Evaluation of Text Entry Methods for Korean Mobile Phones, a User Study

Ivaylo Ilinkin
Gettysburg College
Gettysburg, PA 17325, USA
iilinkin@gettysburg.edu

Sunghee Kim
Gettysburg College
Gettysburg, PA 17325, USA
skim@gettysburg.edu

ABSTRACT
This paper reports the results of a user study designed to evaluate text entry methods for mobile phones used in Korea. At present the keypad layout for Korean mobile phones has not been standardized and different manufacturers produce phones with different layouts. Included in the evaluation are three of the dominant text entry methods: Chon-ji-in, EZ-Hangul, and SKY. The metrics used in the analysis are key strokes per character, words per minute, and total error rate. The results suggest that SKY offers a good balance between speed, effort, and accuracy. The paper also introduces a phrase set that has high correlation with the Korean language and could be used in other experiments on Korean text entry methods.

Author Keywords
Hangul, Korean text entry, mobile phones, evaluation.

ACM Classification Keywords
H5.2. Information interfaces and presentation: User Interfaces – *Evaluation/methodology*.

General Terms
Performance.

INTRODUCTION
This paper presents a formal user study that compares the three dominant Korean text entry methods for mobile phones and introduces a phrase set for Korean that could be used in other experiments. The phrase set is derived from the work of MacKenzie and Soukoreff [4] and has high correlation with the Korean language.

This is the first comprehensive user study, to the best of our knowledge, for evaluating Korean text entry methods. The closest prior work [1] involved 15 participants and used the same two phrases (of length 7 and 32) for both training and testing. The training was considered complete when the participant could type each phrase in three consecutive trials at roughly the same speed (± 5 seconds). During testing each participant entered the phrases three times. To correct errors the participants were instructed to retype the incorrect character. The participants spent between 45 and 80 minutes for the whole experiment. Different phones were used for each method, although an attempt was made to ensure that they were of similar dimensions.

In contrast, the study reported in this paper used the same device across all methods to eliminate any possible device-dependent effects. In addition, each participant entered a total of 400 phrases over 8 sessions conducted on 4 consecutive days. Deletion was allowed for error correction and error rates were measured as described in [7].

KOREAN WRITING SYSTEM (HANGUL)
The Korean alphabet has 24 basic and 16 compound letters (Figure 1). A distinguishing characteristic of Hangul is that the letters within a syllable are stacked according to predefined rules (Figure 2). Thus, both the individual letters and syllables can be treated as independent structural units.

Consonants	basic	ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅇ ㅈ ㅊ ㅋ ㅌ ㅍ ㅎ
	compound	ㄲ ㄸ ㅃ ㅆ ㅉ
Vowels	basic	ㅏ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅠ ㅡ ㅣ
	compound	ㅐ ㅒ ㅔ ㅖ ㅘ ㅚ ㅙ ㅝ ㅟ ㅞ ㅢ

Figure 1. The Korean alphabet. The shaded letters represent two-vowel diphthongs. For the analysis in the subsequent sections the non-shaded letters are considered *characters.*

Word	한글					
Syllables	한			글		
Letters	ㅎ	ㅏ	ㄴ	ㄱ	ㅡ	ㄹ

Figure 2. Example of Korean word and syllable structure.

KOREAN TEXT ENTRY METHODS
The predominant Korean text entry methods are Chon-ji-in, EZ-Hangul, and SKY found on phones by Samsung, LG, and SK Telecom, respectively. Unlike the text entry method for English, which lays out the letters alphabetically along the keypad and within the individual keys, Korean methods often place similar-sounding letters on the same key (e.g. ㄱ ㅋ ㄲ which represent variations of the sound [**g/k**]). The shapes of the letters make such grouping all the more

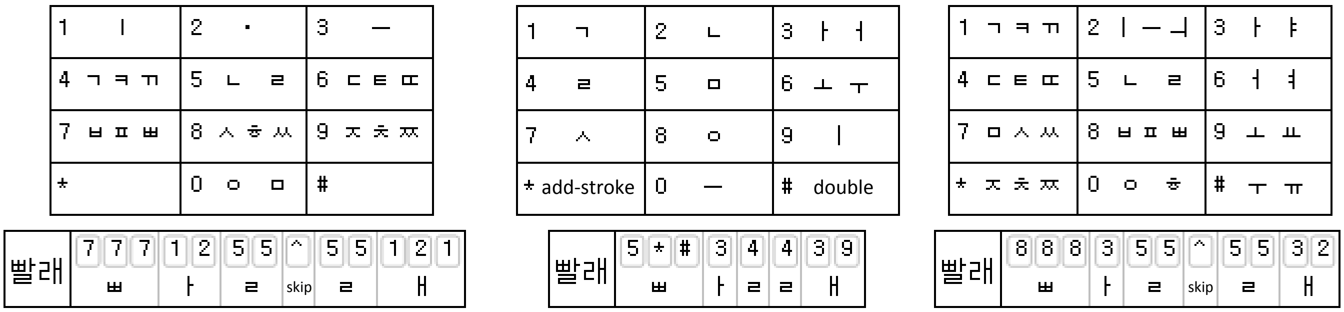

Figure 3. Keypad layouts and example key stroke sequences for methods A (left), B (middle), and C (right).

natural. Alphabetical order is typically maintained across the keypad with respect to the first letter on the keys.

Chon-ji-in (Method A)

This method (Figure 3, left) incorporates cultural elements to aid the user in adopting the technology. The symbols corresponding to the keys 1, 2, and 3 have the meaning of *sky* (|), *ground* (—), and *man* (•), and all vowels are constructed using these keys (*concepts*). The consonants are assigned to the rest of the keys and are produced by multi-tap. The compound consonants are normally not shown since their placement is implicit by their single counterparts.

This layout is considered easy to learn, but in general it requires more key strokes for text entry. It suffers from segmentation of the consonants, i.e., typing consecutively two letters assigned to the same key.

EZ-Hangul (Method B)

This method (Figure 3, middle) uses the concepts of *adding a stroke* and *doubling* to compose the letters based on their shapes. Only six consonants are assigned a number key and are used with the keys * (add a stroke) and # (double) to compose the rest. The vowels are composed similarly.

In general this method requires fewer key strokes. However, since only a subset of the letters is visible on the keypad and the composition rules are not readily apparent, it is considered somewhat difficult to learn. A distinctive feature of this layout is that it does not suffer from segmentation.

SKY (Method C)

This method (Figure 3, right) places the consonants on the left two columns of the keypad and the vowels on the right. Unlike the previous two methods all the basic consonants and vowels are visible on the keypad. Similar to ***A*** this method suffers from segmentation.

This method requires *at most* 3 strokes per vowel – 5 require 1 stroke, 10 require 2 strokes, and 6 require 3 strokes. In contrast, with ***A***, 17 vowels require *at least* 3 strokes; with ***B***, 9 vowels require *at least* 3 strokes.

ANALYTICAL COMPARISON

Two metrics were used to compare analytically the text entry methods: KSPC [7] and KLM-GOMS [5, 6]. The input text was compiled from Korean classics (http://gojun.knu.ac.kr) and consists of 90314 words, 251566 syllables, and 624047 characters.

Key Strokes Per Character (KSPC)

$$\text{KSPC} = \frac{\text{Total Number of Keystrokes}}{\text{Total Number of Characters of Transcribed Text}}$$

Method ***A*** has the highest KSPC (1.51) since it requires more strokes on average for composing the vowels. Method ***C*** has the lowest KSPC (1.30) and ***B*** has KSPC of 1.32.

KLM-GOMS Predicted Time Per Character

The predicted movement time based on KLM-GOMS model [5, 6] is comparable for ***B*** (1003.92 ms) and C (995.82 ms), and they both outperform ***A*** (1111.39 ms). KLM-GOMS penalizes each grapheme (letter) that requires 2 or more keys. Nine letters in ***A*** are not penalized (and only 2 vowels are penalty-free) while 12 letters are not penalized in ***C***. In ***B*** 10 letters are penalty-free and the movement time component is also large because of the need to use the control keys for doubling and adding a stroke.

EXPERIMENT

Participants

Twenty four paid participants (12 male and 12 female) took part in the study. They ranged in age between 18 and 38 with an average age of 24.7 and median age of 24. All participants were right-handed and were native Koreans. The experiment was conducted in Seoul, Korea.

Design

The experiment was a mixed design with one *between-subject* factor, *method* (***A***, ***B***, and ***C***) and one *within-subject* factor, *session* (1 – 8). The participants were divided into three groups of eight (4 male, 4 female) and assigned a method that they had not used in the last five years based on a pre-study questionnaire about their mobile phone history.

Apparatus

The device was a *Nokia 6120 Classic*. The keys are *12mm* x *5mm* with a gap of *1mm* between keys. *Backspace* was mapped to *left-arrow*, *Skip* to *OK*, and *Space* to *right-arrow*.

Procedure

Each participant attended 8 sessions (50 trials per session) on 4 consecutive days (2 sessions a day separated by a 5-minute break). During each trial a phrase was displayed on the screen and the participants typed underneath. The phrase remained on the screen until the participant confirmed the end of the trial via the *right soft* key. The only means of correcting errors was via the *Backspace* key.

During training the participants were asked to type 15 phrases exactly as shown with a maximum of 5 attempts per phrase. During testing they were instructed to type each phrase as fast and accurately as possible, but did not receive feedback about their speed and accuracy. A 2-minute break was enforced after every 15 minutes within each session.

Phrase Set

The phrase set was derived from the one in [4]. The phrases were translated by the second author (a native Korean) and were altered where appropriate to reflect the idiomatic use of the language. The correlation with Korean (Table 1) was analyzed with the software in [4] using the letter frequencies reported in [2].

phrases: 419 minimum length: 8 maximum length: 62 average length: 29 syllables: 4467 letters: 11256	words: 1668 unique words: 1254 minimum length: 2 maximum length: 17 average length: 6
correlation with Korean: 0.9917	

Table 1. Phrase set statistics.

Results and Analysis

In total the participants entered 400 phrases. Of these, 27 were discarded since they corresponded to incomplete trials killed via accidental press of the *End* key. Analysis of variance (ANOVA) was used to analyze the results with Tukey's Honestly Significant Difference (HSD) test for post-hoc pairwise comparison.

Key Strokes Per Character (KSPC)

KSPC averages (Figure 4) were 1.66 (***A***, σ=.29), 1.51 (***B***, σ=.29), and 1.46 (***C***, σ=.27). There was significant effect for $method$, $session$, and $method * session$ (Table 2). Tukey's HSD reported ***C*** < ***B*** < ***A*** ($p < .01$), i.e. the KSPC for ***A*** was significantly higher than the other methods which is consistent with the analytical results.

source	df	*F*	sig.
method	2	404.933	$p < .0001$
session	7	4.094	$p < .0001$
method*session	14	2.511	$p < .01$
Error	8928		

Table 2. ANOVA results for KSPC.

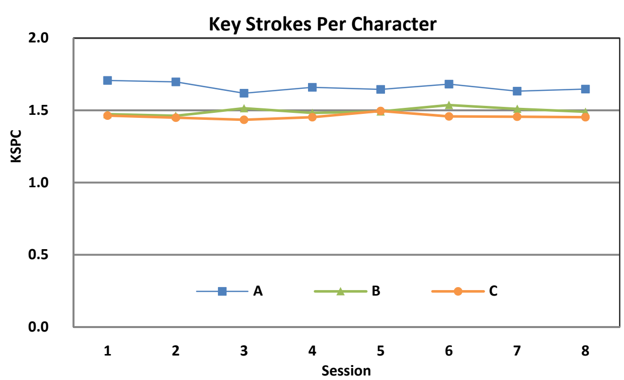

Figure 4. Average KSPC per method per session.

Words Per Minute (WPM) – Learning Curve

For this metric *word* is defined as five characters of transcribed text [3]. The average text entry speeds were 19.13 wpm (***A***, σ=4.98), 19.71 wpm (***B***, σ=5.17), and 19.22 wpm (***C***, σ=5.29). This difference was significant (Table 3). Tukey's HSD showed ***B*** > ***C***, ***A*** ($p < .01$), i.e., ***B*** was significantly faster, while there was no significant difference between ***C*** and ***A***. Participants improved over time which is supported by a significant effect of $session$ and $method * session$.

source	df	*F*	sig.
method	2	15.379	$p < .0001$
session	7	432.778	$p < .0001$
method*session	14	5.608	$p < .0001$
Error	8928		

Table 3. ANOVA results for WPM.

Figure 5 shows fitted learning curves for each method extrapolated to 24 sessions. The crossover points indicate that ***C*** affords ease of learning, since by session 7 it achieves higher speeds than both ***A*** and ***B***, and the extrapolated curves suggest that its advantage will continue. This is consistent with the fact that among the 16 most frequently used letters (which account for 86.1% of total use [2]) 13 require 1 stroke and 3 require 2 strokes with ***C***.

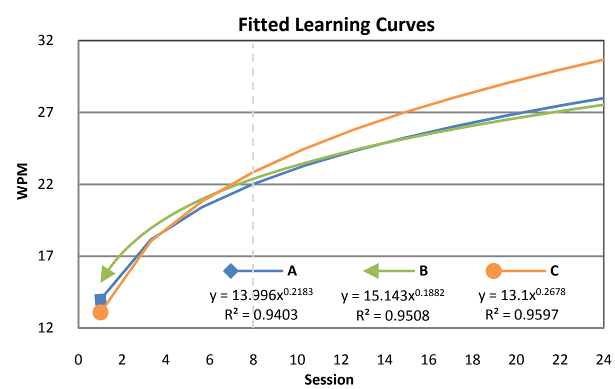

Figure 5. Fitted learning curves extrapolated to 24 sessions.

Surprisingly ***A*** maintains a speed advantage over ***C*** for the first 4 sessions even though among the top 16 letters 7 require 2 strokes and the rules for composing the vowels place higher mental load. Perhaps this can be explained by the fact that half of the participants assigned to ***A*** had used it before (but not within the last 5 years). In general, it is difficult to find participants who have not used ***A***, since it is one of the first methods for Korean text entry, introduced in 1999. Methods ***B*** and ***C*** were introduced in 2005 and none of the participants assigned to ***B*** and ***C*** had used them before.

Method ***B*** had the best initial performance, despite the fact that the rules for composing the letters are complicated, not all letters are visible on the keypad, and 6 of the top 16 letters require 2 strokes. Figure 5 suggests that ***B*** is likely to lose its speed advantage over ***A***, which is in contrast with the common perception in Korea that ***B*** is more difficult to learn but eventually affords faster text entry speed.

Error Rate

Three error metrics were computed as described in [7]:

- *uncorrected error rate*: a measure of the errors that remain in the transcribed text
- *corrected error rate*: a measure of the corrected errors
- *total error rate*: uncorrected plus corrected error rate

Figure 6 shows that the participants tended to correct errors as the uncorrected error rate was very low. The average total error rate was 4.83% (***A***, σ=7.15), 5.74% (***B***, σ=7.52), and 5.92% (***C***, σ=7.52). This difference was significant as shown in Table 4. Tukey's HSD showed ***A*** < ***B***, ***C*** ($p < .01$), i.e., there was no significant difference between ***B*** and ***C***, but there was significant difference between ***A*** vs. ***B*** and ***C***.

source	df	*F*	sig.
method	2	18.270	$p < .0001$
session	7	3.850	$p < .0001$
method*session	14	1.620	$p > .05$
Error	8928		

Table 4. ANOVA results for Total Error Rate.

Figure 6. Average error rates per layout per session.

Participant Conscientiousness (PC)

PC was proposed as "a means to distinguish perfectionists from apathetic participants" [7]:

$$\mathrm{PC} = \frac{\text{Total number of corrected errors}}{\text{Total number of errors}}$$

The participants were very conscientious (average PC=0.95, σ=0.03). Despite the instruction to balance speed and accuracy, there appeared to be a tendency to correct errors.

CONCLUSION

This paper presented the results of a formal user study for evaluating the three most common text entry methods for Korean mobile phones: Chon-ji-in (***A***), EZ-Hangul (***B***), and SKY (***C***). The phrase set used in this study, derived from [4], is shown to have high correlation with Korean and could form the basis of a corpus for further studies.

The text entry methods were evaluated based on KSPC, WPM, and total error rate. ANOVA found *method* to be a significant main effect for all metrics. In terms of KSPC, statistically significant difference was observed between all methods with lowest for ***C*** and highest for ***A***. Text entry was fastest with ***B*** while there was no significant difference between ***C*** and ***A***. However, ***C*** achieved faster text entry speed by the end of the experiment and learning curve analysis suggests that this advantage is likely to continue. The total error rate for ***A*** was significantly lower than that of ***B*** and ***C***. Overall, ***C*** seems to offer a good balance between speed (WPM), effort (KSPC), and accuracy (TER).

A distinguishing characteristic of all three methods is that the consonants and vowels are arranged in two separate sections. In ***A*** the keys for the vowels occupy only the first row, while in ***B*** and ***C*** they are assigned to the right-most column. This grouping appears to facilitate two-thumb typing of *consonant-vowel* sequences that are inherent in the structure of syllables. In fact, the computer keyboards for Korean text entry also exhibit a similar grouping – left half for consonants and right half for vowels.

	2 AIU	3 EOY
4 BCD	5 FGHJ	6 KLM
7 NPQR	8 ST	9 VWXZ

Figure 7. Possible layout for English.

The observations from this study could inform the design of text entry methods for other languages. As a proof of concept, the layout in Figure 7 employs the vowel-consonant split, maintains alphabetical order, and does not hinder recognition, but has considerably fewer conflicts (722 vs. 1229) and lower KSPC (1.82 vs. 2.09) than the standard layout on the phrases in [4].

REFERENCES

1. Kee, D. Evaluation for performance and preference of Hangul entry methods using real mobile phones. *Journal of the Ergonomics Society of Korea*, *25*, 3 (2006), 33-41.
2. Kim, H. and Kang, B. *Frequency analysis of Korean*. Korea University Institute of Korean Culture, 1997
3. MacKenzie, I.S. and Soukoreff, R.W. Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction*, *17* (2002), 147-198.
4. MacKenzie, I.S. and Soukoreff, R.W. Phrase sets for evaluating text entry techniques. *Ext. Abstracts CHI 2003*, ACM Press (2003), 754-755.
5. Myung, R. Keystroke-level analysis of Korean text entry methods on mobile phones. *International Journal of Human-Computer Studies*, *60*, 5-6 (2004), 545-563.
6. Silfverberg, M., MacKenzie, I.S., and Korhonen, P. Predicting text entry speed on mobile phones. *Proc. CHI 2000*, ACM Press (2000), 9-16.
7. Soukoreff, R.W. and MacKenzie, I.S. Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. *Proc. CHI 2003*, ACM Press (2003), 113-120.