2006－HI－117（5）
2006／1／13

ミャンマー語ショートメッセージサービス（SMS）のローマ字入力法の開発
－マルチタップ入力法との比較、シミュレーション、分析－

イェ　チョウ　トゥ　　浦野　義頼
早稲田大学大学院　国際情報通信研究科

ショートメッセージサービス（SMS）は、先進国における人々の生活の上で不可欠なものとなっている。発展途上国のひとつであるミャンマーでは、今年（2005年）、現地の人々を対象としたSMSが始まったばかりである。しかしながら同サービスは現在英語でのみ提供されており、ミャンマー語のSMSが必要となっている。また、ミャンマー語の効率的なSMS入力法はまだ存在しない。それゆえ、本論文は、携帯電話での実用的かつ効率的なミャンマー語のローマ字入力法の提案を試みるものである。

Development of Romanized Input Methods for Myanmar Language Short Messaging Service (SMS):
Comparison with Multitap Input Method, Simulation and Analysis

YE KYAW THU and YOSHIYORI URANO
Graduate School of Global Information and Telecommunication Studies, Waseda University

Short Messaging Services (SMS) has become an integral part of people's lives in developed countries. One of the Developing countries Myanmar (Burma) has just started SMS for local people this year, 2005. However, the service is provided only for English language at present, and SMS for Myanmar language is necessary. And there is no efficient input method for Myanmar language SMS yet. Therefore, this paper is an attempt to propose a practical and efficient romanized input method for Myanmar language on a mobile phone.

1. Introduction

Over the past few decades, wireless devices have revolutionized the communication industry, and text messaging service such as Short Messaging Service (SMS) is getting more and more popular in the global wireless market today. For developing countries like Myanmar, where the expense of lying physical fixed communication lines with the cost of maintenance is huge, the advent of the wireless communication revolution is a welcome development. But there is no standard input method not only for a mobile phone but also for a PC keyboard for Myanmar language at present. In this paper, we propose a possible "Romanized input method" for Myanmar Language SMS. Although this model is based on a romanization process for a language, we neglect the pronunciation of words according to our language nature. In this model, we are trying to pay attention not only to input efficiency but also to the shortness of novice user's learning curve. Here, we will make explaination together with our proposed "Multitap input method". We also developed a simulation programs for both models and made performance camparison between these two models. Here, we only used two formulae (KSPC and WPM) for performance analysis for some reasons. Then, we will mention the merit and demerit of both models according to experiments outcomes. Although we cannot say that this model is the most suitable for Myanmar language, we believe that it is also one of the possible choices.

2. Myanmar Language

Myanmar Script has been a majority language of Myanmar for over 1000 years old. Myanmar writing system derives from the Brahmi-related scripts which flourished in India from about 500 B.C to over 300 A.D. Myanmar script is a system of writing constructed from consonants, consonant combination symbols (i.e. Medials), vowel symbols related to the relevant consonants, and diacritic marks indicating tone level (niggahita, visajjaniya). Overall writing direction is from left to right. Despite great differences in appearance and detail, the Myanmar script follows the same basic principle as Devanagari. Today the Myanmar script looks circular, but it used to look flatter and more decorative many years ago.
Traditionally, Myanmar language alphabet is recognized as containing 33 consonants and 12 basic vowels, viz.

Consonants : က ခ ဂ ဃ င
စ ဆ ဇ ဈ ဉ
ဋ ဌ ဍ ဎ ဏ
တ ထ ဒ ဓ န
ပ ဖ ဗ ဘ မ
ယ ရ လ ဝ သ
ဟ ဠ အ (၃၂)

Independent Vowels: ဣ ဤ ဥ ဦ ဧ ဩ ဪ

Vowels : အ အာ အိ အီ အု အူ ဧ အဲ
ဩ ဪ အံ အိ့

Medials: ျ ြ ွ ှ

Finals : င် စ် တ် ယ် က် ဋ် ဖ် ဒ် လ် etc.

Abbreviations: ဏ္ ဿ ဠ္ ဍ္ဎ

Pali: ကုဋ္ဌ အာဠာဟနဂုဏ္ဍ အဓိဋ္ဌ

3. PC Keyboard Layouts for Myanmar Language

Although there are several PC keyboard layouts for Myanmar language, few of them use an actual keyboard mapping. Most layouts are just font and for display only (i.e. we cannot make saving, sorting or searching in Myanmar language). Unicode table for Myanmar is now updated to version 4.1 and still fixing some problems. Some local ICT companies have been trying to develop Unicode based keyboard mapping for Myanmar language, however, there is no efficient or standard input method yet. Users have to practice for so many hours to type Myanmar language. Therefore, when we try to search efficient input methods for mobile phone keypad, we face many difficulties. There are no digital resources for Myanmar language (e.g Web contents, Newspapers, Magazines, Books etc.) to make analysis such as on usage frequency of Myanmar letters. Today, most web contents of Myanmar text are picture format like JPEG (i.e. scanned magazines, books etc. saved as picture format and uploaded on the internet). So, we have to create Myanmar language contents by ourselves with our own input method. According to Myanmar language nature, spaces are used to separate phrases, not words. Pali, the language for Buddhism religious, is also a part of Myanmar language. Some of the Pali words are used in Myanmar language even in daily conversion. And thus, we need to think about conveniency for inputting Pali words also.

4. Myanmar Language Unicode Tool

For making Myanmar language analysis, such as frequency ranking of Myanmar letters, we have developed "Myanmar Language Unicode Tool ver.1.0". By using this tool, we can create and save Myanmar Unicode text file easily as well as search feature etc. There are two input interfaces in this tool, the first interface is like visual keyboard (Fig. 1) and the second one is our own Romanization input interface (Fig. 2).
Here, we use "Myazedi unicode font" for displaying Myanmar letters but inputting and combination process are done by our own ways.
The Program is developed with Microsoft Visual Basic .Net 2003 for the purpose of creating some Myanmar text files and making analysis on frequency usage of Myanmar letters (in this research we only considered from က to အ) in present Myanmar literature.
In the user interface of visual keyboard, the user can input Myanmar letters by clicking the buttons. When user is typing, Myanmar letters will be shown in the right pane and the equivalent unicode value in the left pane. The program allows saving Myanmar letters unicode values as a text file format but not actual Myanmar texts.

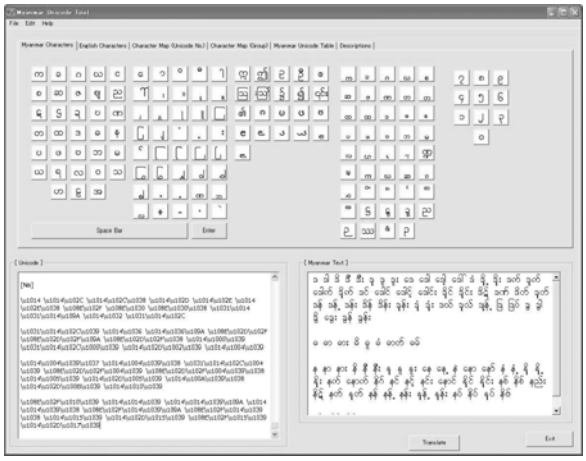

Fig.1. Myanmar Language Unicode Tool Version 1.0

The program can identify contained Myanmar letters, the number of letters and its percentage compared to other letters in a given file. The program can also draw a usage frequency graph of each letter. We believe that this process is very important in designing keypad layout for some input methods such as Fast-tap [22] and SIMKEYS [4].

5. Romanization for Myanmar Language

Currently, there is no standard or officially approved romanization system for Myanmar language yet. There have been attempts to make one, but none have been successful. In Myanmar language, many words are spelled differently from the way they are pronounced. For example, the word for "thief" is pronounced *tha-kou* (သိုး) but spelled *thu-kou* (သူိုး), the word for "ox" is pronounced *na-thou* (နှိုး) but spelled *nwa-thou* (နွားသိုး). Replicating Myanmar sounds in the Roman script is difficult. There is, however, a Pali-based romanization system, but it fails to replicate the consonants in contemporary Myanmar language. In the next section, we will mention briefly about some historically romanization attempts for Myanmar language and our new romanization input approch for PC.

5.1 Transliteration

Myanmar language script is also used for Pali (here, we mean Myanmar's Pali). There is widely accepted romanization system for Pali and this can be applied with fair success to transliterating Myanmar script. So there are not many variations in the different systems using the transliteration method. Such variations as do occur affect the few extra symbols which were added to the alphabet (one or two of them from Sanskrit) to cover features peculiar to Myanmar script. The first transliteration for Myanmar language seems to have been proposed by H.L. St Barbe in 1878. His transliteration method is the same as Duroiselle's transliteration method for Pali which is adopted by the "Pali Text Society for Pali". Duroiselle's system was sutible for old Myanmar script and then virtually unchanged by Luce (1932), Than Than (1959), and Ba Shin (1962). Small modified version may be seen in Yi Yi (1960), Sprigg (1963), Bernot (1965), Okell (1965), and Whitbread (1969) [5].
The following is an example of romanization by Duroiselle's system:

Table 1. Duroiselle's romanization system

Myanmar Words	Duroiselle's Transliteration
ည\၉	ñ
ဝတ္ထု	vatthu
ဒွာရ	dvāra
သုတေသန	sutesana

5.2 Transcription

Although transliteration for Myanmar language can be refered to the rule of Pali romanization, there is no guide-line for transcription. There are several varieties of transcription methods. Some of the sounds are rendering by the same roman letter (e.g. k, t, p, n, m, l, h) but there is several variation in the conventions for representing certain vowels, the finals, tones, and some of the initial consonants, especially the aspirates. The following is the example for rendering "final nasal "and "hush sibilant" in different transcription systems. [5]

Final nasal: n ng ŋ ṅ ñ or +˜
Hush Sibilant: sh hs hy š ʃ or ɕ

5.3 Combined System

In romanization process, some of the scholars have believed that it is impossible to show acccurately the pronunciation of a word as well as its spelling in Myanmar script. According to the Myanmar proverb, "spelling is correct but pronunciation is voiced" is one of the problems. However, Minn Latt's Combined System (the Prague method romanization of Burmese) is famous. He made some minor revisions in his romanization such as "th sh š θ" to "ht hs sh th". In Minn Latt's romanization: "Myanmar" is written as "myánmá". [5] The following table is a

comparison between existing romanization methods for Myanmar word သုတေသန (research).

Table 2. Comparison between romanization methods

Romanized Word	Romanization Method
sutesana	Duroiselle's System
thŏŏ́téthană	Latter's System
thu.te-thăna.	Grant Brown's System (Conventional)
θúteθəná	Stewart's System (IPA)
thu.tei-thana.	Cornyn's System (typewritten)
thutéithăna	Minn Latt's 1966 System

5.4 New Romanized Input Method for PC

After studying several existing romanization systems, we noticed that the difficulty is how to represent the same pronunciation words with different spelling like (ကျ and ကြ), (ဉ , ည and ဥ), (အို, ဩ, ဪ and ဥ, ဦ, ဦး) , (ယောက်and ရောက်) and (ဒီ and ဍီ) etc. If we input according to pronunciation, the spelling of the word might be mistaken. Here, we only concentrate on how to input Myanmar words correctly to PC. And if we make roman character assignment to all possible Myanmar words, it will be difficult for users to memorize and the system has to be intelligent. And thus, we have neglected the pronunciation of words and assigned some English alphabets on Myanmar letters according to their original pronunciation. The following is an example of assignment to Myanmar letters.

Table 3. New method of assignment

Myanmar letters	Romanized
က ခ ဂ ဃ င	ka kha ga gha nga
စ ဆ ဇ ဈ ည ဉ	ca cha ja jha nnya nya
◌ျ ◌ြ ◌ွ ◌ှ ◌ၠ ◌ၡ	yc lgt lgts tn nn
ဋ္ဌ ၍ ၎င်း ၏	Type "sym" then make selection
အ ဣ ဥ ဦ ဧ	i ii u uu e
ဩ ဪ ဿ ၐ ဣ်	First type main consonent, then type "ps" and make selection

This input method idea is based on Myanmar language writing order. For example, to type "ကြာဖြူ" (white lotus): the equivalent roman assignment is "yykaycyyphannl". Here, "yy" for "◌ြ", "ka" for "က", "yc" for "ာ" and "yy" for "◌ျ", "pha" for "ဖ" and "nnl" for "ူ". We assume that this input method will be very easy for native Myanmar people who are already familiar with English keyboard.

We developed a simulation program to experiment this input idea (Fig. 2). Although this input method cannot get (Key Stroke per Character) KSPC = 1, an advantage is that even first experienced user can become familier with the typing method within 5 to 10 minutes. The learning curve is very short in this input method.

In this program, we use "selection method" for several "yayit" (e.g ◌ြ, ◌ျ, ◌ွ, ◌ှ, ◌ၠ, ◌ၡ , ◌ၢ and ◌ၣ) and this constraint will be settled if the system is intelligent enough to be able to detect the following consonant and check possible conbination of letters. Actually, this idea is not fully romanization but an attempt to type Myanmar characters using assigned English alphabets. The following is an example of how to type Pali word "ဝတ္ထု"in this model.

1) type "wa" for "ဝ"
2) type "ta" for "တ"
3) type "ps" and then select "ဿ" from the selection menu
4) type "tnl" for "ု"

Fig.2. Romanized input simulation on PC

6. Comparison with Multitap Input

Here, we will describe different input ideas between "Multitap Input Model" and "Romanized Input Model".

6.1 Multitap Input Model

Fig.3. Multitap keypad layout

The standard keypad used in mobile phones consists of 12 keys: number 0 to 9, * key and # key. In English mobile phone keypads, the letters a-z are spread across the keys 2 to 9 with 0 or # used for SPACE. The first and still most common way to enter text on a mobile phone is the Multitap or Multi-Press method. This method works by cycling through letters on a key with each successive press e.g pressing '2' key once for 'a', twice for 'b' and three time for 'c'. For example, if you want to type "the", the key press sequence will be "84433".
There are two kinds of multitap input method.

1) Multitap with timeout value
2) Multitap with timeout kill button

The first method is based on a time limit during which the second one of the subsequent key presses must happen. During the timeout value, the device will intepret the user's input as selecting one of the letters represented by that key using multiple key presses. When the timeout occurs, the next key press is interpreted as a new letter.
The second method uses timeout kill button (e.g arrow key) instead of timeout value. In this system, a user has to kill the timeout by pressing a timeout kill button. In this research, our model is designed for multitap with timeout kill button method. The simulation program and evaluation is also based on it.
Based on the standard English Multitap keypad layout, we have assigned Myanmar letters for each key. Here, 1-key is assigned for Myanmar letters က, ခ, ဂ, ဃ, င. 2-key is assigned for Myanmar letters စ, ဆ, ဇ, ဈ , ည and so on. Other symbols such as "killer stroke" (်), vowel symbol (ိ), tone-marks (့), (း) etc. are assigned on 8-key and user can select one or more symbols at the same time. "Conjunct consonants" are those which are either written one under the other, e.g. ္က, ်ံ, ္ဆ etc. or in a special form viz. ကြ, ကျ, ဿ. Those letters are also assigned on remains keys on the keypad.

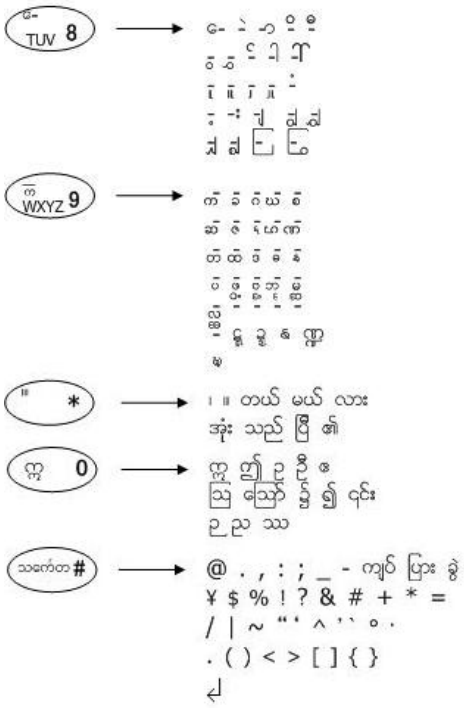

Fig.4. Myanmar letters assignment for selection

Here, multitap input method for Myanmar language that we have considered is not simple as English multitap. This is because Myanmar language has more letters than English language and its writing system uses several combinations between letters, medials and killers etc. The problem is that there are a lot of possible combinations to type one word, but we basically have only 12 keys on the keypad. Our idea is that users first select consonant or main letter to type (i.e by using multitap input method), and then choose the already defined combination (i.e by selection method). The most frequently used Myanmar words like "ပြီ", "၏" etc. are assigned to the "*" key, and "၌", "၍", "ဿ" etc. are assigned to "0" key respectively. Accordingly, users can type these letters easily by the way of "press a key and select". Although this method totally neglects the order of our traditional writing system, this way can be efficient for typing Myanmar language on the mobile phone keypad.

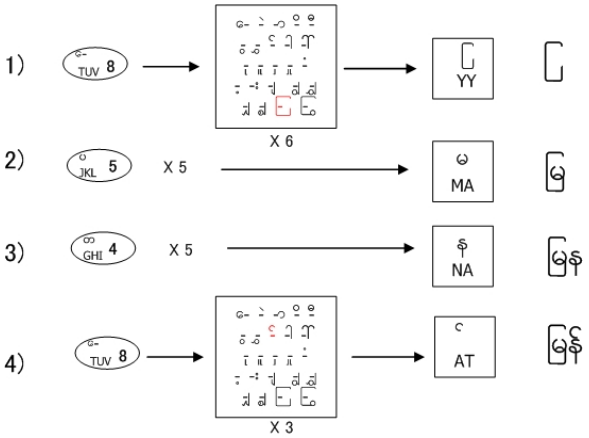

Fig.5. Key sequence required to type မြန်

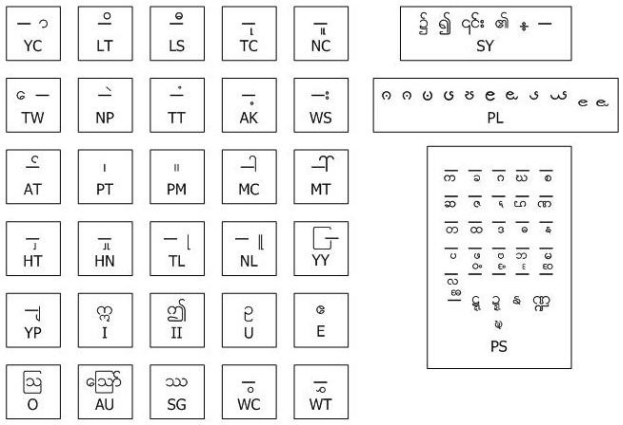

Fig.7. Assignment for other Myanmar letters

6.2 Romanized Input Model

In this model, we consider possible romanization input methods for Myanmar language SMS. Like in PC romanized input ways, we have assigned 1 or 2 English alphabets for each Myanmar letter (e.g. KA for က, KH for ခ, GA for ဂ and O for သ etc.). Comparing with PC model, we have assigned 2 English alphabets maximum for mobile model, so that faster input can be made even though there are limited keys on mobile phone keypad. We call this model "2C Romanized Input Model". By using it, we can type Myanmar letters without changing (i.e encoding of letters on a mobile phone keypad) to existing English keypad layout. Although this idea is very convenient for PC keyboard, it is not so convenient for twelve-key mobile phone keypad yet.

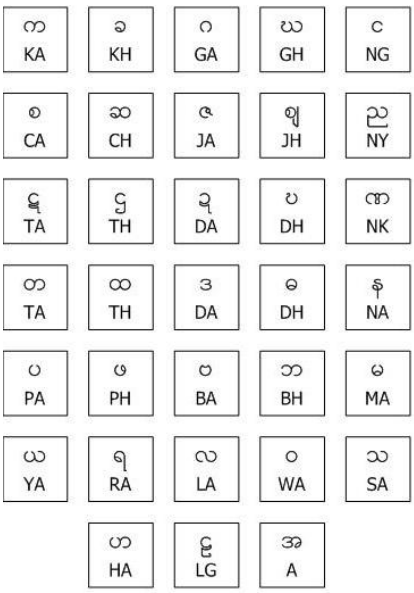

Fig.6. "2C Romanization" for letters "ka" to "a"

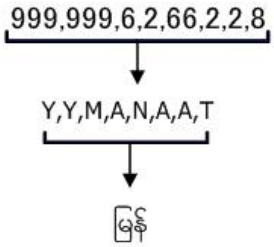

Fig.8. Key sequence required to type မြန်

7. Experiments and Evaluations

We have developed simulation programs for each input model coding with Java 2 for Micro Edition (J2ME) Wireless Toolkit Ver.2.2 [22] and Microsoft Visual Basic .Net. After some experiments on the two input models, we have been making evaluations to find efficient input methods. This evaluation is based on the following two facts.

7.1 Input speed

For evaluating input speed, we use two fomulae; the first one is calculating KSPC for each model. KSPC is an acronym for *keystrokes per character*. It is the number of keystrokes required, on average, to generate a character of text for a given text entry technique in a given language. Here, to type lowercase letters in QWERTY keyboard, we assume that KSPC = 1.

$$KSPC = \sum_{i=1}^{33} Pi \times ni$$

The second formula is used to calculate WPM (Word per Minute).

$$WPM = \frac{60}{CPW \times TPC}$$

Here, CPW means "Characters per Word" (average number of characters that constitute each word in the Myanmar language). TPC is "Time per Character" (average time required to type a character).

7.2 Usability (or) Learning curve

Learning curve (or) the amount of time that users have to practice to become familier with the new input method is also important point of view.
For the proposed two input methods, learning speed of the "Multitap model" is faster than "Romanized input model". Even if the users are already familiar with SMS in English, there is still a difficulty in memorizing assigned English alphabets for Myanmar letters in "Romanized input model". It is a little bit more difficult to memorize 2C romanized input methods compared to PC romanized one.
For the experiments such as Myanmar character combination test, we use some emulator programs. As a result of several experiments, although we can see Myanmar characters on the emulator screen, it cannot be shown on the actual mobile phone device. And thus, we need to create our own special fonts (e.g bitmap font) to show Myanmar characters correctly on real device. For evaluating input speed, we have developed small simulation programs by using Visual Basic and Java programming.

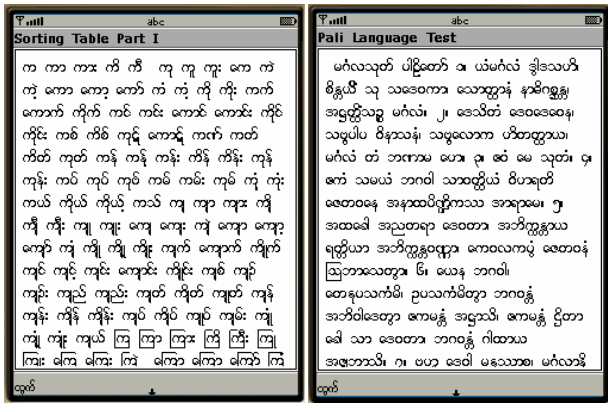

Fig.9. Experiment with J2ME emulator

The following two screens are output of Myanmar character statistic data from the Myanmar Language Unicode Tool Program Ver.1

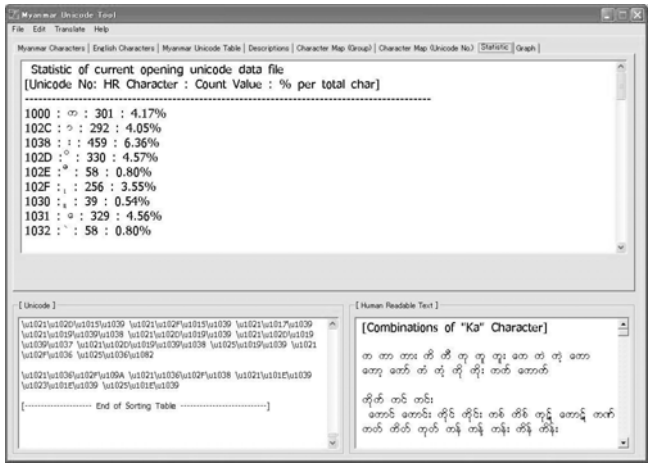

Fig.10. Statistic results for counting Myanmar letters

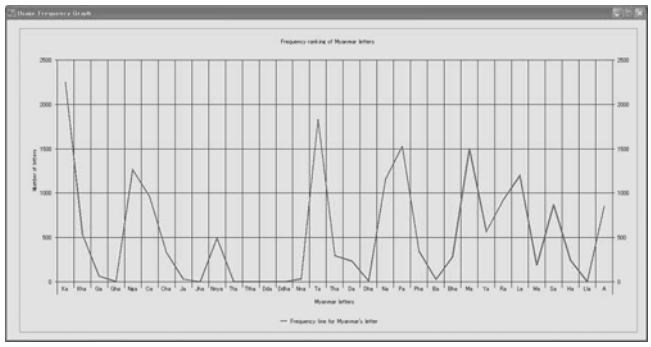

Fig.11. Graph for frequency of usage on Myanmar letters for "Pa Loot Toot cartoon journal"

The following table is a sample keystrokes comparison between romanized and multitap input method. This process is still under research.

Table 4. Keystrokes comparison between Romanized and Multitap Input Method for Myanmar language

Myanmar Words	Romanized	Multitap
ငါး	14	16
စာပေ	15	9
မိဘ	10	14
လှလှ	14	18
ဆရာ	15	8

8. Conclusion

In this paper, we focus mainly on the Multitap and Romanized input methods for Myanmar language. However, prioritizing current usage of Myanmar letters is still under research and evaluation for each model with typing error rate is not completed yet because of some difficulties. When the "Table of frequently used Myanmar letters " come out, assignment of Myanmar letters on the keypad might change. Currently, all of the experiments are based on PC simulation software and mobile phone emulator programs (such as Nokia's emulator). We plan to make expriments on actual mobile devices in the near future in order to get native users' feedback. We have a plan to study other famous input methods such as Less-Tap, MessagEase, SIMKEYS and T9 input models, and then continue searching efficient input methods for Myanmar language on mobile phone.

9. Acknowledgements

This research is supported by Prof. URANO Lab., Graduate School of Global Information and Telecommunication Studies (GITS), Waseda University, Tokyo, Japan. We are grateful to Mr. Ngwe Tun of Solveware Solution for allowing us to use "Myazedi Unicode Font" freely in our thesis.

References

[1] Scott MacKenzie, "KSPC as a Characteristic of Text Entry Techniques", 4th International Symposium on Human Computer Interaction with Mobile Devices, 2002

[2] Robert William Soukoreff, "Text Entry for Mobile Systems: Models, Measures, And Analysis for Text Entry Research", A Thesis Presented to The Faculty of Graduate Studies of York University, June 2002

[3] Shiferraw Abebe and Tewodros Seyum**, "**Ethiopic Keyboard Mapping and Predictive Text Inputting Algorithm in a Wireless Environment", ETC Conference, September 20, 2004

[4] Rick W. Ha, Pin-Han Ho, and Xuemin (sherman) Shen, University of Waterloo, "SIMKEYS: An Efficient Keypad Configuration for Mobile Communications", IEEE Communications Magazine, November 2004

[5] John Okell, *A Guide To The Romanization Of Burmese*, The Royal Asiatic Society of Great Britain and Ireland, 1971

[6] John Okell, "Alphabetical Order In Burmese", JBRS, LI, ii, Dec., 1968

[7] John Okell, and Anna Allott, *Burmese /Myanmar Dictionary of Grammatical Forms*, Curzon press, Richmond, Surrey UK, First Published in 2001

[8] John Okell, *A reference grammar of Colloquial Burmese* I *and* II, Oxford University Press, Ely House, London, 1969

[9] Martin Hosken, SIL Non-Roman Script Initiative, "Myanmar Script Canonical Ordering", from SIL Web Site, November 26, 2002 Rev. 42

[10] J.E. Bridges, *The Burmese Manual*, British Burma Press, Rangoon, 1906

[11] Myanmar Language Commission, *Myanmar-English Dictionary*, Myanmar Language Commission Press, No.27, Pyi Road, Yangon, 1993

[12] Sun Associates, *Practical Myanmar*, Thein Myint Win Offset, 96/11 Street, Yangon, 1996

[13] David Bradley, Burmese Phrase Book, Lonely Planet Publications, PO Box 617, Hawthorn, Vic 3122, Australia, 1997

[14] Michael Morrison, *Beginning Mobile Phone Game Programming*, SAMS, 800 East 96th St., Indianapolis, Indiana, 46240 USA, 2005

[15] Dreamtech Software Team, *Wireless Programming with J2ME*, Hungry Minds, Inc., 909 Third Avenue New York, NY 10022, 2002

[16] Ralph Barbagallo, *Wireless Game Development in C/C++ with BREW*, Wordware Publishing Inc., 2320 Los Rios Boulevard Plano, Texas 75074, USA, 2003

[17] Ye Kyaw Thu, "Developing Efficient Input Methods for Myanmar Language SMS", IEICE Communication Society Conference, Hokkaido University, Sapporo, Japan, Sept 2005

[18] Myanmar Unicode Table Version 4.1, http://www.unicode.org/charts/PDF/U1000.pdf

[19] J2ME Documentation, http://java.sun.com/j2me/docs/index.html

[20] Microsoft Developer Network http://msdn.microsoft.com/default.aspx

[21] Myanmar Unicode Reference Papers http://www.myanmars.net/unicode/doc/index.htm

[22] Fastap Keypads http://www.digitwireless.com/