Keyboard Layout and Input method for Khmer SMS

Developer: Miss. LACH Dalina
Team Leader: Mr. HENG Samedi

6 April 2008

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

Table of Contents

1 Introduction

Normally, users really want to write text fast and enjoy using the keypad which is easy to remember that, it can help them to write faster. In order to respond to this request, we have to consider carefully on assigning keypad layout and the input method for Khmer Unicode on mobile phone. Recently, there is only one keypad layout for Khmer Unicode on cell phone (*Figure 1*) which designed by Nokia and it uses the Multi-tap Method.

In this report, we will describe in detail why we decide to design new keypad layout and implement new input method for Khmer Unicode on mobile phone while there have already existed.

2 Keypad Layout

Due to the huge number of Khmer characters (172 characters); we need to be very careful before deciding to assign them into 12 keys (keypad standard for mobile phone).

2.1 Existing Keypad Layouts

The keypad layout that has been created by Nokia has some points to consider as bellow:

- Khmer vowel signs should not be assigned before consonants; however, consonants should be assigned first (start from key 1) because all Khmer words start from consonants not from vowels.
- ◌ុ (CUOENG) should not be assigned with dependant vowels because it is not easy to remember, however, it is easy to be confused. It ought to be assigned on key 0.
- Independent vowels (ឥ ឦ ឧ ឩ ឪ ឫ ឬ ឭ ឮ ឯ ឰ ឱ ឲ ឳ) are commonly used, so it is better not to assign them at key 0. Users may find it difficult to remember.
- It is no need to assign these (ក ប អ (17A3) អា ឱ ឳ ◌ា ◌ិ ◌ី ◌ឹ ◌ឺ ◌ុ ◌ូ ◌ួ ◌ើ ◌ឿ ◌ៀ ◌េ ◌ែ ◌ៃ ◌ោ ◌ៅ ◌ំ ◌ះ ◌ៈ ◌៉ ◌៊ ◌់ ◌៌ ◌៍ ◌៎ ◌៏ ◌័ ◌៑ ◌្ ◌៓ ◌។ ◌៕ ◌៖ ◌ៗ ◌៘ ◌៙ ◌៚ ◌៛ ◌ៜ ◌៝) while it is not used in Khmer language.
- Because of the absentee of the numeric symbols for divination lore in Nokia's keypad, the fortune teller can't use this phone for the purpose of his/her work.

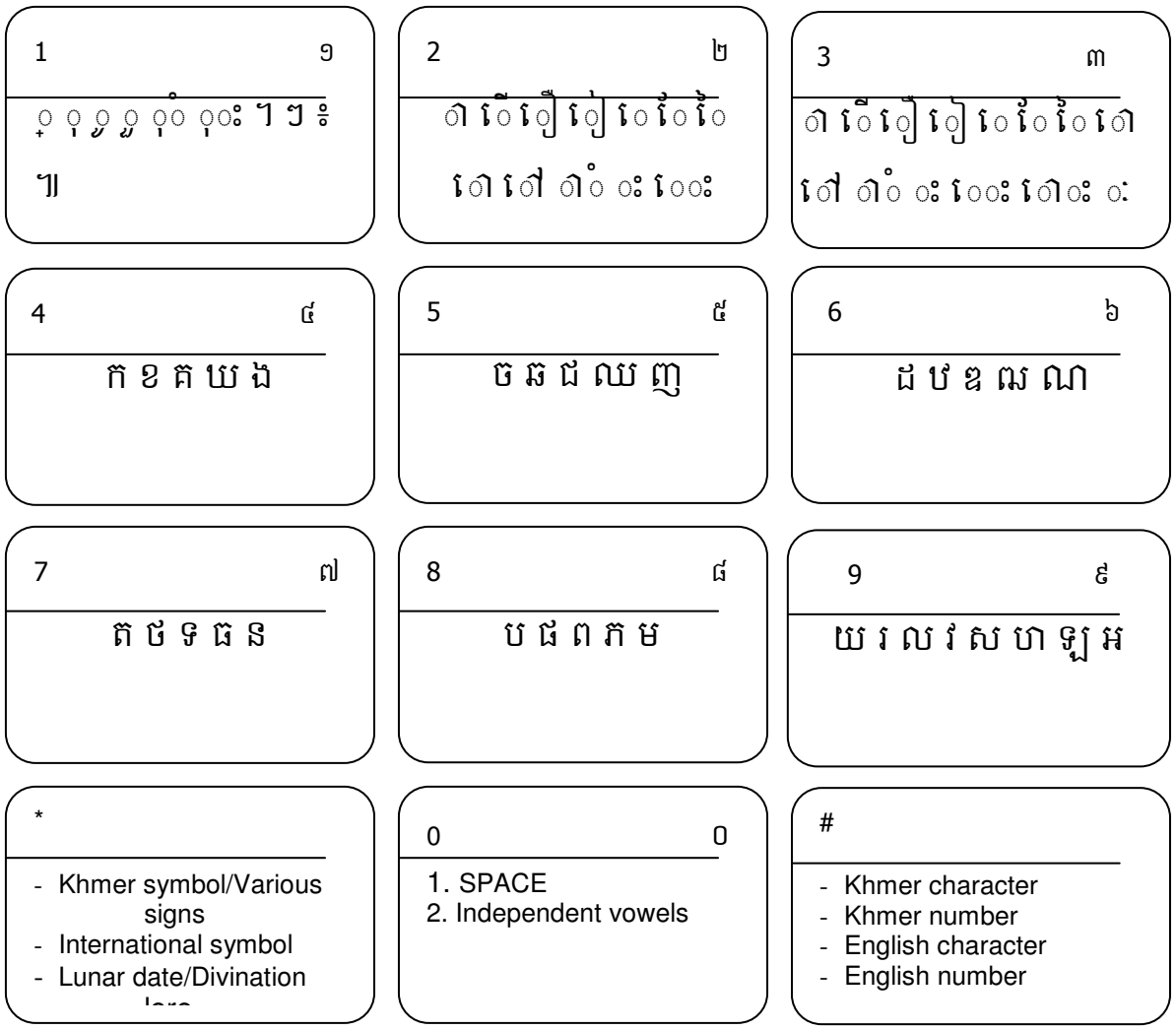

Figure 1: Keypad Layout for Khmer Language [1]

2.2 Suggested Keypad Layouts

We decided to choose this keypad layout (*Figure 2*) due to many reasons as following:

- We choose to assign at most 9 characters per key (from key 1 to 9) because it works with two-key press method (page 6) which requires maximum 9 characters.
- All consonants must be assigned before vowels because all Khmer works begin with consonants.
- Some various signs, independent vowels and currency symbol have to be assigned with consonant due to the huge number of Khmer characters (172 characters).

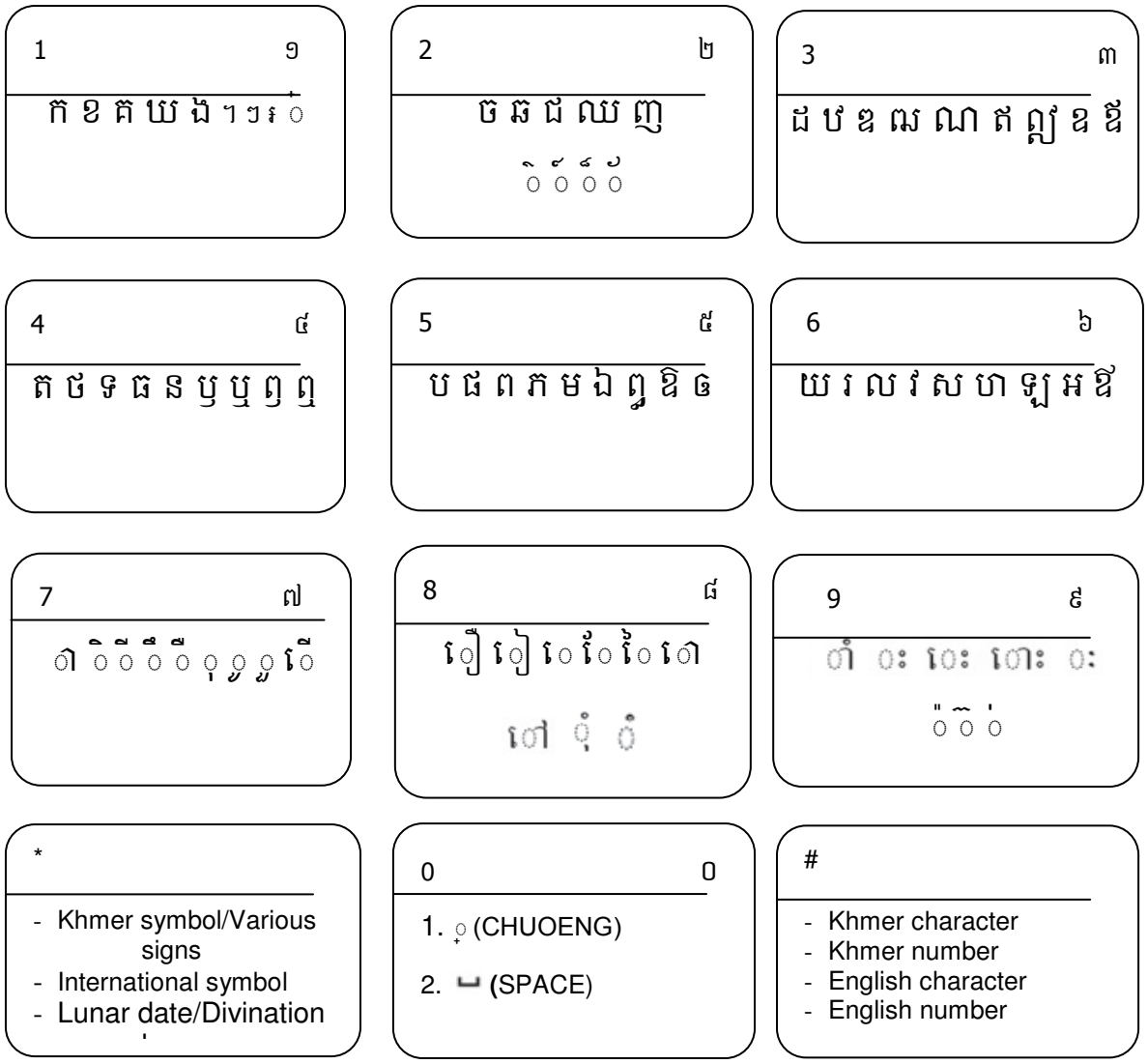

Figure 2: Suggested Keypad Layout for Khmer Unicode

3 Input Method

3.1 Existing Input Method

3.1.1 Multi-press Method

The multi-press methods require more than one keystroke to enter a character. These methods allows for unambiguous typing of characters. They can be used alone or as a fallback for systems using more complex text entry methods. The multi-press methods are well suited to type words not contained in the dictionary. [2]

a. Multi-tap Method

The first and still most common way to enter text on a mobile phone is the multi-tap method. In the case of English keyboard, since ‘a’, ‘b’, and ‘c’ share the same key, the user presses it once to enter an ‘a’, twice to enter ‘b’ and three times to enter a ‘c’. For example, to enter the word dog, the user presses the sequence of keys “36664”. As two consecutive characters of a word can share the same key, as for example the word “no”, where both ‘n’ and ‘o’ are assigned to 6, a timeout or kill-time-out key is needed to determine when to stop shifting the letters and display a new character. [2]

b. Two-key press Method

In the two-key press method, the user presses two keys successively to specify a character. The first key press, as in multi-press method, selects the group of the characters. The second press is to specify position of the character within the group. For example, to enter the character ‘K’, the user presses 5 and then 2. The two-key method is very simple. There are no timeouts or such kill-timeout key. Each character is entered with exactly two key presses. [3]

c. Less tap Method

The less tap is similar to multi-tap, however, it allows the entry of the most frequent letter on each key with one keystroke, the second most frequent letter with two keystrokes and so on.

3.1.2 Single-press Methods

The single-press methods try to reduce the Keystrokes per Character (KSPC) to roughly one. They resort to a dictionary as a means of resolving the ambiguity of the input.

a. T9

The T9 input method, patented by Tegic Communication, Inc, uses dictionary as the basis for disambiguation. The method is based on the same key layout as the multi-press method, but each key is pressed only once. For example, to enter “the”, the user enters the key 8-4-3-0. The 0-key, for the SPACE, delimits words and terminates disambiguation of the preceding keys. T9 compares the word possibilities to its linguistic database to guess the intended word. Naturally, linguistic disambiguation is not perfect, since multiple words may have the same key sequence. In these cases, T9 gives the most common words as a default. To select an alternate word, the user presses a special NEXT function key. [3]

c. iTAP

The iTAP system requires only one key touch to select a letter, and also proposes the next word you intend to add to your message or note. By using past text input and the context of where this text was entered, iTAP goes far beyond simply remembering unique words, slang and abbreviations and suggests entire sentences. [4]

d. eZitext and eZitap

The eZiText and eZiTap input method, developed by Zi Corp (www.zicorp.com), is a predictive text messaging along with word completion. It predicts and displays desired

candidate before all keys are pressed. These methods faced the same ambiguity problem as the T9 method.

e. The Predictive Text Entry Method

With the predictive text entry method, the user presses one key per character and the program matches the key sequence to words in a dictionary. Even if several characters are mapped to the same key, in many cases, only one word is possible given the sequence. This method makes it possible to reduce the KSPC to roughly 1. If the key sequence corresponds to two or more words, the user can browse through the resulting word list and choose the word he/she intended to write. The user, for example, enters the word *come*, by first pressing 2. The program will then propose the word *a* because it matches the entered sequence. When the user presses 6, 6, and 3, the program might propose the words *an*, *con* and finally *come*. The words *bone*, *bond*, and *anod* (and some more), also fit the given sequence. The user can access these words by pressing a next-key. [2]

f. LetterWise

LetterWise is a linguistically optimized technique that is not dictionary-based. When typing, press the key with the letter you want. Most probably, the letter you intend will appear. If it does not, press the NEXT key repeatedly until the right letter appears. In English, the letter‘t’ is often followed by ‘h’ and hardly ever by ‘g’. The program selects the most probable letter knowing the previous one. The main advantage of the method is the small amount of memory needed. Another advantage is the fact that it is just as easy to enter words, which are not in the dictionary. Therefore, this could be a suitable fallback method instead of the multi-tap methods, to produce faster text input. [2]

g. WordWise

Word Wise developed by **Eatoni Ergonomics** uses an auxiliary key. A character on a key is selected explicitly by simultaneously pressing the key corresponding to the character and the auxiliary key indicating the position of the character on the key. This decreases the number of matching words for a key sequence considerably because the user explicitly disambiguates some characters in the sequence. A drawback is that two keys must be pressed concurrently. With a limited space keyboard, this can prove difficult to some users. [2]

3.2 Suggested Input Methods

Since there are only 12 keypads in a standard mobile phone, the main problem with entering Khmer text is that 1 key is mapped to many different characters (9 characters).

Two-key press

In the two-key press method, the user presses two keys successively to specify a character. First, as in multi-press method, the user selects the group of the characters. The system will display the pop up multiple candidates along with its index position. Then, the user presses the index key of the intended character. The system will understand that it is the index of the key unless user presses the key immediately after selecting the group of characters (n seconds). If more than n seconds, the system will pop up other multiple candidates correspond to the key that user had pressed. In the case of Khmer, all the characters can be reached within two key presses, except the subscripts, which require three keystrokes.

For example, if user wants to write a word: "សាលា", he/she will press the sequences of keys as following:

1. User will first press key 6 (យ រ ល វ ស ហ ទ្ឍ អ ឱ)
2. Second he/she presses key 5 (index of "ស") to choose character "ស".
3. Third user presses key 8 and 1 successively for "ា".
4. Forth user presses key 6 and 5 (index of "ល") successively to use character "ល"
5. Finally, user do the same thing as number 3

Notes: If user presses 1 key and then another key for it position lately (after n seconds) the system will display all the characters correspond to that key.

ក = 1 +1

2 keystrokes

ក្ = ◌្ + ក = 0 + 1 + 1

3 keystrokes

ក្ក = ក + ◌្ + ក = 1 + 1 + 0 + 1 + 1

5 keystrokes

ប៉ី = ប + ◌៉ + ◌ី = 5 + 1 + 2 + 8 + 7 + 7

6 keystrokes

សាលា = ស + ◌ា + ល + ◌ា = 6 +5 + 8 + 1 + 6 + 3 + 8 +1

8 keystrokes

ស្ត្រី = ស + ◌្ + ត + ◌្ + រ +◌ី = 6 + 5 + 0 + 4 + 1 + 0 + 6 + 2 + 7 + 7

10 keystrokes

ខ្ញុំទៅសាលា = ខ្ញុំ + ទ + ◌ៅ + ស + ◌ា + ល + ◌ា = * + 4 + 3 + 9 + 4 + 6 +5 + 8 + 1 + 6 + 3 + 8 +1

13 key strokes

In short, we can infer that the number of keystroke of consonants (NBKC) equals the number of characters (NBC) multiply by two (**NBKC = NBC * 2**), the number of keystroke of vowels equals the number of vowels (NBV) multiply by two (**KBKV = NBV * 2**), and the number of keystroke of subscript (NBKS) equals the number of subscript (NBS) multiply by three (**NBJS = NBS * 3**).

4 Conclusion

We are strongly believed that this new keypad layout with new input method (two-key pressed method) will help users to write text easier and faster than before.

5 References

[1]: http://nds1.nokia.com/files/support/apac/phones/guides/Nokia_5000_APAC_UG_km.pdf
January 25, 2009

[2]: HMS: A Predictive Text Entry Method Using Bigrams: By Jon Hasselgren Erik Montnemery Pierre Nugues Markus Svensson January 25, 2009

[3]: PAN Localization: Input Method for Mobile Phone in Khmer Language By Ms. ROS Pich Hemy

[4]: http://www.mobileinfo.com/News_2003/Issue31/motorola_iTap.htm
January 26, 2009