Gathering Text Entry Metrics on Android Devices

Steven J. Castellucci

Department of Computer Science
and Engineering

York University

4700 Keele St.

Toronto, Ontario M3J 1P3 Canada

stevenc@cse.yorku.ca

I. Scott MacKenzie

Department of Computer Science
and Engineering

York University

4700 Keele St.

Toronto, Ontario M3J 1P3 Canada

mack@cse.yorku.ca

Abstract

We developed an application to gather text entry speed and accuracy metrics on Android devices. This paper details the features of the application and describes a pilot study to demonstrate its utility. We evaluated and compared three mobile text entry methods: QWERTY typing, handwriting recognition, and shape writing recognition. Handwriting was the slowest and least accurate technique. QWERTY was faster than shape writing, but we found no significant difference in accuracy between the two techniques.

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

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

Introduction

Text entry on mobile devices is an important research topic, as many people communicate via SMS messages (a.k.a. text messages). Analysts predict more than seven *trillion* SMS messages will be sent worldwide in 2011 [1]. In addition, smartphones facilitate Internet searching, email composition, and document editing. To

Figure 1. The TEMA application (above) can be downloaded from www.cse.yorku.ca/~stevenc/tema.

aid evaluation of mobile text entry methods, we created an application to gather metrics on Android devices: Text Entry Metrics on Android (TEMA).

Android is a mobile operating system developed and marketed by Google. Since Android's initial release in late 2008, it has surpassed iOS, Research In Motion, and Windows Mobile in global smartphone market share [9]. In addition, technology experts believe Android tablet PCs will be very popular in 2011 [3].

Applications to gather text entry metrics exist for other platforms, but not for Android. *TextTest* [10] does not rely on any specific text entry technique, but only works on desktop PCs. An unnamed iOS application [2] uses only the QWERTY keypad. In contrast, Android is ideal for mobile text entry research. Applications are developed using Java syntax and Android libraries. Android allows users to develop interchangeable text input methods (called IMEs in developer parlance). These IMEs can be used system-wide without modifying installed applications. Consequently, TEMA can be run on a vast number of mobile devices and form factors, each capable of using a variety of IMEs.
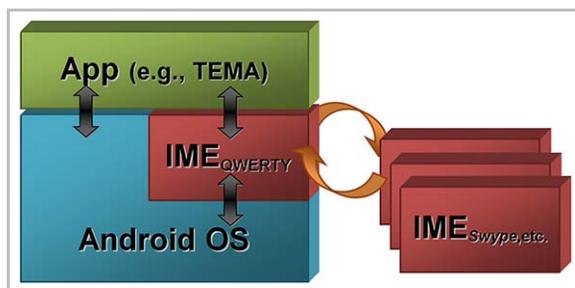

Figure 2. The input method (IME) can be changed without requiring any modification to TEMA.

TEMA Features

In addition to calculating text entry metrics, TEMA has additional features to assist researchers:

Stats log: The "stats" log records the start and end time of each evaluation session. It summarizes entry speed, multiple error rate metrics [8], and intermediate measurements (e.g., presented text, transcribed characters, elapsed time, etc.) for each trial.

Event log: The "event" log contains time-stamped input events for low level, post-study analysis. When connected to a PC, Android devices typically appear as removable drives, thus simplifying log retrieval.

Landscape orientation: Although TEMA will rotate its screen layout, landscape orientation is only recommended on devices with a physical keyboard. Most onscreen IMEs in landscape orientation obscure the presented text field – participants would not see the phrase to enter.

Set of 500 phrases: The text presented for transcription is randomly chosen from a 500-phrase set [7].

Interruption timer: Although interruptions are not recommended during evaluation sessions, TEMA measures the duration of interruptions (e.g., an incoming phone call, task switching, etc.) and deducts it from the transcription time.

Trial management: Text entry trials can be refreshed (with a new phrase) or reset (with the same phrase) if a participant gets distracted or pauses unnecessarily during a trial; all measurements are reset. The ignored trial appears in the events log, but not in the stats log.
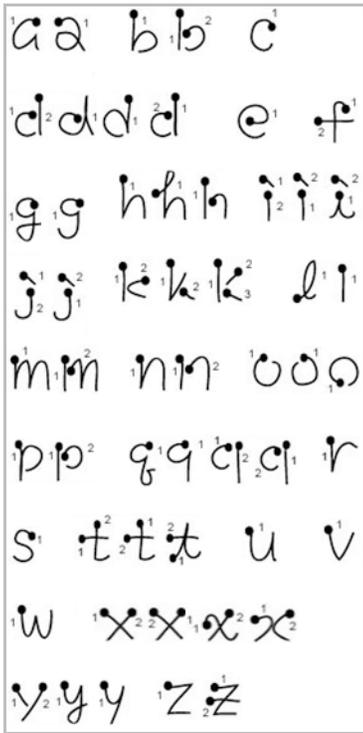
Figure 3. The *DioPen* alphabet for lowercase letters [4].

Method

To demonstrate the practicality of TEMA, we conducted a small user study comparing three of the IMEs installed on an Android smartphone.

Participants

Six volunteer participants (five males, one female) were recruited from the local university campus. Ages ranged from 24 to 35 years ($\mu = 29$; $\sigma = 4.17$). Two participants were left-handed. Although participants were knowledgeable about handwriting and shape writing techniques, none used them regularly. Instead, participants used a QWERTY keypad. With the prevalence of smartphones, we were unable to find participants who were novices with QWERTY keypads.

Apparatus

The TEMA application ran on a Samsung *Galaxy S Vibrant* (GT-I9000M) smartphone (seen in Figure 1 with the default QWERTY keypad) with Android OS v2.1. The touch screen measured 4.0 inches diagonally and had a resolution of 480×800 pixels. The phone was held in portrait orientation throughout the study. No screen protector or case was used to affect touch screen sensitivity or device handling. The phone's wireless radios were disabled to eliminate disruptions due to incoming calls or text messages.

Three of the IMEs included with the phone were evaluated with TEMA: Samsung's default QWERTY keypad, *DioPen*, and *Swype*. For each IME, the input language was set to English (US) and options for auto-spacing, auto-capitalization, and word prediction were deactivated.

DioPen (www.diotek.com) is a handwriting recognition technique. Users enter letters by tracing gestures on the input area with their finger (Figure 4). The gestures resemble handwriting and can be composed of up to three separate strokes (Figure 3). Some letters are associated with multiple gestures to allow for variations in handwriting input.

Swype (www.swypeinc.com) uses shape writing recognition to perform word-based input. Users draw a path on the QWERTY keypad starting at the first letter of the desired word and intersecting each subsequent letter (Figure 5). The resulting sequence (including unintentional letters along the path) forms a shape that (ideally) is unique to the desired word. If a collision occurs (i.e., the shape matches multiple words), the user selects the desired word from a short list.

Procedure

Participants entered ten phrases using each IME. They were instructed to enter text as quickly as possible, to correct errors if noticed immediately, but to ignore errors made two or more characters back. To type on the QWERTY keypad, all participants chose to use two thumbs. With *DioPen* and *Swype*, they held the device in their non-dominant hand and used a finger on their dominant hand to input gestures.

Before each condition, participants were instructed on how to use the corresponding technique. A practice session followed, consisting of three random phrases. Study sessions typically lasted 20 minutes and took place in a quiet office, with participants seated at a desk.

Figure 4. The letter "e" being entered using *DioPen*.

Figure 5. The word "the" being entered on the *Swype* keypad.

Design

The experiment employed a within-subjects factor, technique, with three levels: QWERTY, *DioPen*, and *Swype*. The order of testing was counterbalanced using a Latin Square. Each condition involved ten phrases of text entry. Phrases were chosen randomly (without replacement) from a 500-phrase set [7]. The phrases were converted to lowercase letters and did not contain any numbers or punctuation.

The dependent variables were entry speed and accuracy. Entry speed was 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 length, including spaces [11]). The entry speed was averaged over the ten phrases and reported in words-per-minute (wpm).

Accuracy was measured according to the total error rate (TER), corrected error rate (CER), and uncorrected error rate (UER) metrics [8]. CER reflects the errors that the participant corrected during transcription, while UER reflects the errors that the participant did not correct. TER characterizes general input accuracy and is the sum total of CER and UER. Error rates were averaged over the ten phrases and reported as a percent.

Results and Discussion

Accuracy

Although there was a significant effect of technique on total error rate (TER) ($F_{2,10} = 10.76$, $p < .005$), there was no significant difference between QWERTY and *Swype* ($p > .05$). The TER of *Swype* was the lowest, at 6.2%. Interestingly, an evaluation of *ShapeWriter*,

another shape writing technique, revealed a similar TER value of 6.7% [6, pp. 65-66]. The QWERTY TER of 11.8% is also comparable to the 10.4% measured using the *iPhone*'s QWERTY keypad [2]. Unfortunately, *DioPen* had the worst TER, at 25.0%. An evaluation of *Graffiti 2* handwriting recognition, reported elsewhere, revealed an error rate of 19.4% [5].
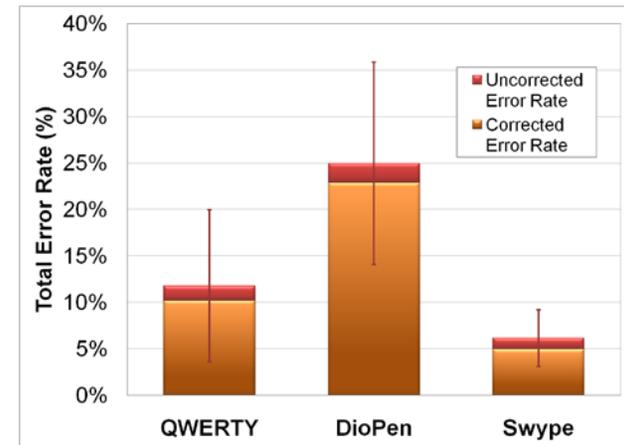

Figure 6. Accuracy values gathered by TEMA. Error bars represent ±1 standard deviation.

In general, the low UER results of our study suggest participants were diligent in correcting errors. Unfortunately though, the high CER values for QWERTY and *DioPen* indicate many errors were committed during input. Fortunately the TEMA event logs allowed further analysis of participants' input.

With QWERTY, some participants missed typing a space character. Instead, they typed "v", "b", or omitted the space entirely. Against our instructions, the participants then deleted entire words to insert the missing space.

We later determined that the position of our space-bar varied by about 1 mm (0.04 inches) from the QWERTY keypad on the participants' phones. That difference seems to have impacted participants' performance.

The *DioPen* event logs revealed multiple attempts to enter characters (i.e., participants entered an incorrect character, backspaced, entered another incorrect character, backspaced, etc.). Many gestures were not recognized correctly. One participant mentioned that *DioPen* was difficult because the gesture alphabet did not resemble his own handwriting. Other participants seemed to write the gestures at an angle, which likely affected recognition.

Although *Swype* was the most accurate, it would occasionally produce an incorrect word. To correct it, participants would tap backspace repeatedly to delete the word. We have since learned that pressing and holding backspace deletes the last word. Using this method likely would have reduced *Swype*'s CER.

Entry Speed
There was a significant effect of technique on entry speed ($F_{2,10} = 65.17$, $p < .0001$). The QWERTY rate of 21.4 wpm is the fastest in our study. It is even higher than the 15.9 wpm reported for the *iPhone*'s QWERTY keypad [2]. *DioPen* was the slowest technique at 6.1 wpm. This is probably related to the high rate of gesture misrecognition, which required users to correct their input. A study evaluating *Graffiti 2* yielded a slightly better rate of 9.2 wpm [5]. Our *Swype* entry speed of 17.4 wpm is consistent with a *ShapeWriter* study that reported 15 wpm after five minutes of practice and 20 wpm after twenty minutes of practice [6, pp. 65-66].
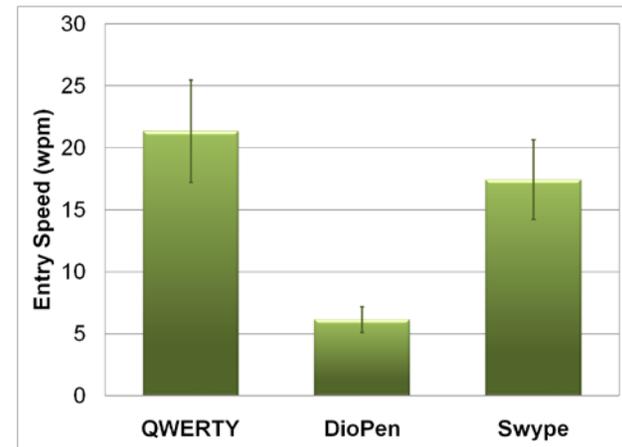

Figure 7. Entry speed values gathered by TEMA. Error bars represent ±1 standard deviation.

Future Work
The metrics-gathering functionality of TEMA is complete. We plan to test it on other devices and IMEs, and to investigate the following enhancements:

File browser for log directory: Currently, researchers must ensure the default log directory exists and is writable. Researchers specify a different directory by typing its entire path. With an integrated file browser, researchers could easily navigate the file system to select a directory or create a new one.

Log file viewer: TEMA logs are created as comma separated vector (CSV) files. There exist CSV viewer applications. However, an integrated log viewer would simplify software installation and could also summarize session statistics.

Custom default parameters: Allowing researchers to specify default study parameters would minimize their involvement during evaluation sessions.

Auditory and/or haptic feedback: Providing feedback using sound or vibration could affect text entry speed or accuracy. Allowing this option in TEMA would provide additional evaluation conditions.

Automatic IME switching: To evaluate numerous IMEs, researchers must change OS settings when needed. By allowing researchers to specify an IME sequence, TEMA could switch IMEs after a specified number of phrases.

Conclusion

The conducted study demonstrated how TEMA can benefit researchers. The stats logs summarized entry speed and accuracy metrics for three distinct IMEs, while the event logs revealed participant tendencies and IME shortcomings. Our TEMA application gathers text entry metrics on Android devices, regardless of the input method used.

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] Arif, A. S., Lopez, M. H., and Stuerzlinger, W., Two new mobile touchscreen text entry techniques. Poster at the *36th Graphics Interface Conference.* 2010, CEUR-WS.org/Vol-588 22-23.

[3] Brustein, J., Rivals to the iPad say this is the year, in *The New York Times*, January 3, 2011, p. B1. http://www.nytimes.com/2011/01/03/technology/personaltech/03tablet.html.

[4] DIOTEK, *DioPen K for Android (english manual)*. (Accessed on December 28, 2010.) http://www.diotek.com/?ui=customer_DownManual&Div=DioPen.

[5] Költringer, T. and Grechenig, T., Comparing the immediate usability of Graffiti 2 and virtual keyboard. In *Extended Abstracts CHI 2004*, ACM Press (2004), 1175-1178.

[6] Kristensson, P. O., *Discrete and continuous shape writing for text entry and control*. Unpublished doctoral dissertation, Linköping University, Linköping, Sweden, 2007.

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

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

[9] Tudor, B. and Pettey, C., *Gartner says worldwide mobile phone sales grew 35 percent in third quarter 2010; smartphone sales increased 96 percent*, Gartner, Egham, UK, November 10, 2010. http://www.gartner.com/it/page.jsp?id=1466313.

[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.