A Practical Examination of Multimodal Feedback and Guidance Signals for Mobile Touchscreen Keyboards

Tim Paek[*], Kenghao Chang[**], Itai Almog, Eric Badger, Tirthankar Sengupta

Microsoft Research[*] and Microsoft Corp.
One Microsoft Way
Redmond, WA 98052
{timpaek | ebadger | itaia | tsengupt}@microsoft.com

University of California, Berkeley[**]
281 Hearst Memorial Mining Bldg 1764
Berkeley, CA 94720
kenghao@eecs.berkeley.edu

ABSTRACT
Mobile devices with touch capabilities often utilize touchscreen keyboards. However, due to the lack of tactile feedback, users often have to switch their focus of attention between the keyboard area, where they must locate and click the correct keys, and the text area, where they must verify the typed output. This can impair user experience and performance. In this paper, we examine multimodal feedback and guidance signals that keep users' focus of attention in the keyboard area but also provide the kind of information users would normally receive in the text area. We evaluated whether combinations of multimodal signals could improve typing performance in a controlled experiment. One combination reduced keystrokes-per-character by 8% and correction backspaces by 28%.

Categories and Subject Descriptors
H5.2 [**Information Interfaces and Presentation**]: User Interfaces - *Graphical user interfaces (GUI).*

General Terms
Human Factors

Keywords
Multimodal feedback, touchscreen, soft keyboard, mobile device

1. INTRODUCTION
Mobile devices with capacitive or resistive touch capabilities often utilize an on-screen, virtual keyboard, or *touchscreen keyboard* for text input (see [7] for a general survey). Because touchscreen keyboards are software-based, they can be easily adjusted for different languages, screen orientation, and key layouts. Furthermore, they can be augmented with widgets for word prediction and disambiguation candidates. On the other hand, touchscreen keyboards have a significant disadvantage in that they lack the tactile affordances of physical hardware. In particular, tactile feedback contributes to the consistency of finger movements during typing [14] and lets users know when they have touched, clicked and slipped away from a key [5]. Without tactile feedback, users often have to switch their focus of attention between the *keyboard area*, where they must locate and hit the correct keys, and the *text area*, where they must verify the typed output. This switching can impair typing user experience and performance. For example, as users focus on targeting in the keyboard area, they may miss typing errors or auto-corrections in the text area. If errors compound, users will have to spend more time engaged in post-hoc editing, which is both challenging on a touchscreen [14] and mentally disruptive. Indeed, researchers have found that users generally type slower on a touchscreen keyboard than on a physical keyboard [5], and fail to notice typing mistakes as often [2].

In this paper, we examine different types of multimodal feedback and guidance signals that keep users' focus of attention in the keyboard area but also provide the kind of information users would normally get in the text area. Because our goal is to deploy a commercial product that can be easily adopted, we consider only multimodal signals for QWERTY keyboards. One of the signals has already been shown in previous research to improve typing performance. However, for commercial deployment, we need to identify combinations of multimodal signals that enhance the overall typing user experience. Given our practical imperative, this paper consists of two contributions. First, we explore three types of multimodal feedback and guidance signals[1] that keep users focused on the keyboard area. Second, we evaluate whether combinations of signals can improve typing performance in a controlled experiment.

2. MULTIMODAL SIGNALS
A great deal of previous research has explored the benefits of equipping mobile devices with tactile feedback [2][5]. While equipping touchscreen keyboards with tactile feedback is certainly a promising direction, researchers have not thoroughly examined whether similar results can be achieved augmenting the standard touchscreen keyboard with more visual and auditory signals. From a practical perspective, visual and auditory signals are also much easier to deploy and cheaper than hardware innovations.

With no tactile feedback on mobile touchscreen keyboards, users have to monitor their fingers to make sure they are targeting the right keys, but when they do, they can miss important feedback in

[1] See [13] a longer technical report which includes more signals as well as a usability study aimed at 1) refining their interaction design and 2) finding combinations of signals that users prefer. We also discuss UX design implications.

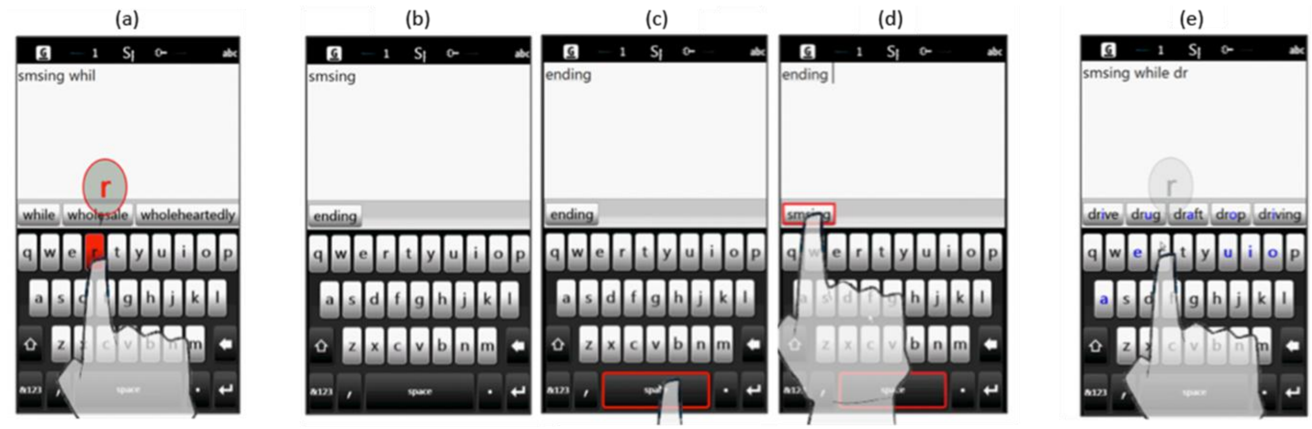

Figure 1. Visual design for the (a) unexpected-key feedback, (b)(c)(d) auto-correction feedback, and (e) key prediction guidance.

the text area. As such, we sought multimodal signals around the keyboard area that could also convey text area information. We investigated three types of signals which answer the following questions: 1) Did I just type a word incorrectly? 2) Did a word I typed just change? 3) Where is my next key? The first two signals provide *feedback* about events users would normally discern when they are monitoring the text area. The third signal provides *guidance* about how to avoid typing mistakes.

2.1 Did I Just Type a Word Incorrectly?

Many touchscreen keyboards utilize a *candidates area* above or below the keyboard area where they display widgets containing word candidates (e.g., HTC and Android-based smartphones). Following [12], candidates are typically word predictions, though nowadays word disambiguation candidates are commonly included. As visual signals, the presence of widgets in the candidates area conveys to users that they can quickly auto-complete a word or auto-correct an incorrect word by touching a widget [3]. These visual signals belie sophisticated typing intelligence technologies for dealing with noisy input [4] and can be used to alert users to unexpected keys and possibly an incorrect word. Users would normally discern when they have typed an incorrect word by constantly monitoring their typed output. As a consequence, we decided to create a multimodal signal called "*unexpected-key feedback*" in the keyboard area to alerts users to possible errors so that they can immediately switch their focus of attention to the text area or to the candidates area. For this signal, we piggybacked the design on the tooltip balloon (e.g. iPhone [6]) and added a distinct auditory signal. To our knowledge, no prior research has explored this kind of unexpected-key signal.

Figure 1(a) shows the signal's visual design, which evolved from a series of usability refinements [13]. As the user types an 'r' after 'whil' in Figure 1(a), the key and the tooltip balloon turn red, both of which slowly fade back to their original grey color. Furthermore, instead of the usual "click" sound for the fingertip-click event, a distinct "clunk" sound is played. Note that some usability participants found the sound to be sufficient feedback, whereas others preferred just the visual, and still others both.

In terms of implementation, a key was considered "unexpected" when the characters entered so far did not match the prefix of a word that existed in our typing intelligence dictionary. Our dictionary is a professionally reviewed and morphologically inclusive set of over 78K English words and acronyms.

2.2 Did a Word I Typed Just Change?

In attending to the keyboard area, users sometimes fail to see auto-corrections in the text area that may be replacing legitimate words such as proper nouns and technical terms that do not exist in the dictionary. For example, on the iPhone, as the user types an unknown word, a predicted word appears below the typed output which then replaces the unknown word at a word boundary. This can lead to tremendous frustration, especially if users do not notice the text replacements until later and have to edit.

Figure 1(b)(c)(d) depict how we ultimately designed a multimodal signal for "*auto-correction feedback*". After the user has typed the unknown word 'smsing' in Figure 1(b), as the user clicks the space bar, a red border appears around the button (see Figure 1(c)) and a distinct "swish" sound is played (as if something was quickly replaced). The audio signal here is absolutely essential because fast typists are not likely to notice the visual feedback. If the user desires to put back their replaced word, they can click the replaced word, which now appears with a red border in the candidates area, as shown in Figure 1(d). This reverses or undo's the replacement. Our design for the auto-correction feedback is similar to how Kristensson and Zhai [8] visually highlighted auto-corrected words in their elastic stylus keyboard.

2.3 Where is my Next Key?

Besides feedback signals that provide information normally conveyed in the text area, we decided to examine a guidance signal acclaimed in the research literature. In particular, previous studies explored the text entry benefits of highlighting the next predicted key. Perhaps the most conspicuous guidance signal was utilized by Al Faraj et al. [1] in BigKey, a mobile QWERTY soft keyboard, where they dynamically adjusted the visual size of the next likely keys by their probabilities. Despite the constant adjustment of the keyboard layout, users of BigKey were surprisingly 25% faster and more accurate. Given such prior success, we decided to implement a signal for "*key-prediction guidance*". We hypothesized that this signal might guide user who are uncertain about how to spell a word into the correct characters. In this way, key-prediction guidance is closely linked to word prediction. Indeed, we made this link explicit in our visual design.

Figure 1(e) shows the signal for key-prediction guidance, where the next likely keys are colored blue in the letters on the buttons. Initially, we colored the entire key button blue but some usability participants found this too distracting. By coloring just the letter, we found a subtle visual cue which users who were looking for guidance could grab hold of and those who were not could ignore. To prevent the entire keyboard from turning blue, we showed the visual cue only after the second letter of a word, and highlighted up to five letters at most. These letters had to correspond to word prediction candidates in the candidates area. We did not give any auditory signals. In terms of implementation, we generated prediction candidates by performing prefix matches against word entries in our 78K+ English dictionary [13] and then highlighted the next likely character based on the top-ranked candidates.

3. Experiment

Before examining text entry performance, we conducted a usability study in which 11 participants were asked to type phrases using a variety of multimodal signals and combinations thereof (see [13] for details). We asked them to identify which multimodal signals they would leave on by default and why. Overall, we found that all participants wanted the auto-correction feedback on by default because they were frustrated to discover unwanted auto-corrections. As such, we decided to deploy the auto-correction feedback. Furthermore, we found that participants did not perceive any conflict with combining the auto-correction feedback with either the unexpected-key feedback or the key-prediction guidance, both of which garnered praise from some usability participants who said that they perceived improved performance. With respect to the unexpected-key feedback, some participants remarked on how it made the candidates area more useful – that is, by alerting them to disambiguation candidates that corrected their text. With respect to the key-prediction feedback, some participants explained how they relied on it for spelling.

In order to examine whether the multimodal signals could in fact improve text entry beyond perceived performance, we conducted a controlled text entry experiment comparing three *SignalType* conditions, our primary independent variable: (1) unexpected-key feedback combined with auto-correction feedback (**UnexpectedKey**+), (2) key-prediction guidance combined with auto-correction feedback (**KeyPredict+**), and (3) auto-correction feedback alone as a baseline (**Baseline**). We included auto-correction feedback in every condition because we had already decided to deploy the signal. We used this experiment to decide whether to deploy either the unexpected-key feedback or the key-prediction guidance as well. Indeed, (1) and (2) allowed us to gauge the text entry performance of the combination of signals.

As our dependent variables, we examined the efficiency measure *keystrokes-per-character* (KSPC) [10] and the number of times users pressed the backspace key. Because we did not allow users to place the cursor onto their typed text for editing and selecting, pressing backspace was the only way users could correct text. Hence, the number of backspaces is a proxy for corrections.

We recruited 18 participants (9 males and 9 females) between the ages of 21-39 using the same professional contracting service as before. Participants came from a wide variety of occupational backgrounds. All participants were compensated for their time. During recruiting, all participants answered that they were familiar with the QWERTY layout and could type on a normal-size keyboard without frequently looking at the keys.

For stimuli, we utilized MacKenzie and Soukoreff's [11] phrase set. To ensure that participants had a chance to hit every letter on the keyboard, we wrote a script to select the shortest sequences of phrases that covered the entire alphabet from a–z. For each condition, subjects received 8 practice and 20 stimuli items. The practice items were introduced to avoid a learning effect.

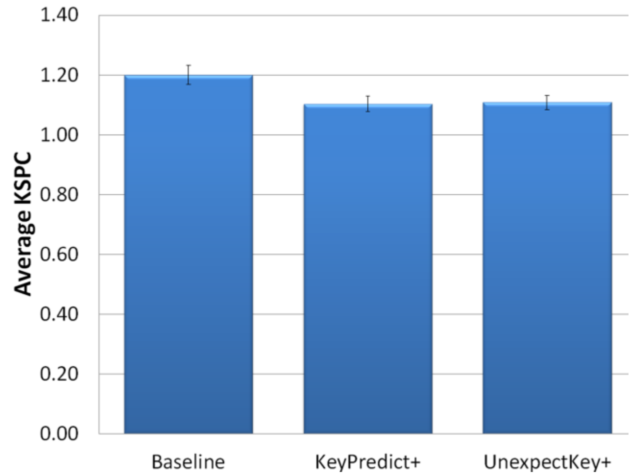

Figure 2. Average keystrokes-per-character (KSPC) for *SignalType* conditions with standard errors of the mean.

Participants were then asked to enter text into the mobile device according to the following procedure. We first displayed a target phrase on a desktop computer screen and asked participants to memorize it with as much time as they needed. We asked them to memorize the phrases to mimic the experience of entering intended text. When participants felt they were "ready", their task was to type the phrase into the mobile device "as quickly and as accurately as possible". Timing began as soon as they entered the first letter of the phrase and ended when they hit the 'Enter' button twice. The entire experiment took slightly under 2 hours.

Overall, we conducted a repeated measures design study where all participants received all *SignalType* conditions as a within-subjects variable in different counter-balanced orders.

3.1 Results

In terms of KSPC, we hypothesized that **UnexpectedKey+** would exhibit lower KSPC than the **Baseline** because if users do in fact use feedback to select disambiguation candidates in the candidates area, then that should save them keystrokes. Likewise, we hypothesized that **KeyPredict**+ would exhibit would lower KSPC than the **Baseline** because guidance into the correct spelling should save participants erroneous keystrokes. Indeed, we found a significant main effect for *SignalType* ($F_{2,712} = 5.25$, $p<.01$). As shown in Figure 2, **UnexpectedKey+** ($\mu = 1.11$) had significantly lower KSPC than the **Baseline** ($\mu = 1.20$; $p < .01$) and so did **KeyPredict**+ ($\mu = 1.10$; $p < .01$). However, the two were not statistically different.

In terms of corrections, we hypothesized that **UnexpectedKey+** would result in fewer backspaces than the **Baseline** by alerting users to incorrect keys before they continue to add more characters. We also hypothesized that **KeyPredict**+ would reduce the number of backspaces by steering users away from incorrect spellings. Indeed, we found a main effect for *SignalType* ($F_{2,712} = 5.01$, $p < .01$). However, the only significant difference was between **UnexpectedKey+** ($\mu = 2.46$) and the **Baseline** ($\mu = 3.41$; $p < .01$). Figure 3 shows the average number of backspaces for the *SignalType* conditions.

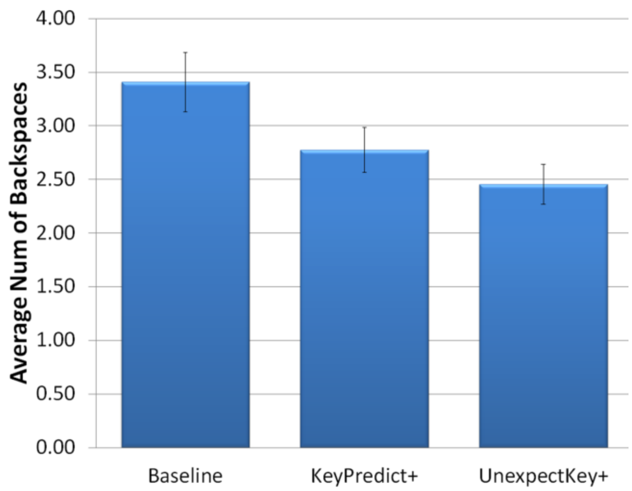

Figure 3. Average number of backspaces for the *SignalType* conditions with standard errors of the mean.

After the user experiment, we asked participants to pick their favorite *SignalType* condition and to rank-order which of the three they would leave on by default. 13/18 participants picked the **UnexpectedKey+** condition as their favorite. No one picked the **Baseline**. With respect to rank-ordering, 9/18 listed **UnexpectedKey+** at the top and 8/18 participants listed **KeyPredict**+ at the top. The fact that only 9, and not 13, of the participants said they would leave **UnexpectedKey+** on by default implies that although some participants found that particular combination of signals to be their favorite condition, they might prefer to use it only as desired.

3.2 Experiment Discussion

Despite the fact that the **Baseline** condition for *SignalType* included the Auto-correction feedback, which in theory could have made it harder to find significant differences, we still managed to find differences for time to enter text, KSPC and number of backspaces. For all of the two dependent variables, **UnexpectedKey+** emerged as the best combination of signals. In summary, **UnexpectedKey+** reduced KSPC by 7.7%, and reduced the number of backspaces by 27.9%.

With respect to limitations, our results are limited by the form factor of our test device. As shown recently by Lee & Zhai [9], the type of touch sensor can affect the performance of touchscreen widgets. For our studies, we used a resistive touchscreen primarily because that was the only available prototype device for our product. In terms of other directions for future research, it is best to conduct longitudinal studies to verify our performance differences over the long-term. Although we provided plenty of practice for users to learn each *SignalType* condition, performance differences may fade away with accumulated learning.

4. Conclusion

In this paper, we introduced and motivated the need for multimodal signals that provide feedback and guidance to users in the keyboard area. We described three implemented multimodal signals that answer different questions. Unexpected key feedback answers “Did I just type a word incorrectly?”, auto-correction feedback answers, “Did a word I typed just change?”, and key-prediction guidance answers “Where is my next key?” The first two signals provide feedback about events users would normally discern when they are monitoring the text area. The third signal provides guidance about how to avoid typing mistakes. We evaluated whether two combinations of signals, unexpected-key feedback + auto-correction feedback and key-prediction guidance + auto-correction feedback, could also improve typing performance in a controlled experiment. The former significantly reduced keystrokes-per-character by 8% and reduced backspaces by 28%. Finally, we summarized everything we learned about designing multimodal signals with design implications.

5. REFERENCES

[1] Al Faraj, K., Mojahid, M., & Vigouroux, N. 2009. BigKey: a virtual keyboard for mobile devices. *Proc. Of Intl HCI*, 3-10.

[2] Brewster, S., Chohan, F., & Brown, L. 2007. Tactile feedback for mobile interactions. In *Proc. of CHI*, 159-162.

[3] Chan, C. 2009. HTC Hero: Software & HTC Sense review. Posted 10-15-2009: http://www.androidcentral.com/htc-hero-software-htc-sense-review. Retrieved 1-25-2010.

[4] Goodman, J., Venolia, G., Steury, K., & Parker, C. 2002. Language modeling for soft keyboards. In *Proc.of AAAI*, 419-424.

[5] Hoggan, E., Brewster, S., & Johnston, J. 2008. Investigating the effectiveness of tactile feedback for mobile touchscreens. In *Proc. of CHI*, 1573-1582.

[6] http://www.apple.com/iphone/

[7] Kölsch, M. & Turk, M. 2003. Keyboards without keyboards: A survey of virtual keyboards. *Technical Report 2002-21*. University of California, Santa Barbara.

[8] Kristensson, P. & Zhai, S. (2005) Relaxing stylus typing precision by geometric pattern matching. *Proc. of IUI*, 151-158.

[9] Lee, S. & Zhai, S. 2009. The performance of touch screen soft buttons. *Proc. of CHI*, 209-318.

[10] MacKenzie, I. S., & Tanaka-Ishii, K. 2007. *Text entry systems: Mobility, accessibility, universality*. San Francisco: Morgan Kaufmann Publishers.

[11] MacKenzie, I. S. & Soukoreff, R. 2003. Phrase sets for evaluating text entry techniques. In *Extended Abstracts of CHI 2003*, 754-755.

[12] Masui, T. 1999. POBox: An efficient text input method for handheld and ubiquitous computers. H. Gellersen, Ed. *Lecture Notes in Computer Science*, 1707, 288-300.

[13] Paek, T., Chang, K., Almog, I., Badger, E., Sengupta, T. 2010. Multimodal feedback and guidance for mobile touchscreen keyboards. MSR-TR-2010-76. http://research.microsoft.com/apps/pubs/?id=132400

[14] Rabin, E. & Gordon, A.M. 2004. Tactile feedback contributes to consistency of finger movements during typing. *Experimental Brain Research*, 155(3), 1432-1106