowels and consonants, but these segments can be observed only as aspects of syllable. A syllable can also be divided for descriptive purpose into its **onset** and **rhyme,** as shown in the following figure. The rhyming part of a syllable consists of the vowel and any consonants that come after it – a fairly familiar notion. Any consonants before the rhyme form the onset of the syllable. The rhyme of a syllable can be further divided into the **nucleus**, which is vocalic part, and the **coda**, which consists of any final consonants.

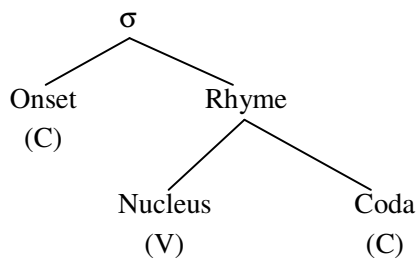


Figure 4-1 Syllable representation

In Khmer language, the syllables are divided into 6 groups: CV, CVC, CCV, CCVC, CCCV, and CCCVC; as shown in the following examples:

CV	[kɑ:]	"neck"	
CVC	[kɑ:j]	"to scratch (the ground)"	
CCV	[c^{h}mɑ:]	"cat"	
CCVC	[cbɑh]	"clear"	
CCCV	[sas.strɑ:.ca:]	"professor"	
CCCVC	[streh]	"stress"	(loan word)

There are not many syllables that are in the last two groups. The general form of Khmer syllable can be written as: C(C)(C)V(C) which shows that the maximum number of constant cluster is three and the number of the consonant after the vowel is either one or zero. Moreover, the general form shows that, in a syllable, there is at least one consonant in the onset.

The table below presents a list of possible consonant cluster which will be important for the syllabification:

Consonant cluster	**Example**		
pd	ប្ដូរ	[pdɔ:o]	"to change"
p^{h}t	ផ្ទី	[p^{h}ti:]	"a type of edible plant"
p^{h}d	ផ្តល់	[p^{h}dɑl]	"to provide"
p^{h}c	ផ្ចាញ់	[p^{h}caɲ]	"to defeat"
p^{h}k	ផ្កា	[p^{h}kɑ:]	"flower"
p^{h}n	ភ្នំ	[p^{h}num]	"mountain"
p^{h}ɲ	ភ្ញាស់	[p^{h}ɲoah]	"to hatch"

pʰŋ	ផ្ងារ	[pʰŋaː]	“to face up”
pʰs	ផ្សា	[pʰsaː]	“market”
pʰl	ផ្លូវ	[pʰlɔːow]	“street”
pr	ព្រែក	[prɛːk]	“small river”
pʰj	ព្យាង្គ	[pʰjiːəŋ]	“syllable”
tb	ត្បាល់	[tbal]	“small mill”
tʰk	ថ្គាម	[tʰkiːəm]	″molar″
tʰm	ថ្ម	[tʰmɑː]	″brick″
tʰn	ធ្នូ	[tʰnuː]	″bow, december″
tʰŋ	ថ្ងូរ	[tʰŋɔːo]	“to moan”
tʰw	ធ្វើ	[tʰwəː]	“to do”
tʰl	ថ្លា	[tʰlaː]	“clear (water)”
tr	ទ្រូ	[truː]	“a kind of fishing instrument”
tʰj	ធ្យូង	[tʰjuːŋ]	“coal”
cʰp	ផ្ការ	[cʰpuəh]	“to”
cb	ច្បាប់	[cbap]	“law”
cd	ផ្ការ	[cdao]	“ingot (of gold)”
cʰk	ឆ្កែ	[cʰkaːɛ]	“dog”
cʰm	ឆ្មា	[cʰmaː]	“cat”

cʰn	ឆ្នាំ	[cʰnam]	“year”
cʰŋ	ឆ្ងល់	[cʰŋɑl]	“to wonder”
cʰw	ឆ្វេង	[cʰweːŋ]	“left”
cʰl	ឈ្លី	[cʰliː]	“to crush with the fingers”
cr	ជ្រៃ	[crɛ̰ː]	“oblique”
kʰp	ខ្ពស់	[kʰpuəh]	“tall”
kb	ក្បាល	[kbaːl]	“head”
kʰt	ខ្នះ	[kʰtɛəh]	“pan”
kd	ក្តាន់	[kdan]	“dear”
kʰc	ខ្ចី	[kʰcej]	“borrow”
kʰɲ	ខ្ញុំ	[kʰɲum]	“I (first person singular)”
kŋ	ក្ងាន	[kŋaːn]	“goose”
kʰw	ឃ្វាល	[kʰwiːəl]	“to tend (animal)”
kʰs	ខ្សែ	[kʰsaːɛ]	“rope”
kʰl	ឃ្លី	[kʰliː]	“billiard ball”
kr	ក្រែ	[krɛː]	“bed”
kʰj	ខ្យល់	[kʰjɑl]	“wind”
mpʰ	ម្ភៃ	[mpʰej]	“twenty”
mt	ម្ទេស	[mteːs]	“chili”
md	ម្តាយ	[mdaːj]	“mother”

mc	ម្ចាស់	[mcas]	“owner”
mk	ម្កាក់	[mkaʔ]	“otaheite apple”
mk^{h}	ម្ខាង	[mk^{h}aːŋ]	“one side”
mn	ម្នាក់	[mnɛaʔ]	“one person”
mɲ	ម្ញីកម្ញ៉ក់	[mɲeʔ.mɲaʔ]	“coquettish”
mŋ	ម្ង៉ៃ	[mŋaj]	“one day”
ms	ម្សៅ	[msaw]	“powder”
ml	ម្លូ	[mluː]	“betel (leaf)”
mr	ម្រេច	[mric]	“pepper”
mj	ម្យ៉ាងទៀត	[mjaːŋ.tiːɜt]	“moreover”
mh	ម្ហូប	[mhob]	“food”
sp	ស្ពឺ	[spœː]	“tooth of a gear wheel”
sb	ស្បៃង	[sbaːoŋ]	“sack”
st	ស្តាំង	[stɛaŋ]	“falcon”
sd	ស្តាប់	[sdap]	“to listen”
sk	ស្កាត់	[skat]	“to intercept”
sm	ស្មា	[smaː]	“shoulder”
sn	ស្នា	[snaː]	“crossbow”
sɲ	ស្ញា	[sɲaːo]	“motionless because of illness”
sŋ	ស្ង៉ា	[sŋaːo]	“to boil”

sw	ស្វាយ	[swaːj]	“mango”
sl	ស្លាប់	[slap]	“to die”
sr	ស្រា	[sraː]	“alcohol”
sj	ស្យាម	[sjaːm]	“Siem (Thailand)”
k^hm	ខ្មោច	[k^hmaːoc]	“ghost”
k^hn	ខ្ញើយ	[k^hnaːəj]	“pillow”
lp	ល្ពៅ	[lpəw]	“pumpkin”
lb	ល្បី	[lbej]	“famous”
lk	ល្ងឹកណា	[lkɨkˑnaː]	“when”
lk^h	ល្ខោន	[lk^haːon]	“theater”
lm	ល្មម	[lmɔːm]	“suitable”
lŋ	ល្ងង់	[lŋɔŋ]	“stupid”
lw	ល្វា	[lwiːə]	“kind of fig tree”
lh	ល្ហុង	[lhoŋ]	“papaya”

Table 4-1: All possible set of consonant clusters with 2 consonants in Khmer language

Most of Khmer words contain only one syllable and only a little number of words has more than one syllable. This makes a general conclusion that Khmer is a mono-syllable language. Generally, Khmer words contain at most four syllables. The words with more than four syllables are usually words from Bali and Sanskrit. Below are examples of words which have different number of syllables ranging from one to six.

1 syllable : [kbaːl] “head”, [p^hhnejk] “eyes”
2 syllables : [cɑŋˑkaː] “chin”, [krɑːcɑːk] “nail”
3 syllables : [ʔɑmˑbaŋˑmeɲ] “in a moment”, [k^hjaːˑdɑmˑrej] “scorpion”
4 syllables : [ʔɑmˑpəlˑʔɑmˑpɛːc] “firefly”

5 syllables : [tʰa:.nɑn.taʔ.raʔ.saʔ] “rank”, [saʔ.kɑl.vət. tʰji:ə.laj] “university”

6 syllables : [wiʔ.ca:.rəʔ.naʔ.kəʔ.tha:] “editorial”

IV.2. Syllabification rules

For words with one syllable, there is no problem concerning syllabification. For words with more than one syllable, a window is defined. To detect syllable boundary in a word, we scan the word from left to right using the window which we will define shortly. In Khmer language, four forms of window are identified. They are: …CVCV…, …CVCCV…, …CVCCCV…, and …CVCCCCV…. Thus the syllabification should be defined based on each of these forms.

A general rules:

(1) When two identical consonants appear next to each other, one of them is considered as coda of the previous syllable and another is onset of the next syllable.
(2) The onset contains at least one consonant.
(3) The maximum number of consonant found in the onset is three.
(4) The coda contains at most one consonant.
(5) If the nucleus is a short vowel, then the coda must contain one consonant.
(6) All these consonants /ɲ ŋ n j w/ cannot be the first consonant of the cluster.
(7) Consider …C_1VC_2…, if C_2 is one of the nasal consonants / ɲ ŋ n m/, then C_2 must be coda of that syllable.

i. Form … CVCCCCV …

We can find syllable boundary of this form based on rule (3) presented above. Thus the …CVCCCCV… will be syllabified into …CVC.CCCV… Here is an example of the words of this form:

[**sahstra:**ca:] “professor” → [**sah.stra:**ca:]
[saŋskret] “Sanskrit” → [sa:ŋ.skret]

ii. Form …CVCV…

We can find syllable boundary of this form based on rule (2) presented above. Hence, the form …CVCV… will be syllabified into …CV.CV… Here are some examples of the words of this form:

[kɑ:kɑ:] “sediment” → [kɑ:.kɑ:]

[kɑ:ŋi:ə] "job" → [kɑ:.ŋiə]

iii. Form …CVCCV…

Possible syllabifications of this form are: …CV.CCV… and …CVC.CV…Usually, we can apply the general rules to syllabify most of the words that have this form.

We have noted that in Khmer language there are a possible set of consonant that can be clustered as shown in **Table 4-1**. So, if the CC can be a consonant cluster, then the CC will be the onset of the next syllable i.e. the syllabification of this form is …CV.CCV… else the syllabification is …CVC.CV…

Examples:

(5) [capdɑ:ʔ] "pull out" → [cap.dɑ:ʔ]

(6) [si:ɜwp^{h}əw] "book" → [siɜw.p^{h}əw]

(7) [kɑ:ŋmi:əh] "golden bracelet" → [kɑ:ŋ.mi:əh]

[kambet] "knife" → [kam.bet]

(Valid consonant clusters)

[kɑ:**k^{h}w**ɔ:k] "dirty" → [kɑ:.**k^{h}w**ɔ:k]

[nę:**tr**a:] "eyes" → [nę:.**tr**a:]

[pi:**pr**uəs] "because" → [pi:.**pr**uəs]

(Invalid consonant clusters)

(5) [k^{h}e**tk^{h}**ɔ:m] "try" → [k^{h}e**t.k^{h}**ɔ:m]

[dɑ:**kk^{h}**em] "decorative plant" → [dɑ:**k.k^{h}**em]

iv. Form …CVCCCV…

Possible syllabifications of this form are: …CV.CCCV… and …CVC.CCV… By using the general rules, we can syllabify some of the words whose form is …CVCCCV… Additionally, we can use the possible set of clustering consonant shown in **Table 4-1**. So, if the CCC is one of these **/str, skr/**, then the syllabification must be …CV.CCCV… Else, the syllabification must be …CVC.CCV…

Examples:

(2) [ʔɔːostraːliː] “australia” → [ʔɔːo.straː.liː]
[kamp^{h}laːəŋ] “gun” → [kam.p^{h}laːəŋ]

IV.3. Conclusion

Many words with deferent number of syllables have been analyzed and we have found rules necessary to detect syllabification boundary of Khmer words. The test has not yet done to determine accuracy of theses rules. We will do the accuracy test in the next phase.

V. Stress Analysis

V.1. Introduction

A stressed syllable is usually produced by pushing more air out of the lungs in on syllable relative to others. A stressed syllable thus has greater respiratory energy than neighboring unstressed syllables. It may also have an increase in laryngeal activity. Stress can always be defined in terms of something a speaker does in one part of an utterance relative to another [1].

It is difficult to define stress from a listener’s point of view. A stressed syllable is often, but not always, louder than an unstressed syllable. It is usually, but not always, on a higher pitch. Then most reliable thing for a listener to detect is that a stressed syllable frequently has a longer vowel than that some vowel would be if it were unstressed. But this does not mean that all long vowels are necessary stressed [1].

In this section we will talk about how to detect the stress in Khmer language and what are the properties for Khmer stress. The stress in Khmer language basically depends on the syllables which have a longer vowel. There would have more than one stress in a word but mostly of Khmer worlds are stressed on the last syllable [2]. We found that in Khmer language there is not only one stress on the last syllable but also on others syllables.

V.2. Stress for Khmer language

Khmer language is similar to French language regarding stress. Words always receive stress at the last syllable [2]. For example, in French language, the words “travailler [tra.va.ˈje] (to work), desiˈrable [de.zi.rabl] (preferable)” and in Khmer language, the words “ជីតា [ciː.ˈtaː] (grandfather), ឪពុក [ʔəw.ˈpok] (father)”.

In some longer words, it might seem as if there is more than one stress in the single word. For example, the word “ប្រជាប្រិយ [praː.ˈciː.ə.ˈprej] (popular), ទូលំទូលាយ [tuː.ˈlum.tuː.ˈliː.əj] (large)”, we see that the first word is stressed on the second syllable “ciːə˝ and the third syllable

“prej” and the second word, it is stressed on the second syllable “lum” and the fourth syllable “li:əj”.

It is not easy to detect stressed syllables in the word from the point of view of the listener. In Khmer language the detection of stressed syllables in the word is based on the vowel’s duration of the stressed syllable and the unstressed syllable.

In general we found that vowel in unstressed syllable is about 0.1 ms shorter than the one in stressed syllable.

Vowel	Words	Unstressed Syllable(ms)	Stressed Syllable(ms)
◌ា [a:]	ការងារ[k**a:.**ˈŋiə](job) / កា[ˈk**a:**](cup)	0.068	0.163
◌ី [əj]	កីឡា[k**ej**.ˈla:](sport) / ថ្មី[ˈthm**ej**](new)	0.071	0.122
◌ែ [e:]	រ៉ែ[rɨ:.ˈrɑ:] / ការែ[ka:.ˈrɨ:](sqare)	0.139	0.258
◌ី [i:]	ជីតា[ci:.ˈta:](granfather) / បញ្ជី [baɲ.ˈci:] (list)	0.213	0.112

In fluent speech, vowel in unstressed syllable usually is reduced to /ə/ and the coda in that same syllable, if it exists, is deleted.

Examples:

ការែម [ka:.ˈre:m] (ice cream)	→	[kə.ˈre:m]
សង្រ្គាម [sɑŋ.ˈkri:əm] (war)	→	[sə.ˈkri:əm]
សៀវភៅ [si:ɜw.ˈp^{h}əw] (book)	→	[sə.ˈp^{h}əw]
ទំលាប់ [tum.ˈloap] (hobby)	→	[tə.ˈloap]

ꯇꯨꯝ [tum.ˈnę:] (free) → [tə.ˈnę:]

1. **Word of one syllable:**

   The word which only contains one syllable has stress on itself. For example: ˈka: “cup”, ˈta: “grandfather”, ˈkom “not”, ˈpʰsa: “market”, ˈkbal “head”, ˈkʰɲom “I”, ˈkʰna:əj “pillow”, ˈpʰtɛah “house”.

2. **Word of two syllables:**

   The word which contains two syllables, the stress is on the last syllable of the word. For example: ka:.ˈre:m “ice cream”, sa:.ˈla: “school”, kam.ˈbet “knight”, paʔ.ˈtʰɑ:m “firstly”, kam.ˈpʰlɨ:ŋ “gun”.

3. **Word of three syllables:**

   The word which contains three syllables can be a combination of two words of one syllable and two syllables or only one word with three syllables.

   For the former case (combination of two words), the stress is determined by the stress in the constituent words. For example:

   tɑm.ˈrɑ:ŋ “filter” + ˈno:m “urine” → tɑm.ˈrɑ:ŋ.ˈno:m “kidney”
   ʔɑm.ˈbəl “salt” + ˈmte:h “pepper” → ʔɑm.ˈbəl.ˈmte:h “salt mixed with pepper”
   ˈko:l “base” + kum.ˈnɨt “idea” → ˈko:l.kum.ˈnɨt “idea”

   For the latter case (single word), the stresses are on the first and last syllable. For example: ˈtʰaʔ.wiʔ.ˈka “money”, ˈʔaʔ.nuʔ.ˈwoat “pratice”, ˈsap.pə.ˈda: “week”

4. **Word of four syllables:**

   The word containing four syllables has stresses on the second and last syllable. For example: ʔɑm.ˈpɨl.ʔɑm.ˈpę:c “firefly”, kru:ˈli:ŋkru:ˈlɔoŋ “kind of gray and black bird”.

5. **The word contains more than 4 syllables:**

   The word contain more than four syllables are the new words or it is taken from *Bali*. The stresses are on every two syllable in the word i.e. 2nd, 4th, 6th ... syllable. If the number of syllables is odd, the last syllable is also stressed. For example: sa:.ˈkɑl.wi.ˈtʰji:ə.ˈlaj “university”, sa:.ˈkɑl.pʰi:ə.ˈwu:.paʔ.ˈni:.jəʔ.ˈkam “globalization”

V.3. Conclusion

Every word of Khmer language has at least one stress at the last syllable. Word with one syllable has stress. Khmer word is monosyllable word. The word with two or more than three syllables has stresses on every two syllable and the last syllable always gets stress. For word with three syllables the stresses are on the first and the last syllable. Finally, for word which is the combination of other words, the stresses are identified based on the constituent words.

VI. Intonation Analysis

VI.1. Introduction

The intonation is for expressing different emotions of the speaker felling, for example: excitement, anger, suspicion, fear, sad, sarcasm, emphasis, speaking rate…etc. The intonation is defined as the pitch movement in spoken utterances. The intonation contour is defined as the rise and fall of pitch throughout the utterance [3].

The intonation is depending on the type of sentences. It goes up and down to make the sentence and its sense more comprehensible. For example, the question should go up at the end of the sentence so the listeners understand well that it is a question. If every sentence is pronounced in the same way, it is difficult to know that it is a question or normal sentence. Sometime it could be boring to listen to a talk which the intonation is not well used.

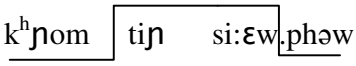
This part focus on the intonation of Khmer sentences, such as declarative, question, interrogation…etc. We do not research on the intonation with the difference felling of the speaker.

VI.2. Intonation for Khmer language

There are several types of sentences in Khmer language such as declaration, exclamatory, interrogative, imperative, and parenthesis. This section shows about how the intonations of these sentences are produced.

VI.2.1. Declaration

i. Sentence with one rhythm:



I buy book
"I buy a book."

kʰɲom mɨn cɑŋ.ba:n tę:
I not want no
I don't want to get it.

ii. Sentence with more than one rhythm:

cun.piʔ.ka: kɑm.puc.ci:ə nɨŋ mi:ən sɑŋ.kʰɨm ci:ə.nic
Immobility Cambodian will have hope always

Cambodian immobility will always have the hopes.

VI.2.2. Exclamatory

i. Sentence with one rhythm:

kdaw nah
hot too
It's too hot!

ii. Sentence with more than one rhythm:

si:ɛw.phəw nih tʰlaj nah !
book this expensive very
This book is very expansive!

VI.2.3. Interrogative

i. Full doubt question: It is semilar to yes-no question in english language.

nɛaʔ cɑŋ təw lę:ɲ kɑm.pɔŋ sa:om tę: ?
you want go play Kampongsom no
Do you want to go to Khmpongsome?

nɛaʔ jul tę: ?
you understand no
Do you understand?

ii. Question with some doubt:

ɳɛaʔ tʰwə: ka: pon.ma:n tʰŋaj mu:əj ʔa:.təʈ ? ba:n ponma:n ?
You do work how many day a week get how much
How many days do you work per week? How much do earn?

VI.2.4. Imperative

cɔ:ol mɔ:k !
come in
Come in!

cʰup trɑŋ nəŋ haəj !
Stop there
Stop there!

VI.2.5. Open and close sentences

i. Close sentence:

cɔ:ol mɔ:k kɔ:on !
come in child
Come in boy!

ii. Open sentence:

ma:.nɨt cʰup trɑŋ nəŋ haəj !
Manith stop there
Manith stop there!

VI.2.6. Parentesis

koat prɑ:.ha:ɛl.ɟi:ə mɨn mɔ:k tɛ: ta:m ɲom sma:n pi:.pruəh koat mɨn sɔov sruəl kʰlu:ən
He maybe not come as I guess because he do not have good heath
He maybe does not come, I guess, because he is not well!

VI.3. Conclusion

The intonation can change the stress on any word in the phrase. In Khmer language, the intonation is normally goes down at the end of the sentence except, the interrogative with full doubt question and exclamatory which goes up at the end of the sentence.

VII. Sound Change Rules

VII.1. Introduction

Sound change includes any processes of language change that affect pronunciation (phonetic change) or word structures (phonemic change). Sound change can consist of the replacement of one speech sound (or, more generally, one phonetic feature) by another, the complete loss of the affected sound, and (rarely) even the introduction of a new sound in a place where there previously was none. Sound changes can be environmentally conditioned, meaning that the change in question only occurs in a defined sound environment, whereas in other environments the same speech sound is not affected by the change.

Sound change is assumed to be usually regular, which means that it is expected to apply mechanically whenever its structural condition is met, irrespective of any non-phonological factors (such as the meaning of the words affected). On the other hand, sound changes can sometimes be sporadic, affecting only one particular word or a few words, without any seeming regularity [4].

In this section, we will present sound change rules in Khmer that we have identified. We use "." to represent syllable boundary and "#" word boundary. We also use brace "{, }" to represent alternative i.e. {a, b} represents a or b.

VII.2. Sound Change Rules

1. Final obstruent devoicing: voiced obstruent in the syllable coda or at the end of a word becomes voiceless. In Khmer, there are only 2 voiceless consonants which have their voiced counterparts: [p] and [t] and their voiced counterparts [b] and [d].

   b → p / _{., #}
   d → t / _{., #}
   [+voice] [-voice]

cab	→	cap	'to catch'
məʔ.kod	→	məʔ.kot	'crown'

2. Final obstruent unaspirating: aspirated obstruent in the syllable coda or at the end of a word becomes unaspirated. In Khmer, there are only 4 voiceless unaspirated consonants that have their voiceless aspirated counterparts: [p, t, c] and [k] and their aspirated counterparts [p^h, t^h, c^h] and [k^h]. Note that [c^h] never occurs in the syllable coda.

   t^h → t / _{., #}
   p^h → p / _{., #}
   k^h → k / _{., #}
   [+spread] [-spread]

prɑ:.ma:t^{h}	→	prɑ:.ma:t	‘to look down’
rum.lo:p^{h}	→	rum.lo:p	‘to rape’
muk^{h}	→	muk	‘face’

3. Initial obstruent devoicing: voiced obstruent at the beginning of consonant cluster in the onset becomes voiceless.

b	→	p / {., #}_C
d	→	t / {., #}_C
[+voice]		[-voice]

bra:ə	→	pra:ə	‘to use’
bdɔ:o	→	pdɔ:o	‘to change’
bʔɔ:on	→	pʔɔ:on	‘younger’

blɑn →plɑn →p^{h}lɑn ‘to rob’

(With rule number 4 below: plɑn → p^{h}lɑn)

We have only one word in which [d] is at the beginning of the consonant cluster.

dbet	→	tbet	‘because of’

4. Initial obstruent aspirating: unaspirated obstruent [p] at the beginning of the consonant cluster in the onset followed by the consonants [k, n, j] or [l] becomes aspirated.

p	→	p^{h}/ {., #}_{k, n, j, l}
[-apirated]		[+aspirated]

pkol	→	p^{h}kol	‘a type of plant with small yellow flowers’
pnɔ:ŋ	→	p^{h}nɔ:ŋ	‘the Phnong, a Mon-Khmer of north-eastern Cambodia’
pjuh	→	p^{h}juh	‘tempest’
pluk	→	p^{h}luk	‘cremator’

5. Initial obstruent aspirating: unaspirated obstruent [t] at the beginning of the consonant cluster in the onset followed by the consonants [n, m, l, w] or [j] becomes aspirated.

t	→	t^{h} / {., #}_{n, m, l, w, j}
[-apirated]		[+aspirated]

da:əm.tna:ot	→	da:əm.t^{h}na:ot	‘palm tree’
tma:t	→	t^{h}ma:t	‘vulture’
tlok	→	t^{h}lok	‘clown’
twi:ə	→	t^{h}wi:ə	‘door’
ni:.tja:.nuʔ.kɔ:ol	→	ni:.t^{h}ja:.nuʔ.kɔ:ol	‘legitimate’

6. Initial obstruent aspirating: unaspirated obstruent [k] at the beginning of the consonant cluster in the onset followed by the consonants [m, w, t, s, n] or [l] becomes aspirated.
   k → k^h / {., #}_{m, w, t, s, n, l}
   [-apirated] [+aspirated]

kme:ŋ#broh	→	k^hme:ŋ#broh	‘little boy’
kwa:en	→	k^hwa:en	‘intelligent’
kt*ɑ*:m	→	k^ht*ɑ*:m	‘a kind of tree’
ksat	→	k^hsat	‘king’
knoŋ	→	k^hnoŋ	‘in’
klaŋ	→	k^hlaŋ	‘strong’

7. Initial obstruent aspirating: unaspirated obstruent [c] at the beginning of the consonant cluster in the onset followed by the consonants [n] or [w] becomes aspirated.
   c → c^h / {., #}_{n, w}
   [-apirated] [+aspirated]

cnaj	→	c^hnaj	‘to polish gems’
d*ɑ*m.lɔ:oŋ#cwiə	→	d*ɑ*m.lɔ:oŋ#c^hwiə	‘a kind of sweet potato’

8. Initial obstruent unaspirating: aspirated obstruent [p^h] at the beginning of the consonant cluster in the onset followed by the consonants [d] or [ʔ] becomes unaspirated.
   p^h → p / {., #}_{d, ʔ}
   [+aspirated] [-aspirated]

p^hdac	→	pdac	‘to cut off’
p^hʔa:ʔ	→	pʔa:ʔ	‘to postpone’

9. Initial obstruent unaspirating: aspirated obstruent [c^h] at the beginning of the consonant cluster in the onset followed by the consonants [ʔ] becomes unaspirated.
   c^h → c / {., #}_ʔ
   [+aspirated] [-aspirated]

c^hʔen	→	cʔen	‘to be cooked’
c^hʔəŋ	→	cʔəŋ	‘bone’

10. [s] in the syllable coda or at the end of a word is changed to [h].
    s → h / _{., #}

das	→	dah	‘to wake someone up’
cas	→	cah	‘old’

11. [k] in the syllable coda or at the end of a word preceded by [ɑ, ɔ, a, ɑ:, ɔ:, a:, e, ɨə, iɜ, iə] or [ɛa] is changed to [ʔ].

k → ʔ / {ɑ, ɔ, a, ɑ:, ɔ:, a:, e, ɨə, iɜ, iə, ɛa}_{., #}

kɑk	→	kɑʔ	‘to book’	
p^hɔk	→	p^hɔʔ	‘mud’	
prak	→	praʔ	‘money’	
kɑ:k	→	kɑ:ʔ	‘to freeze’	
rɔ:.lɔ:k	→	rɔ:.lɔ:ʔ	‘wave (water)’	
k^ha:k	→	k^ha:ʔ	‘to spit’	
kɑ:.k^hlek.kɑ:.k^hlak	→	kɑ:.k^hleʔ.kɑ:.k^hlaʔ	‘unorganized’	
baɲ.ca:.nɨək	→	baɲ.ca:.nɨəʔ	‘a kind of dessert’	
trɑ:.ciɜk	→	trɑ:.ciɜʔ	‘ear’	
ŋi:ək#kra:oj	→	ŋi:əʔ#kra:oj	‘to turn back’	
rɛak	→	rɛaʔ	‘shallow’	

For Khmers Kampuchea-Krom (indigenous ethnic Khmer minority living in southern Vietnam), this rule is not applied. They pronounce the final [k] as [k] not [ʔ].

12. [k] in the syllable coda or at the end of a word preceded by [e:, ę:] or [a:e] is changed to [c].

k → c / {e:, ę:, a:e}_{., #}

ce:k	→	ce:c	‘banana’
cra:ən#pę:k	→	cra:ən#pę:c	‘too much’
ca:ek	→	ca:ec	‘to divide’

13. [ŋ] in the syllable coda or at the end of a word preceded by [e:, ę:, ɛ:, i:] or [a:e] is changed to [ɲ].

ŋ → ɲ / {e:, ę:, ɛ:, i:, a:e}_{., #}

pre:ŋ	→	pre:ɲ	‘oil’
kę:ŋ	→	kę:ɲ	‘to sleep’
wɛ:ŋ	→	wɛ:ɲ	‘long’
mi:ŋ	→	mi:ɲ	‘aunt’
ka:eŋ	→	ka:eɲ	‘angle’

14. r-deletion: [r] in the syllable coda is deleted

r → **ø /** _{., #}

da:ər	→	da:ə	‘to walk’
ka:r	→	ka:	‘to marry’

15. m-deletion: [m] at the beginning of consonant cluster in the coda preceded by [a] and followed by [ŋ] is deleted.

m → **ø** / a_ŋ{., #}

pramŋ	→	praŋ	‘dry season’
cramŋ	→	craŋ	‘bank (river)’

16. Consonant at the end of consonant cluster in the coda is deleted.

C → **ø /** C_{., #}

pe:tj	→	pe:t	‘medical doctor’
br*a*:.pɔnt^{h}	→	br*a*:.pɔn	‘wife’

This rule can apply several times.

sastr	→	sast	→	sas	‘science’

Note that rule 15 must be applied before this rule.

17. [i] in close syllable is changed to [ɨ].

i → ɨ / _C{., #}

kit	→	kɨt	‘to think’
cit	→	cɨt	‘near’

But in Siem Reap province, north-west of Cambodia, this rule is not applied. People in Siem Reap pronounce the above words as [kit] (to think) and [cit] (near).

18. [e] in close syllable is changed to [ə].

e → ə / _C{., #}

ken	→	kən	‘to grind’
cet	→	cət	‘to slice’

We have found only one word in which /e/ in close syllable does not change to /ə/:

* k*a*:.k^{h}leʔ.k*a*:.k^{h}laʔ → k*a*:.k^{h}ləʔ.k*a*:.k^{h}laʔ ‘unorganized’

All the above rules can be applied in any order and as many times as possible except rules 15 and 16 which are to be applied in this order: rule 15 first and then rule 16. The output of one rule is taken as the input of another rule.

Example:

			bre:ŋ	‘oil’			
Rule (13)	bre:ŋ	→	bre:ɲ	Rule (3)	bre:ŋ	→	pre:ŋ
Rule (3)	bre:ɲ	→	pre:ɲ	Rule (13)	pre:ŋ	→	pre:ɲ
			bramŋ	‘dry season’			
Rule (3)	bramŋ	→	pramŋ	Rule (15)	bramŋ	→	braŋ
Rule (15)	pramŋ	→	praŋ	Rule (3)	braŋ	→	praŋ

For bramŋ ‘dry season’, if we apply rule 16 before rule 15 it will give incorrect result:

*Rule (16) bramŋ → bram

VII.3. Conclusion

Many rules have been identified for sound change in Khmer. These rules can be apply several times and in any order except rule 15 and 16 which has to respect this order: rule 15 first and then rule 16. It is important to know some rules for those who learn foreign languages like English or French. If they understand these rules, they can pronounce quickly and correctly words in these foreign languages. For example the rule that changes /s/ to /h/ if it is in the coda. In Khmer /s/ in the coda is always pronounced as /h/. So if they are not aware of this rule, they will also apply this rule for these foreign languages i.e. they will pronounce /beis/ as /beih/ “base”.

VIII. Results and conclusion

The fundamental approach of our Khmer TTS system is diphone concatenation. The study of Khmer phonemic inventory leads to the identification of a list of diphones. The identification of acoustic cues of vowels and consonants is very helpful for diphone splitting and sound corpus tagging.

From the research on syllabification of Khmer words, a set of rules have been defined to predict syllable boundary of a given word. The accuracy of these rules will be tested in the next phase.

Research on stress analysis in Khmer has found that Khmer word has stress. For words with one or two syllables, we have only one stress and it is at the last syllable. For words with three syllables, stresses are on the first and last syllable. For words with more than three syllables,

stresses are on every two syllable (2^{nd}, 4^{th}...) and the last syllable is also stressed. For words which are combination of other words, stressed are defined based on the stresses of the constituent words.

Many rules have been identified to do sound change in Khmer. These rules can be used to do letter-to-sound conversion, which is one of the main modules in TTS system.

VIII. References

[1]. Petter Landefoged, A Course in Phonetics, fourth edition, Replica Press Pvt. Ltd., 100% EOU, Delhi 110 040, 2001.

[2]. Dr. Prum Mul, Fundamental of Phonetics and Phonology, Phnom Penh, Cambodia, 2003.

[3]. Rafia Bokhari, Intonation, Center for Research in Urdu Language Processing, National University of Computer and Emerging Sciences, Lahore, Pakistan.

[4]. http://en.wikipedia.org/wiki/Sound_change, Feb 10 2008.

[5]. Chun Nat, Khmer Dictionary, fifth edition, Phnom Penh, Cambodia, 1967.

[6]. Khen Sok, Khmer Grammar, first edition, Phnom Penh, Cambodia, 2007.

[7]. Carlo Gussenhoven and Haike Jacobs, Understanding Phonology, second edition, Oxford University Press, Inc., 198 Madison Avenue, New York, NY 10016, 2005.

[8]. J.M. Pickett, The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology, A Viacom Company, 160 Gould Street, Needham Heights, MA 02494, 1999.

[9]. Asanka Wasala and Kumudu Gamage, Research Report on Phonetics and Phonology of Sinhala, Language Technology Research Laboratory, University of Colombo School of Computing, Sri Lanka.