Encouraging Consistency in Mobile Text Entry Evaluations

Steven J. Castellucci
Dept. of Computer Science and Engineering
York University
4700 Keele Street
Toronto ON M3J 1P3 Canada
stevenc@cse.yorku.ca

I. Scott MacKenzie
Dept. of Computer Science and Engineering
York University
4700 Keele Street
Toronto ON M3J 1P3 Canada
mack@cse.yorku.ca

CHI'12, May 5–10, 2012, Austin, Texas, USA.
Workshop on Designing and Evaluating Text Entry Methods

Abstract
We developed TEMA, an application to gather text entry speed and accuracy metrics on Android devices. TEMA is intended to bring consistency to the evaluation of text entry methods for Android devices. As a demonstration of this, we present an evaluation of four mobile text entry methods: two-thumb QWERTY typing, one-finger QWERTY typing, handwriting recognition, and shape writing recognition. The two QWERTY techniques were the fastest, with no statistically significant difference between them in entry speed or accuracy. Shape writing was slightly slower, but similar in accuracy. Handwriting was the slowest and least accurate technique**.**

Author Keywords
Text entry, metrics, entry speed, accuracy, Android OS

ACM Classification Keywords
H.5.2 Information interfaces and presentation (e.g., HCI): User Interfaces---evaluation/methodology.

General Terms
Human Factors, Performance, Measurement

Introduction
Mobile devices are ubiquitous in contemporary society and interacting with them often involves text entry. To facilitate consistent evaluations of mobile text entry

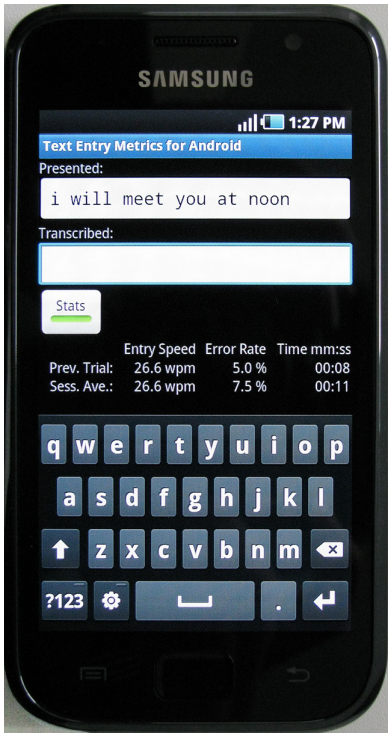

Figure 1. The TEMA application installed on a smartphone.

methods, we created an application to gather metrics on Android devices: Text Entry Metrics on Android (TEMA) [2].

This position paper describes TEMA's features, our motivation for developing it, and our position on conducting consistent, empirical evaluations of mobile text entry methods. In addition, we summarize a currently unpublished user study conducted using TEMA.

Motivation

Mobile devices are often used for SMS text messaging and social networking. An estimated 7 trillion text messages were sent in 2011 [1] and more than 350 million users access Facebook from a mobile device [3]. Thus, investigating and developing methods for mobile text entry is a significant research topic.

The prevalence of mobile devices encouraged us to develop mobile versions of *Graffiti*, H4-Writer [7], and SAK [5] text entry techniques. This necessitated a program to gather performance metrics – a mobile equivalent to *TextTest* [10] for the PC. The result, TEMA, is an application to aid mobile text entry researchers using Android devices.

The choice to target Android devices was simple. The Android operating system powers over 250 million [8] mobile devices, including smartphones and tablets. Text entry is accomplished using physical keyboards, soft keyboards, shape writing [4], handwriting recognition, or voice recognition. This variety exists because anyone can freely develop and distribute an Android text input method (IME in developer parlance). Android is the only popular mobile platform to allow third-party text entry methods.[1] These IMEs can be used system-wide, without modifying installed applications. Consequently, TEMA can run on a vast number of mobile devices and form factors, each capable of using a variety of IMEs.

The ability to evaluate text entry techniques in a consistent manner is very important. Such evaluations allow for meaningful comparisons between input methods and between studies. Unfortunately, some published user studies differ on how entry speed and accuracy are measured. Sometimes, accuracy measures are omitted completely. We hope the availability of TEMA facilitates the consistent gathering and reporting of established entry speed and accuracy metrics (detailed in the Features section below).

A consistent methodology for evaluation should be adopted. In situations where an established convention might not suffice (e.g., non-prose or non-alphabetical text entry), researchers should detail how metrics were calculated.

TEMA Features

TEMA (Figure 1) is a ready-made application to aid researchers gathering text entry metrics on Android devices. It occupies little storage space (only 125 kB) and has the following features:

Entry speed metric: Entry speed is calculated by dividing the length of the transcribed text by the entry time (in seconds), multiplying by sixty (seconds in a minute), and dividing by five (the accepted word

[1] *ShapeWriter* was approved for the *iPhone* [12], but is no longer available.

length, including spaces [11]). Thus, the result is reported in words-per-minute (wpm).

Accuracy metrics: Accuracy is evaluated according to the total error rate (TER), corrected error rate (CER), and uncorrected error rate (UER) metrics [9]. TER characterizes general input accuracy and is the sum total of CER and UER. CER reflects the errors that the participant corrected during transcription, while UER reflects the errors that the participant did not correct. All three error rates are reported as a percent.

Stats log: The "stats" log summarizes entry speed, the accuracy metrics mentioned above, and intermediate measurements (e.g., presented text, transcribed characters, elapsed time, etc.) for each trial.

Event log: The "event" log contains time-stamped (in milliseconds) input events for low level, post-study analysis. Both event and stats logs are saved to the Android device's internal storage.

Set of 500 phrases: The text presented for transcription is randomly chosen from a 500-phrase set [6].

User Study

The TEMA pilot study [2] was small, with only six participants of varying experience. However, the study summarized here involved sixteen novice participants and included four counterbalanced conditions: QWERTY using two thumbs; QWERTY using a finger; *DioPen* (www.diotek.com), a handwriting technique; and *Swype* (www.swypeinc.com), a shape writing technique. The single-finger QWERTY condition represents an alternative QWERTY input method and allows comparisons with the single-finger handwriting and shape writing input techniques. Entry speed and accuracy results appear in Figure 2 and Figure 3, respectively.

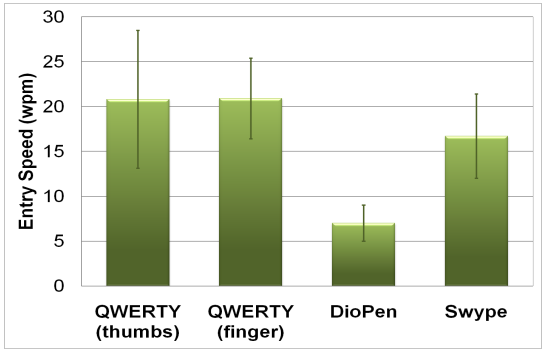

Figure 2. Entry speed values gathered by TEMA. Error bars represent ±1 *SD*.

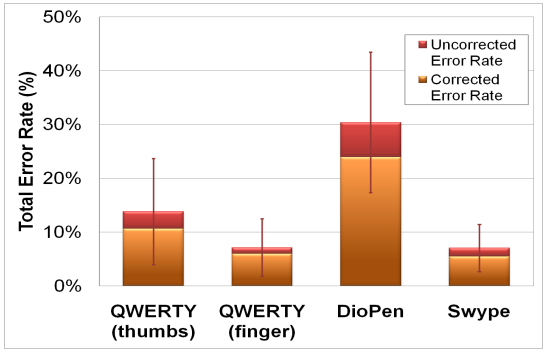

Figure 3. Accuracy values gathered by TEMA. Error bars represent ±1 *SD* of total error rate.

An analysis of variance revealed a significant effect of technique on entry speed ($F_{3,36} = 71.17$, $p < .0001$). Surprisingly, there was no significant difference between the two QWERTY conditions; many believe two-

thumb input to be a faster method. Every other pairwise comparison proved to be significant. There was also a significant effect on total error rate ($F_{3,36}$ = 41.66, p < .0001). However, post hoc analysis indicated a significant difference only between *DioPen* and all other conditions.

Conclusion

To facilitate meaningful comparisons of text entry methods, researchers must evaluate techniques using the same metrics. TEMA provides a consistent platform for mobile text entry research on Android devices. It includes hundreds of phrases for text entry, measures timings, calculates established entry speed and accuracy metrics, and generates easily viewable log files for post-study analysis.

The conducted study demonstrated TEMA's utility. Despite the perceived advantage of two-thumb input, there was no statistically significant difference between the two QWERTY conditions with respect to either entry speed or accuracy. Shape writing was slightly slower, but not significantly less accurate. Handwriting was both slow and error-prone.

We hope other researchers will find TEMA and the results derived from our user study beneficial to their mobile text entry research. Additional TEMA information is available at www.cse.yorku.ca/~stevenc/tema.

References

[1] ABI Research, *More than seven trillion SMS messages will be sent in 2011*. (Accessed on December 29, 2010.) http://www.abiresearch.com/press/3584-More+than+Seven+Trillion+SMS+Messages+Will+Be+Sent+in+2011.

[2] Castellucci, S. J. and MacKenzie, I. S., Gathering text entry metrics on android devices. In *Ext. Abs. CHI 2011*, ACM (2011), 1507-1512.

[3] Facebook, *Statistics*. (Accessed on January 3, 2012.) http://www.facebook.com/press/info.php?statistics.

[4] Kristensson, P. O., Discrete and continuous shape writing for text entry and control. Doctoral dissertation. *Dept. of Computer and Information Science*, Linköping University, 2007.

[5] Mackenzie, I. S. and Felzer, T., SAK: Scanning ambiguous keyboard for efficient one-key text entry, *ACM Transactions on Computer-Human Interaction (TOCHI)*, *17*(3), 2010, 39 pages.

[6] MacKenzie, I. S. and Soukoreff, R. W., Phrase sets for evaluating text entry techniques. In *Ext. Abs. CHI 2003*, ACM Press (2003), 754-755.

[7] MacKenzie, I. S., Soukoreff, R. W., and Helga, J., 1 thumb, 4 buttons, 20 words per minute: Design and evaluation of H4-Writer. In *Proc. UIST 2011*, ACM Press (2011), 471-480.

[8] Schonfeld, E., *Android phones pass 700,000 activations per day, approaching 250 million total*. (Accessed on January 3, 2012.) http://techcrunch.com/2011/12/22/android-700000/.

[9] Soukoreff, R. W. and MacKenzie, I. S., Recent developments in text-entry error rate measurement. In *Ext. Abs. CHI 2004*, ACM Press (2004), 1425-1428.

[10] Wobbrock, J. and Myers, B., Analyzing the input stream for character-level errors in unconstrained text entry evaluations, *ACM Transactions on Computer-Human Interaction*, *13*(4), 2006, 458-489.

[11] Yamada, H., A historical study of typewriters and typing methods: From the position of planning Japanese parallels, *Journal of Information Processing*, *2*(4), 1980, 175-202.

[12] Zhai, S., Kristensson, P. O., Gong, P., Greiner, M., Peng, S. A., Liu, L. M., and Dunnigan, A., ShapeWriter on the *iPhone*: From the laboratory to the real world. In *Ext. Abs. CHI 2009*, ACM Press (2009), 2667-2670.