Evaluation of Korean Text Entry Methods for Smartwatches

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

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

ABSTRACT
This paper presents results from a user study designed to evaluate the effectiveness of Korean text entry methods for smartwatches. Specifically, the study compares the four popular text entry methods for smartphones in the context of smartwatch use (three multi-tap 3x4 keypad methods and a QWERTY-like method). A distinctive feature of text entry in Korea is that traditionally different manufacturers have developed their own text entry methods starting with particular physical layouts on feature phones that are now available as soft keypads on smartphones. This research considers the next step in this progression by studying the viability of adopting these text entry methods on smartwatches. The results from the user study indicate that existing methods can be effective for text entry on smartwatches; analysis of the data offers suggestions for improving the effectiveness of the methods.

ACM Classification Keywords
H.5.2. User Interfaces: Evaluation/Miscellaneous,Input devices and strategies

Author Keywords
text entry; soft keyboard; smartwatch.

INTRODUCTION
The introduction of smartwatches has brought back the challenge of providing efficient and effective techniques for small screen sizes [1]. Recent work on text entry methods for smartwatches (e.g. DualKey [4], ZoomBoard [12], Swipeboard [3], SplitBoard [7]) has focused on adapting the QWERTY layout to the constraints of the small screen size. This work aims to evaluate the effectiveness of existing 3x4 layouts for text entry; a QWERTY-like layout was also included as a base line.

The context of this work is Korean text entry. Korean users have long faced multiple choices for text entry on their mobile devices. Starting with feature phones different manufacturers offered their own fixed keypad layouts for the traditional 3x4 keypad. This trend continued with the introduction of smartphones, with the added flexibility that the users can now choose their preferred text entry method. The larger screen size of smartphones has made it possible to adopt a QWERTY-like layout for Korean smartphones, but its effectiveness on a smartwatch has been questioned due to the limited space in which to fit 26 alphabetic keys. Previous work has studied Korean text entry on feature phones [8], which represent a significantly different form factor and mode of interaction (physical vs. soft keys, two-thumb vs. single-finger typing).

Korean text entry methods were chosen for the study, since there are multiple alternatives that have been proven effective and have been adopted by Korean users. In fact, some manufacturers are already providing a soft keyboard for a smartwatch based on one of the layouts considered in this study. While these are essentially multi-tap methods, they do have unique characteristics and offer innovative solutions to mapping an alphabet to a small set of keys. Note that Korean is an alphabetic language, so the results in this work could inform research on text entry methods for English.

CHI 2017, May 06–11, 2017, Denver, CO, USA.
 ISBN 978-1-4503-4655-9/17/05...$15.00.
DOI: http://dx.doi.org/10.1145/3025453.3025657

KOREAN TEXT ENTRY
Figure 1 shows the layouts that were analyzed in this study: (*i*) Sky (SKY), which was introduced by SK Telecom; (*ii*) *Chon-Ji-In* (CJI), which was popularized by Samsung; (*iii*) *Na-Rat-Gul* (NRG), which was developed by LG; and (*iv*) *QWERTY-like* (2SET), which is the Korean layout for a standard keyboard. In the rest of the paper the term *layout* will be used for *text entry method*.

The Korean language is alphabetic and consists of 40 letters: 14 basic consonants, 5 double consonants, 10 basic vowels, and 11 vowel diphthongs. The basic forms can be seen in Fig. 1(2SET) with the vowels occupying the right half of the keypad. The 11 vowel diphthongs are typically not mapped directly to keys on a layout, but are formed by combining the basic vowels (e.g. ㅗ+ㅏ= ㅘ, ㅏ+ㅣ=ㅐ).

A unique feature of the Korean writing system is that the letters are “packed” into syllables of 2 to 4 letters, which gives Korean writing a distinctive appearance — for example, the translation of the phrase *Korean script* has the letter sequence ㅎㅏㄴㄱㅡㄹ, but is written as 한글.

Finally, Korean letters for similar sounds tend to have similar shapes: ㄱ/ㅋ (g/k), ㄷ/ㅌ (d/t), ㅏ/ㅑ (a/ya). This has offered opportunities for creative solutions of layout design.

The rest of this section introduces the 4 layouts used in this study. The key mappings and an example key sequence to enter the word 뺏속 are shown in Fig. 1 and Fig. 2, respectively.

SKY CJI NRG 2SET

	Phrases	Words
total:	400	1402
min len:	8	2
max len:	35	19
avg len:	25.42	6.54
letters: 10169		
correlation with Korean: 0.993		

Figure 1. The four layouts used in this study and phrase set statistics; the numeric labels are only for reference and do not appear on the interface. The *backspace* key deletes only the last letter of the last syllable.

ㅃ	ㅕ	ㅅ		ㅅ	ㅗ	ㄱ
888	66	77	↲	77	9	1

SKY

ㅃ	ㅕ	ㅅ		ㅅ	ㅗ	ㄱ
777	221	8	↲	8	23	4

CJI

ㅃ	ㅕ	ㅅ	ㅅ	ㅗ	ㄱ
5⊕⊗	33⊗	7	7	6	1

NRG

Figure 2. Key sequences for the word 뼛속 assuming the standard numbering on a 3x4 keypad; ↲ is used for key segmentation in SKY and CJI.

SKY Layout
SKY is a 3x4 multi-tap layout. Letters are grouped on the individual keys based on similarity of sound. The vowels occupy the last column and the middle key on the top row. To enter a letter the corresponding key is tapped as many times as indicated by the position of the letter on the key (e.g. 8 (ㅂ), 88 (ㅍ), 888 (ㅃ)). SKY requires a method for segmentation when two letters on the same key need to be entered after each other. For this study the only means for key segmentation was via a single tap of the ↲ control key (timeout was not used).

CJI Layout
CJI shares some similarities with SKY with respect to the consonants. Its distinguishing characteristic is that the vowels are not explicitly represented on the keypad. Instead they are composed with the keys on the top row [ㅣ • ㅡ], since each vowel either has a horizontal or a vertical main stroke. The • is used to complete a vowel by adding the shorter stroke(s) and the order relative to the main stroke is important. For example, ㅣ + • = ㅏ, but • + ㅣ = ㅓ; similarly, ㅡ + • = ㅜ, but • + ㅡ = ㅗ.

CJI also requires a means for segmentation. This layout has the advantage of reducing the number of keys required to compose the vowels. The disadvantage is that it has higher gestures per vowel and increased mental load due to the order in which the gestures are to be performed.

Note that CJI does not use the bottom-left/right letter keys. This seems rather wasteful for a smartwatch where space is limited. While it would have been possible to try a variation that uses all keys (for example, distributing ㅎ, ㅇ, ㅁ on the bottom row, so that each requires one gesture), ultimately the standard layout was used for the following reasons:

1. Manufacturers are likely to offer the standard form on a smartwatch (e.g. LUNA G from SK Telecom) and it would be useful to discover its performance characteristics.
2. The availability of the two extra keys makes it possible to have immediate access to punctuation symbols and emojis.
3. To some extent SKY is a variant of CJI that uses all keys.

In fact, the results suggest that leaving out the two keys did not affect the performance of CJI relative to the other layouts.

NRG Layout
NRG is conceptually different from the previous layouts. Its design is inspired by the structure of Korean letters and uses the concepts of *adding a stroke* (⊕) and *doubling* (⊗), which are assigned to the bottom row [⊕ — ⊗]. Only a small subset of letters is available on the keypad and the rest are composed via these two transformations. For example:

ㅁ ⊕ = ㅂ, ㅂ ⊗ = ㅃ

ㅂ ⊕ = ㅍ

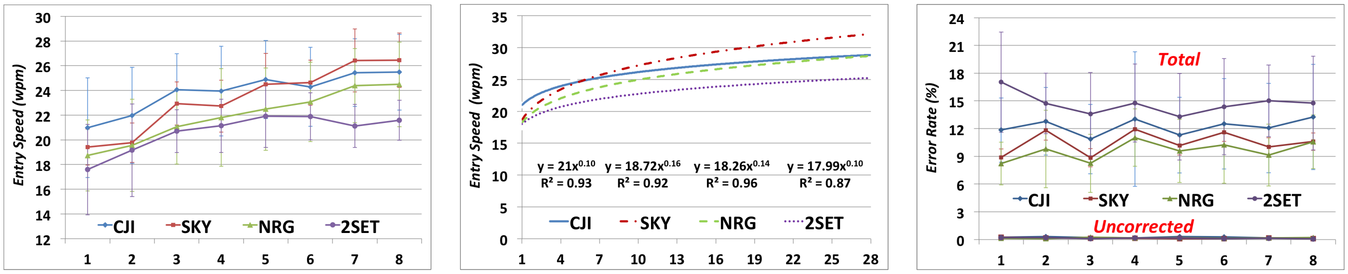
NRG does not require segmentation. It has been considered to be faster than the other layouts for expert users, but it has higher mental load due to the rules of composing the letters.

2SET Layout
2SET is essentially the analog of QWERTY for Korean keyboards. The letters are mapped on a regular keyboard, including 4 of the diphthongs. The 5 double consonants and 2 of the diphthongs are produced by first tapping a *shift* key (⇑).

2SET has the advantage that most letters require just one gesture and Korean users are already familiar with it. As with the QWERTY layout for English, the challenge of offering a full 2SET layout on a smartwatch is that the small size of the keys could make it impractical for text entry. 2SET was included in the study as a baseline for this and future work.

EVALUATION
The performance characteristics of the four layouts were assessed through a formal user study that involved a text copy task. The study was conducted at a university in Seoul, Korea.

Participants
A total of 32 participants (16 male, 16 female; age range [20, 26], μ = 22.6) were recruited for the study. All participants were native Koreans and none had used a smartwatch previously. Only one participant was left-handed.



(a) (b) (c)

Figure 3. Performance averaged over all users per session with 95% CI: (a) text entry speed; (b) speed extrapolated to 28 sessions; (c) error rates.

Apparatus

The device used was an LG G Watch with a 29.6 x 29.6 mm touchscreen. The presented phrase was shown at the top of the screen and the participants typed underneath.

Phrase Set

The phrases used in this study were derived from the phrase set in [8], which is a translation of the phrase set in [11]. The phrases in [8] were adapted by the second author (a native Korean), so that no phrase was longer than 14 syllables. Thus, the longest presented phrase with 35 letters could fit within the screen due to the rules of the Korean script of packing letters into syllables. Figure 1 shows the phrase set statistics computed using the software in [11] and the letter frequencies reported in [9] for estimating the correlation with Korean.

Design

The study was gender balanced and used between-subject design with 8 participants per layout. The participants were assigned a layout that they had not used previously or were least familiar with based on a self-report questionnaire. The one exception was 2SET, since all participants were familiar with it and 21 participants were using 2SET on their smartphones.

The recorded data was used to analyze text-entry speed (WPM), and total error rate (TER). Participants also completed a NASA-TLX survey [5] (Korean translation [10]).

Procedure

Each participant attended 8 sessions over 4 consecutive days and entered 50 phrases per session for a total of 400 phrases per participant. Each day consisted of 2 sessions separated by a 5 minute break; within a session there was a forced 2 minute break after 15 minutes. A typical session lasted 15–20 minutes. The participants wore the watch on their non-dominant hand and entered the phrases with their dominant hand.

The first day included short training to familiarize the participants with their assigned layout. After the participants were briefed about the purpose of the study and signed a consent form they were given a manual describing the layout and were asked to enter 15 phrases. During training a new phrase was shown only when the previous one had been entered correctly or the participant exceeded 5 attempts.

During testing only one attempt was allowed per phrase. The participants were instructed to enter the phrases as accurately and as quickly as possible and to ignore errors beyond the last 2-3 syllables. Nevertheless, the participants tended to correct most errors as observed in [13, 14].

RESULTS

Text-Entry Speed

Figure 3a shows the average text-entry speed per session for each layout. 2SET is slowest across all sessions achieving 21.6 WPM by the end of the study ($\mu = 20.6, \sigma = 3.12$), while SKY starts to dominate around session 6 and achieves 26.4 WPM ($\mu = 23.4, \sigma = 3.42$). CJI is the fastest in the first half of the study, but ends up slightly behind SKY with 25.5 WPM ($\mu = 23.6, \sigma = 4.15$). Finally, NRG is ahead of 2SET achieving 24.5 WPM ($\mu = 22.0, \sigma = 4.26$).

ANOVA revealed significant effect of *layout* and *session* on *text-entry speed* (Table 1), the latter confirming the presence of learning effect. A post hoc Tukey HSD showed that CJI was faster than NRG and 2SET, and that SKY was faster than 2SET at $p < .05$. The other differences were not significant. In contrast NRG was faster than CJI and SKY on feature phones [8], which might be due to the advantage afforded by two thumb typing that is lost on a smartwatch—on average NRG requires longer travel distance with single-finger typing to compose a letter with the ⊕ and ⊗ keys.

Source	df	*F*	*p*	*F*	*p*
layout	3	10.43	< .001	15.44	< .001
session	7	9.91	< .001	1.12	> .05
layout*session	21	0.52	> .05	.48	> .05
Error	224				
		WPM		TER	

Table 1. ANOVA results for WPM and TER.

Figure 3b shows the learning curve extrapolated to 28 sessions. The data suggests that NRG may continue to improve and achieve performance similar to CJI. Despite the expectation that 2SET will be impractical on a smartwatch the participants achieved an average speed over 20 WPM and seemed to plateau at 21.5 WPM around session 5. This is fairly good considering the constraints on key size. Also, CJI ranked at the top even though it only uses 10 of the available 12 keys.

Error Rate

Figure 3c shows the average TER and UER per session for each layout. As mentioned previously, despite the instructions

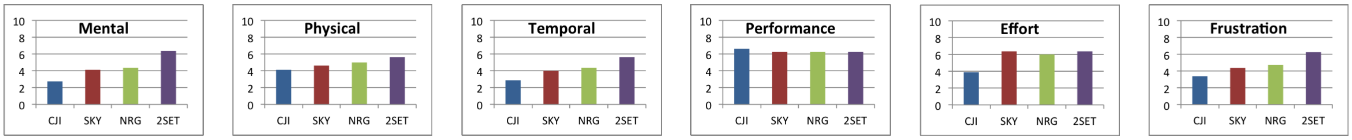


Figure 4. Summary of the qualitative feedback from the 10-point scale NASA-TLX questionnaire.

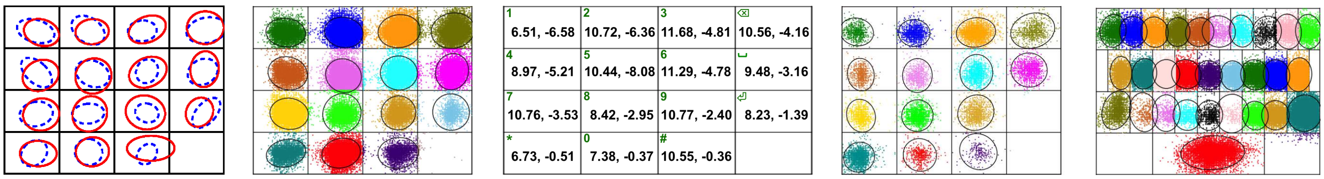


Figure 5. Touch distributions with 95% confidence ellipses: (a) SKY session 1 (blue, dashed) and 8 (red, solid); (b) SKY all sessions; (e) average (dx, dy) offsets in pixels between key and ellipse centers for SKY; (d) NRG left-handed all sessions; (e) 2SET all sessions.

to balance speed and accuracy the participants tended to correct most errors, which is reflected in the very low UER. As expected 2SET had the highest TER ($\mu = 15.1, \sigma = 5.31$) due to the very small key sizes. NRG had the lowest TER ($\mu = 9.6, \sigma = 4.18$) across all sessions followed by SKY ($\mu = 10.5, \sigma = 3.98$) and CJI ($\mu = 12.4, \sigma = 5.76$).

ANOVA revealed a significant effect of *layout* on *total error rate* (Table 1). A post hoc Tukey HSD showed that 2SET had higher TER than the other methods, and CJI had higher TER than NRG, and this difference was significant at $p < .05$.

Qualitative Feedback

After the study each participant filled out a NASA-TLX questionnaire to rank the various dimensions of the task workload (Fig. 4). Of particular interest are the responses to the Mental, Temporal, and Frustration dimensions. (The responses to Performance and Effort suggest that the participants did not have a good frame of reference of what constituted good performance.) As expected 2SET had the highest expressed levels of workload, followed by NRG. Interestingly, CJI was perceived as less demanding across all dimensions, despite the reduced set of available keys and higher difficulty for composing the vowels. SKY, which is a fairly straightforward layout, was found more demanding than CJI.

Touch Distributions

Interesting patterns were observed in the data from the touch distributions. In general, for the 3x4 layouts the spread of the distributions was larger in the last session compared to the first (Fig. 5a). The larger spread seems to correspond to faster text entry speed — in the first session the participants were slower and tentative, which resulted in a more concentrated area of touches on the keys. By the end the increase in speed resulted in less precise, though still effective, key taps.

A second observation for the 3x4 layouts is that, in general, there was a skew toward the bottom-right corner of each key (Fig. 5b,c — only SKY shown due to space constraints); for the single left-handed participant the skew was toward the bottom-left corner (Fig. 5d). This corresponds to the relative orientation between the line of sight and direction of approach to the keypad, and seems to be a result of the participants' intent to compensate for the occlusion problem. The data for 2SET showed an overall tendency to touch below the key centers (Fig. 5e), which is consistent with the findings in [6, 2].

A possible direction for future work could be to study the impact on error rate of the compensation functions described in [6] that shift the coordinates of the touch by a fixed or dynamically computed offset. The results in [6] for QWERTY layout on smartphones showed a decrease in error rate and it would be important to confirm this effect for smartwatches.

DISCUSSION

This paper presented results from a user study on the effectiveness of 3x4 and QWERTY-like layouts for Korean text entry on smartwatches. The results show that these are viable options that achieve competitive speed at reasonable error rate.

Although the study was on Korean text entry, it could inform work on English, particularly 3x4 layouts, which have been studied less extensively. Note that the Korean layouts arrange the letters so that the consonants and vowels occupy two distinct areas (vowels in top row of CJI, right column of SKY and NRG, and right half of 2SET). This arrangement is well suited for two thumb typing on a smartphone, since within a phrase consonants and vowels alternate which generates a rhythm for typing. This might also enhance text entry on a smartwatch, since the vowels are accessible in a fixed location that facilitates recall. Here are possible modifications of the standard English layouts inspired by the consonant-vowel split with only minimal changes and treating y as a vowel:

ae	**io**	**uy**
bc	df	gh
jk	lm	npq
rs	tvw	xz

q w r t **y a e u i o**
s d f g h j k l p
z x c v b n m

Similarly, one might expect comparable performance for the English QWERTY layout. While the participants expressed higher levels of frustration with 2SET, they still achieved over 20 WPM, albeit with 14.7 TER. Since 2SET has higher demand than QWERTY (7 letters require *shift* key), it seems that QWERTY should be closer to 2SET in speed. The data reported for QWERTY [7, 12] does not support this observation, which might be due to differences in experimental setup.

REFERENCES

1. Ahmed Sabbir Arif and Ali Mazalek. 2016. *A Survey of Text Entry Techniques for Smartwatches*. Springer International Publishing, Cham, 255–267. DOI: http://dx.doi.org/10.1007/978-3-319-39516-6_24
2. Shiri Azenkot and Shumin Zhai. 2012. Touch Behavior with Different Postures on Soft Smartphone Keyboards. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '12)*. ACM, New York, NY, USA, 251–260. DOI: http://dx.doi.org/10.1145/2371574.2371612
3. Xiang 'Anthony' Chen, Tovi Grossman, and George Fitzmaurice. 2014. Swipeboard: A Text Entry Technique for Ultra-small Interfaces That Supports Novice to Expert Transitions. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 615–620. DOI: http://dx.doi.org/10.1145/2642918.2647354
4. Aakar Gupta and Ravin Balakrishnan. 2016. DualKey: Miniature Screen Text Entry via Finger Identification. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 59–70. DOI: http://dx.doi.org/10.1145/2858036.2858052
5. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183. DOI: http://dx.doi.org/10.1016/S0166-4115(08)62386-9
6. Niels Henze, Enrico Rukzio, and Susanne Boll. 2012. Observational and Experimental Investigation of Typing Behaviour Using Virtual Keyboards for Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2659–2668. DOI: http://dx.doi.org/10.1145/2207676.2208658
7. Jonggi Hong, Seongkook Heo, Poika Isokoski, and Geehyuk Lee. 2015. SplitBoard: A Simple Split Soft Keyboard for Wristwatch-sized Touch Screens. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1233–1236. DOI: http://dx.doi.org/10.1145/2702123.2702273
8. Ivaylo Ilinkin and Sunghee Kim. 2010. Evaluation of Text Entry Methods for Korean Mobile Phones, a User Study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 2023–2026. DOI: http://dx.doi.org/10.1145/1753326.1753633
9. H. Kim and Kang B. 1997. *Frequency analysis of Korean*. Korea University Institute of Korean Culture.
10. Wonsup Lee, Sujin Kim, Sunghye Yoon, Jangwoon Park, Dalho Lee, Seikwon Park, Byunggil Kang, Jooho Uem, and Heecheon Yoo. 2008. Development of a comprehensive pilot workload assessment model for evaluation of helicopter cockpit design. In *Proceedings of the Fall conference of the Ergonomics Society of Korea (ESK '08)*. 306–309.
11. I. Scott MacKenzie and R. William Soukoreff. 2003. Phrase Sets for Evaluating Text Entry Techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. ACM, New York, NY, USA, 754–755. DOI: http://dx.doi.org/10.1145/765891.765971
12. Stephen Oney, Chris Harrison, Amy Ogan, and Jason Wiese. 2013. ZoomBoard: A Diminutive Qwerty Soft Keyboard Using Iterative Zooming for Ultra-small Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2799–2802. DOI: http://dx.doi.org/10.1145/2470654.2481387
13. R. William Soukoreff and I. Scott MacKenzie. 2003. Metrics for Text Entry Research: An Evaluation of MSD and KSPC, and a New Unified Error Metric. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 113–120. DOI: http://dx.doi.org/10.1145/642611.642632
14. Jacob O. Wobbrock, Brad A. Myers, and Htet Htet Aung. 2004. Writing with a Joystick: A Comparison of Date Stamp, Selection Keyboard, and EdgeWrite. In *Proceedings of Graphics Interface 2004 (GI '04)*. Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 1–8. http://dl.acm.org/citation.cfm?id=1006058.1006059