Rendering Engine For
Khmer Script in Mobile Phone

Version 0.1

Developer: Mr. NETH Sovathena
Team Leader: Mr. HENG Samedi

20 September 2008

Cambodia Country Component
PAN Localization Project
PAN Localization Cambodia (PLC) of IDRC

Table of Contents

1 Introduction

Khmer Script, also known as Aksar Khmer (Khmer Letters), is the official script of the Kingdom of Cambodia. Khmer Script is a complex script; the order of writing is not the same as the order of its spelling. For example, the spelling of the word គេ [ke] (they) is ក + េ. In general, it is written from left to right and downwards when horizontal space runs out [1]. Khmer alphabet contains Consonants, Subscripts, Independent Vowels, Dependent Vowels, Various Sign, Digits, Khmer Digit for divination lore, and other symbols.

This report is about the rendering engine for Khmer script by using Bitmap font for smart typing method (write as spelling) in mobile phone. It is not the rendering for the hold system of mobile phone to support Khmer script but it is the third-party application to support Khmer script.

2 Stat of the art

2.1 Khmer Font support

The available of Khmer font in computer are all in TTF (True Type Font) and OTF (Open Type Font). The TTF font, such as Limon, ABC…etc, does not support the Unicode and also does not support the smart typing. Microsoft had developed Rendering Engine supporting Khmer Script covered by the Unicode Standard. This engine is known as Khmer Open Type shaping engine, use the Open Type Font, which is the Windows component responsible for text layout [2].

2.2 Mobile phone support Khmer script

Recently, Nokia has presented ten models of Khmer supported mobile [3]. They are Khmer Script built-in but the display is still not good enough for viewing the language based on the writing system of Khmer. For example, some characters are displayed in wrong position or overlap others. Below is a list of models of Khmer built-in mobile from Nokia.

N°	Model
1	Nokia 7100 Supernova
2	Nokia 7070
3	Nokia 5130 Xpress Music
4	Nokia 5000
5	Nokia 2760
6	Nokia 2680 Slide
7	Nokia 2630
8	Nokia 2600
9	Nokia 1680
10	Nokia 1650

Table 1: Model of mobile phone support Khmer script

3 Khmer script analysis

The structure of Khmer writing is seen in 5 lines or 5 layers as the figure below.

Layer 1:	Above Vowel
Layer 2:	Consonant Shifter / Above Vowel / Various Sign
Layer 3:	Consonant / Left Vowel / Right Vowel/ Independent Vowel
Layer 4:	Below Subscript/ Below Vowel
Layer 5:	Below Vowel

Figure 1: Layers of Khmer Script

Khmer script can be grouped into 6 main categories: consonants, consonant shifter, subscripts, dependent vowels, Independent scripts, and various sign scripts.

3.1 Consonant

Consonant is generally known as the main character of every Khmer syllable [3] which stands in layer 3 of Khmer writing structure. It can be independent; however, it can be written with other types of Khmer script like subscript and/or dependent vowel. Furthermore, consonants can be classed into three series. First, the consonants such as ង, ម, រ, ឈ, វ, ញ are those lack light voice whereas second such as ហ, ស, ប, អ lack heavy voice. The rest of these are in the third series. The first series including ប can be used with MUUSIKATOAN (ö); the second can be used with TRIISAP (õ); while the third cannot be used with MUUSIKATOAN or TRIISAP at all.

3.2 Consonant Shifters

Consonant Shifters are the two Khmer sign: MUUSIKATOAN (ö) and TRIISAP (õ) that are located in layer 2 over some of Khmer consonants for shifting between two sounds— /qakhoosaq/ (voiceless) and /khoosaq/ (voiced). /qakhoosaq/ sounds –a while /khoosaq/ sounds –o.

- MUUSIKATOAN can stand over consonant of the first series (ង, ម, រ, ឈ, វ, ញ) to shift to sound

  –a, voiceless. For instance, ង៉, ម៉, រ៉, ឈ៉, វ៉, ញ៉.

- TRIISAP can stand over consonant of the second series (ហ, ស, ប, and អ) to shift to sound –o,

  voiced. For instance, ហ៊, ស៊, ប៊, អ៊.

3.3 Subscripts

Every Khmer consonant has its own subscript. Subscripts can be categorized into three small groups: below subscript, right subscript, and left subscript because of their positions render with their main consonant.
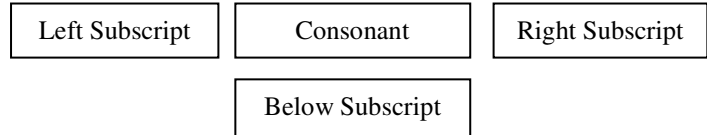
Left Subscript	Consonant	Right Subscript

	Below Subscript	

Figure 2: Khmer Subscript rendering relative to its consonant
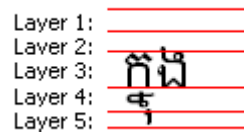
3.4 Dependent Vowel

Like its name, Dependent Vowel is Khmer character that depends on main consonant, i.e. it cannot stands alone while Independent Vowel can stand alone by itself. Dependent Vowels can be classified as five groups according to their position rendering with main consonant or subscript of the consonant. They are below vowel, left vowel, right vowel, above vowel, and two-part vowel.

- **Below Vowels** are those standing below main consonant or below subscript. Hence, they can locate in layer 4 or layer 5. There are only 3 characters of this type: ◌ុ,◌ូ,◌ួ.
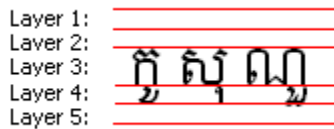
Below vowel in layer 4 under main consonant	Below vowel in layer 5 under a subscript of a consonant

Figure 3: Below vowel in Layer 4 and 5

- **Above Vowels** are those standing above main consonant or consonant shifter. They locate in layer 1 only if there is a consonant shifter stays in layer 2 over the consonant in layer 3. In Khmer script, there are 5 above vowels such as ◌ិ, ◌ី, ◌ឹ, ◌ឺ, and ◌ំ.

1. Above Vowel
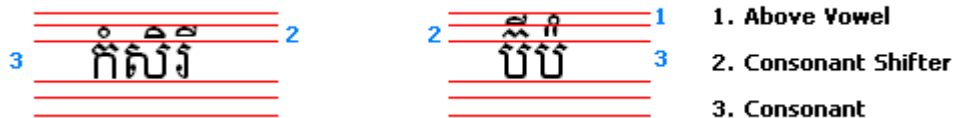2. Consonant Shifter
3. Consonant

Figure 4: Above vowels in layer 1 and 2

- **Left Vowels** are those standing on the left side of main consonant or left subscript. Only 3 characters of Khmer script are left vowel such as េ◌, ែ◌, and ៃ◌.
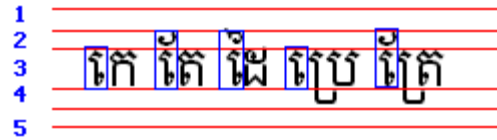
Figure 5: Left Vowels

-   **Right Vowels** are those standing on the right side of main consonant or right subscript. They locate in layer 3 with the consonant as well. They are ◌ា, ◌ះ, and ◌ៈ

-   **Two-part Vowels** are those standing on both right and left side of main consonant. They are ◌ើ, ◌ឿ, ◌ៀ, ◌ោ, and ◌ៅ.

3.5 Independent scripts

This category contains all Khmer independent characters including Khmer Independent Vowels, Khmer digits, Khmer numeric symbols for divination lore, currency symbol, etc.

3.6 Various sign

This category contains all Khmer dependent scripts such as ◌់, ◌៉, ◌៊, ◌់, ◌៌, ◌៍, ◌៎, and ◌ ̤ except vowels.

They cannot be used with consonant shifter. Thus, they can position in only layer 2 or layer 4.

4   Methodology

The rendering engine of Khmer Script for mobile phone is developed by using j2me, source code and library developed by a company named ESOCO (http://www.esco.net) under open source project. These projects are including microEWT, which is a graphical, event-driven user interface library for j2me; and another one is J2ME-Lib, a utility library for j2me applications [5].

microEWT has feature that allows the application to display and render font using an horizontal image in PNG format as a Bitmap font. The image, is transparence, contains all Khmer glyphs or shapes arranging horizontally with special markers on the above edge [5]. This is possible to customize the application to present custom font by designing our own image as Khmer font to replace the old one that is being used by microEWT.

4.1 Font design for mobile

As above mentioned, Bitmap font is in form of a horizontal image. In order to design Khmer bitmap font, we initially choose the shape of a font named *KhmerKep* since it is generally used among others in series of Khmer Unicode font. Moreover, KhmerKep has small and thin body that is the best for displaying Khmer shape on mobile screen.

According to the development of mobile phone, there are many mobile screen sizes such as 128x160, 176x220, 240x320, 800x480, 208x320, etc. However, the screen size 240x320 is selected. This results KhmerKep font should be in size of 14 px for the best displaying, i.e. its size is not too big and not too small. Ultimately, KhmerKep with size 14 px is preferred for the design of Khmer Bitmap font.

Our own custom font contains all usable Latin characters append with all characters and all possible ligatures of Khmer script.
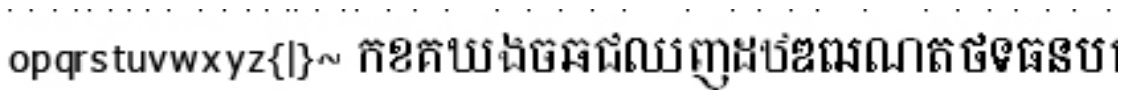
opqrstuvwxyz{|}~ កខគឃងចឆជឈញដឋឌឍណតថទធនបរ

Figure 6: Khmer Bitmap Font

4.2 Rendering engine

Due to Khmer language has complex writing system, rendering, which will be analyzed later in this section, is needed to do in order to display Khmer script correctly. Rendering means the process of displaying Khmer characters properly as their glyph or ligature in the correct position based on the characteristic of Khmer language writing. Likewise, rendering engine generates reordering or changing one shape to another. For example, a consonant meets a left vowel; the vowel will reorder its position to the left of the consonant like គ + ើ → គើ.

Khmer Script rendering can be defined into six basic features: Pre-base, Below-base, Above-base, Post-base, Two-Part base and ligature.

4.2.1 Pre-base feature

Left vowels and Left subscripts have rendering position in Pre-base feature. This feature causes these characters render by reordering their position.
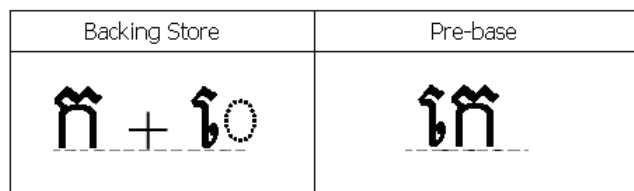
Backing Store	Pre-base
គ + ើ	គើ

Figure 7: Pre-base rendering

4.2.2 Post-base feature

Right Vowels and Right subscript have rendering position in Post-base feature, i.e. they locate on the right side of consonant. Furthermore, Right Vowel--Sign AA (ា)--will render to form ligature that will be explained below.

Figure 8: Post-base rendering

4.2.3 Above-base feature

Above Vowels, Consonant Shifter and Various Signs (except ̐) have rendering position in Above-base feature.

Figure 9: Above-base rendering

4.2.4 Below-base feature

Below Vowels, Below Subscripts, and Various sign 17D2 (̦) have rendering position in Below-base feature.

Figure 10: Below-base rendering

4.2.5 Two-Part base feature

This feature can be found when a character has either Pre-base or Post-base feature. Two-Part vowel is applied with this feature. Furthermore, two of Two-Part Vowel-- ើ and ៀ --also render to form ligature which will be explained soon.

Figure 11: Two-Part base rendering

4.2.6 Ligature

Khmer Unicode has some ligatures that are new shapes derived from combination of two characters. It means that two characters rendered by changing their shape to a new one. As general, Ligature of Khmer Unicode can be found only if consonant or right subscript meets left vowel ៀ or two-part vowel: ៀ, ៀ. This feature is illustrated as the figure below.

Figure 12: Ligature rendering

Note: Since Khmer Unicode has no unique code for each subscript, U+17D2 combines with each Unicode of consonant are used to identify subscript, i.e. subscript can be seen when U+17D2 meets a consonant. For example, 17D2 + ក → ្ក, 17D2 + ឡ → ្ឡ.

Figure 13: Subscript rendering
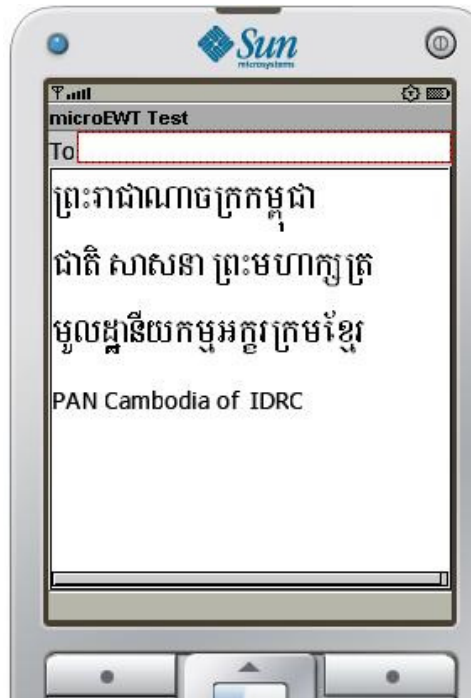
5   Conclusion and Result

Figure 14:First version of Khmer Script on mobile phone

In conclusion, Khmer script can be displayed properly unless it bases on a rendering rule which contains all Khmer rendering features. The development of Khmer rendering engine; therefore, is really playing an important role to be capable of supporting Khmer language in mobile devices.

6   References

[1]. Khmer, http://www.ancientscripts.com/khmer.html, 30/03/2009

[2]. Khmer OpenType specification, http://www.microsoft.com/typography/OpenType%20Dev/khmer/intro.mspx, 30/03/2009

[3]. Nokia KTH, www.kth.com.kh, 30/03/2009

[4]. Encoding Conversion Utility for Khmer, Chea Sok Huor, Atif Gulzar, Ros Pich Hemy, Neak Longchrea

[5]. ESOCO, http://www.esoco.net, 30/03/2009