# Multimodal Feedback and Guidance Signals for Mobile Touchscreen Keyboards


Tim Paek[*], Itai Almog, Eric Badger, Tirthankar Sengupta
Microsoft Research[*] and Microsoft Corp.
One Microsoft Way
Redmond, WA 98052
{timpaek | ebadger | itaia | tsengupt}@microsoft.com

Kenghao Chang
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 get in the text area. We first conducted a usability study to assess and refine the user experience of these signals and their combinations. Then we evaluated whether those signals which users preferred could also improve typing performance in a controlled experiment. One combination of multimodal signals significantly improved typing speed by 11%, reduced keystrokes-per-character by 8%, and reduced backspaces by 28%. We discuss design implications.



## 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 [13] for a general survey). Without the requirement of a physical keyboard, touchscreen keyboards enable larger displays for videos, web pages, email, etc. [9]. 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 [26] of physical hardware, which makes noisy input much more likely [7]. In particular, tactile feedback contributes to the consistency of finger movements during typing [28] and lets users know when they have touched, clicked and slipped away from a key [9]. 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 [27] and mentally disruptive. Indeed, researchers have found that users generally type slower on a touchscreen keyboard than on a physical keyboard [9], and fail to notice typing mistakes as often [4].

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. Although different keyboard layouts can improve text entry on a touchscreen keyboard (see [23] for a review), these layouts have not yet become widespread. Because our goal is to deploy an easily adopted commercial product, we consider only multimodal signals for QWERTY keyboards. Some of the signals have already been shown in previous research to improve typing performance. However, for commercial deployment, typing performance in an experimental setting is not enough. We need to identify combinations of multimodal signals that enhance the overall typing user experience. Given our practical imperative, this paper consists of four contributions. First, we investigate four types of multimodal feedback and guidance signals, and describe related research in general and with respect to each signal. Second, we discuss the results of a usability study we conducted in order to assess and refine the user experience of these signals, and combinations thereof. Third, we evaluate whether two combinations of signals users preferred in the usability study could actually also improve typing performance in a controlled experiment. Fourth, we conclude with implications of our findings on design of touchscreen keyboards.

## 2. RELATED RESEARCH
A great deal of previous research has explored equipping mobile devices with tactile feedback, such as a piezo-electric tactile display that responds to a stylus [12], and a stylus that generates tactile feedback [16]. For touchscreen keyboards, researchers in

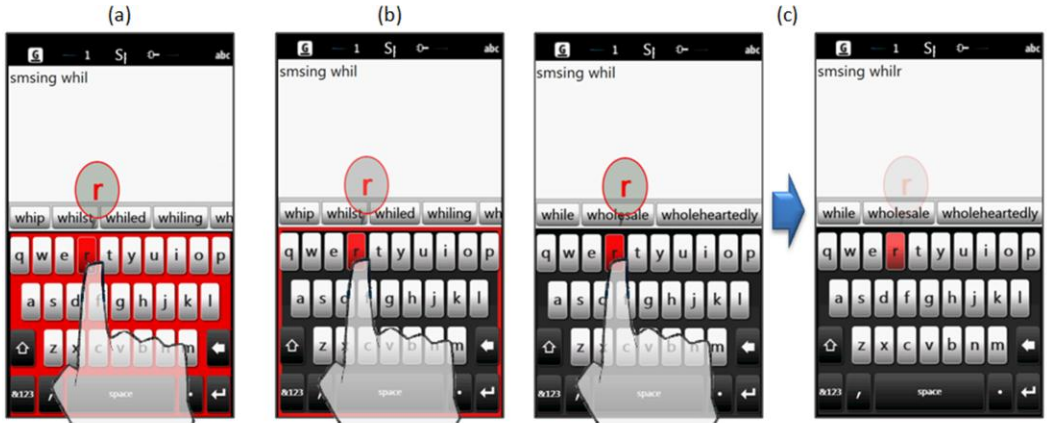


**Figure 1. Evolution of the visual design for the unexpected-key feedback. As users type an unexpected key, they not only see red coloring but they also hear a "clunk" sound.**

two related studies [9][4] demonstrated that inducing artificial tactile feedback via vibration actuators can indeed improve text entry performance in both static and mobile environments (e.g., while typing on a train). In [9], Hoggan et al. created Tactons [5] for three types of keyboard events: a *fingertip-over* event to signal when a fingertip has touched a key, a *fingertip-click* event to signal that a key has been registered, and a *fingertip-slip* event to signal when the fingertip moved over the edge of a key. With these Tactons, and a fourth one for indicating where the home 'F' and 'J' keys were located, their tactile touchscreen keyboard obtained accuracies close to those of a physical keyboard.

While equipping touchscreen keyboards with tactile feedback is certainly a promising direction, Hoggan et al. did not examine whether similar results were achievable by providing more visual and auditory signals to the standard touchscreen keyboard. Indeed, participants in their study even complained that their fingertips obfuscated the visual feedback. However, most commercially available touchscreen keyboards (e.g., iPhone [10]) offer some visual and auditory feedback for the three keyboard events: typically, a tooltip balloon appears and disappears at the fingertip-over and fingertip-slip events respectively, and an audible click is played at fingertip-click events.

Although it has been argued that visual and auditory signals can be ineffective in mobile settings due to small screen size, outside noise, social restrictions and other circumstantial demands [19], it has not been empirically established that artificial tactile feedback alone can impart enough information to realize full parity with physical keyboards. Furthermore, some users find tactile feedback annoying [9]. Because individual differences may account for users' preference of one modality over another, it is incumbent upon the research community to analyze a variety of tactile, visual and auditory feedback techniques, and mixtures thereof, for touchscreen keyboards. Because prior work has already highlighted the benefits of tactile feedback, here we primarily explore the possible benefits of visual and auditory signals. From a practical perspective, visual and auditory signals, as software solutions, are also much easier to deploy and cheaper than hardware innovations.

# 3. MULTIMODAL SIGNALS

With no tactile feedback on mobile touchscreen keyboards, users cannot just monitor their typed output as their fingers find the right keys – i.e., they cannot touch-type as they would on a physical keyboard. They have to periodically monitor their fingers in the keyboard area. To help users with this task, current touchscreen keyboards typically employ tooltip balloons and click sounds as feedback about what keys were actually clicked. This led us to ponder what other kinds of multimodal signals could be leveraged in the keyboard area. In particular, we sought multimodal signals that could also provide the kind of information users would normally get in the text area. We concentrated on the keyboard area because the lack of tactile feedback and the small form factor of mobile devices require users to actively look at the keyboard due to the difficulties of targeting by touch [27]. By providing text area information in the keyboard area, we sought to minimize the constant switching of users' focus-of-attention.

We investigated four types of signals which could help users answer the following questions:

1. Did I just type a word incorrectly?
2. Did a word I typed just change?
3. Am I still on the right track?
4. Where is my next key?

The first three signals provide *feedback* about events users would normally discern when they are monitoring the text area. The fourth signal provides *guidance* about how to avoid typing mistakes. In describing each multimodal signal, wherever appropriate, we point to related research specific to the signal. We also discuss how we refined the design of each signal with respect to the usability study, which is covered in more detail in the next section.

## 3.1 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 [25], 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 a possibly erroneous sequence of characters by touching a widget (sometimes, the most likely candidate is marked further, as for example on [6]). These visual signals belie sophisticated typing intelligence technologies for

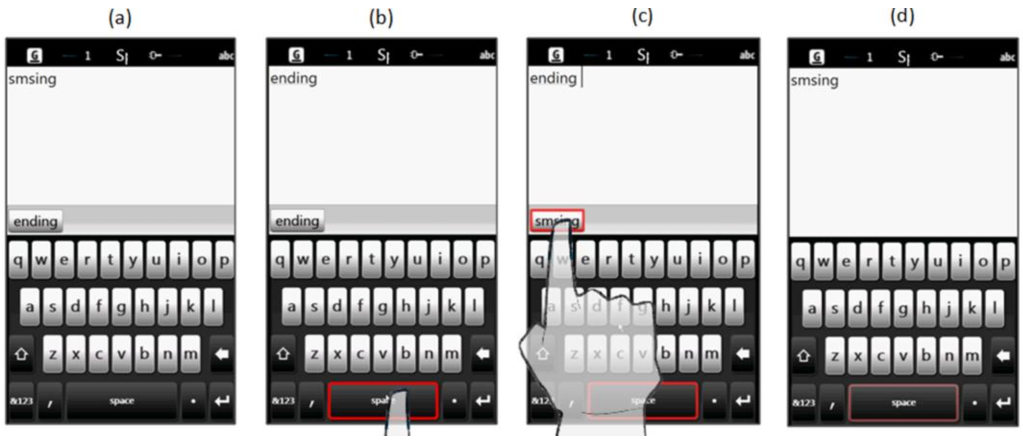


**Figure 2. The auto-correction feedback in action. When unknown words are auto-corrected, users see a red border around the space and hear a "swish" sound.**

dealing with noisy input [7]. Indeed, the same technologies can be utilized to alert users to when they are typing an unexpected key and therefore a possibly incorrect word. Users would normally discern when they have typed an incorrect word by constantly monitoring their typed output. However, we decided to create a multimodal signal called "*unexpected-key feedback*" in the keyboard area which would alerts users to possible errors so that they could immediately, only when necessary, switch their focus of attention from keyboard area to the text area or to the candidates area. For this signal, we piggybacked the design on the tooltip balloon and added a distinct auditory signal.

Figure 1 shows the evolution of the unexpected-key feedback's visual design. As the user types an 'r' after 'whil' in Figure 1(a), the key, the keyboard background, and the tooltip balloon turn red. Furthermore, instead of the usual "click" sound for the fingertip-click event, a distinct "clunk" sound is played. In exhibiting this to participants in the usability study, we discovered that Figure 1(a) was "too disruptive", so we toned down the red keyboard background to a red border around the keyboard, as in Figure 1(b). Users still found this distracting, so we removed the background coloring. As evident in Figure 1(c), only the clicked key and its corresponding balloon are red, both of which slowly fade back to their original grey color. Most users found this combination of visual and auditory feedback agreeable and certainly more pleasant than previous iterations.

In terms of implementation, a key was considered "unexpected" when the characters entered so far (up to the previous word boundary) 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. Our use of a dictionary is similar to the TypeRight keyboard [8], which adjusts the resistance of its buttons to make it momentarily harder to press keys that are deemed likely to lead to typing errors, according to its dictionary and typographic rules. Unlike the TypeRight keyboard, however, our unexpected-key feedback signal functions more as a feedback signal rather than a guidance signal.

Using a dictionary or language model to handle unexpected keys on a software-based keyboard was first utilized by Goodman et al. [7] in their source-channel formulation [29] of the noisy input problem. Although they mostly utilized their formulation to resize the underlying hit target areas of next likely keys, their soft keyboard did immediately correct previously accepted characters. This visual feedback was very subtle and occurred only in the text area.

## 3.2 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 much later and then have to engage in post-hoc editing.

Figure 2 depicts how we designed a multimodal signal for "*auto-correction feedback*". After the user has typed the unknown word 'smsing' in Figure 2(a), as the user clicks the space bar, a red border appears around the button (see Figure 2(b)) 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 2(c). This reverses or undo's the replacement. When we exhibited this to users in the usability study, we received only positive comments about the design.

Our design for the auto-correction feedback is similar to what Kristensson and Zhai [15] employed for their elastic stylus keyboard. They pattern-matched stylus tap traces against ideal point templates that represent entire words to auto-correct noisy words. Auto-corrected words were then visually highlighted with a colored background in the text area. When tapped, a widget containing a list of n-best candidates would appear, allowing users to change the word (see [14] for more details).

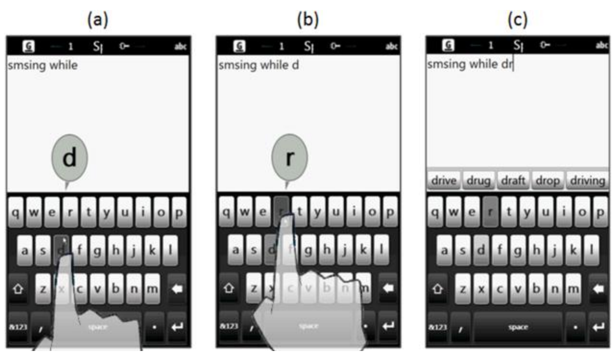


**Figure 3. The key-trail feedback in action. As users press a key, the button turns darker and slowly fades to its original color, even as new buttons are pressed.**

## 3.3 Am I Still on the Right Track?

Users typing quickly do not always notice the tooltip balloons. To help users keep track of the keys they just typed in the keyboard area – a task normally accomplished by monitoring the text area, one visual feedback Baudisch et al. [3] explored is to highlight past user actions on a widget with an afterglow effect. For example, the on-screen keyboard in the Windows 7 Operating System utilizes an afterglow effect whereby pressed keys briefly light up and gradually fade away [11]. As users type quickly, they observe a lit-up trail of key-presses fading away like fingerprints on foggy glass. We wanted to see if this afterglow effect on a smaller mobile keyboard could convey the same kind of useful information as monitoring typed output in the text area, so we decided to implement a "*key-trail feedback*". Note that in some ways the key-trail feedback provides complementary information to the unexpected-key feedback in that when users monitor text, they register not only unexpected keys but also expected, correct key sequences.

Figure 3 shows the signal in action. As the user clicks a 'd' in Figure 3(a) followed by a 'r' in Figure 3(b), the buttons turn dark and slowly fades back to their original color. The 'dr' trail is still visible in Figure 3(c). We did not add an accompanying auditory signal we thought would match the visual feedback. In the usability study, we tested various fading rates for the trail. For fast typists, if the trail lingered too long, the keyboard would be almost entirely marked. After adjusting the rates, we finally settled on having buttons return to their original color by .75 second, which seemed to satisfy fast and slow typists.

## 3.4 Where is my Next Key?

So far we have only examined feedback signals that provide information normally obtained in the text area. We decided to include one guidance signal primarily because of its acclaimed benefits in the research literature. In particular, previous studies explored the text entry benefits of highlighting the next predicted key. Magnien et al. [24] found that by bolding the next likely letter for keyboard layouts unfamiliar to users, showing correctly predicted keys increased speed by nearly 40%. However, they did not carry out their experiment on a QWERTY layout. Likewise, for typing using only eye gaze on a soft keyboard (as measured by an eye-tracker), MacKenzie and Zhang [22] found that next key prediction combined with their gaze fixation algorithm could significantly reduce error rates by nearly 35%. Perhaps the most conspicuous guidance signal was utilized by Al Faraj et al. [2] 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*". However, it is worth asking why any user would ever want to know "Where is my next key?" After all, users familiar with the QWERTY layout by definition know where the next key is and do not need to reduce their visual search space, as with non-QWERTY keyboards [24]. The answer is 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 with word prediction. Indeed, we make this link explicit in our visual design.

Figure 4 shows how we evolved the signal for key-prediction guidance. Note that as a design principle, we decided not to dynamically change the layout of the keys, as in BigKey [2]. In Figure 4(a), the next likely buttons were colored blue, whereas in Figure 4(b), only the actual letters on the buttons were colored. In our usability study, participants found Figure 4(a) to be "too distracting". On the other hand, those participants who used Figure 4(b) for guidance perceived and followed the blue coloring, whereas those who did not, stated that they were not bothered. With key-prediction guidance, it is important to limit the number of colored keys so that the entire keyboard does not turn blue. As such, we showed the blue visual cue only after the second letter of a word (which is also when word prediction candidates appear), and highlighted up to five letters at most. These letters had to correspond to word prediction candidates in the candidates area. As can be seen in Figures 4(a) and (b), these word prediction candidates also colored the predicted letter in blue. Note that the predicted 'e' corresponds to a candidate that is not shown and must be accessed by swiping left to right on the candidates area. For key-prediction guidance, we did not give any auditory signals.

In terms of implementation, similar to Magnien et al. [24], our method for generating predictions was to match prefixes with word entries in our 78K+ English dictionary. We also considered a character-based n-gram language model, similar to [7], but decided not to use it because the predicted keys did not always correspond to word prediction candidates. We felt this would engender user confusion, so we abandoned this method.

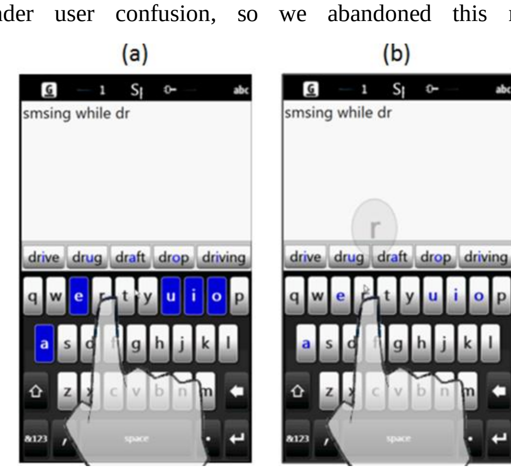


**Figure 4. Two designs for key-prediction guidance. After clicking a word, the next predicted letter is colored blue.**

Furthermore, we did not assume that previously typed letters were correct. As Masui [25] demonstrated with pen-based text entry, word disambiguation can leverage word prediction techniques. In our case, we implemented a matching algorithm that searched through combinations of the previously typed letters and their adjacent keys on the QWERTY layout and matched words in our dictionary based on word probabilities and edit distance [1]. Note that the guidance afforded by the signal constitutes the converse of that provided by the TypeRight keyboard [8], which exerts resistance on unlikely keys.

# 4. Usability Study

The multimodal signals we implemented all occur in the keyboard area as users interact with the keys. As such, when we developed and iteratively refined the visual design of the signals, we paid special attention to how they would combine and interact with each other. For deploying a commercial product, text entry performance is not enough reason to incorporate a multimodal signal. It has to work well with all other signals that are shipping. In fact, it has to work well with all other aspects of the keyboard to create an overall positive typing user experience. Our usability study was conducted with this practical imperative in mind. In this section, we discuss how usability participants reacted to each of the multimodal signals and to their combinations.

## 4.1 Method

4 female and 7 male participants from the local metropolitan area were recruited by a professional contracting service. 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. 3 of the participants owned touchscreen phones, 5 owned a QWERTY phone at some point in their lives, and 3 owned only 12-key numeric phones. The participants came from a wide variety of occupational backgrounds, from housewife to IT professional. The age range varied from 19–39, which approximates our target demographic. The average age was 25.3. The usability study lasted 1.5 hours and all participants were compensated for their time.

All participants were first taught the basics of using a touchscreen keyboard on a mobile device – in particular, a prototype before-market phone with a 3.5 inch (as opposed to 3.7 inch on the iPhone) resistive screen with 800x480 WVGA resolution. Participants were shown a short phrase from the well-known MacKenzie and Soukoreff's [21] phrase set on a desktop computer screen. The set contains 500 short English phrases with no punctuation, varying from 16–43 characters with a high letter frequency correlation with an English corpus (see [21] for more statistics). We supplemented the phrase set with news headline phrases culled from the Internet containing words not found in our dictionary (e.g., "*smsing* while driving is risky", "*obama* is inaugurated"). We randomly selected 4 phrases and 1 supplemental phrase for each of 16 conditions: 1 for each multimodal signal (4-choose-1), plus combinations thereof (4-choose-2, 4-choose-3, and 4-choose-4). We tested these combinations in order to inspect if any of the multimodal signals conflicted with each other, and to investigate how people felt about having a signal be absent in another condition. Because this was not an experiment, we did not counter-balance the order of the conditions.

Using the "think-aloud" protocol [18], participants were asked in each of the 16 conditions to type in the phrases "as quickly as possible" on the mobile device but that they should pause and verbalize any new thoughts they had about the multimodal signals at any moment. In entering text, participants were told to hold the device in any way that was comfortable but they were not allowed to place the device on the table for typing. As users proceeded to type, we also asked specific questions about each signal, and at the end of the study, we had participants rank-order any signals they would leave on by default.

## 4.2 Results

Because this was a usability study aimed at refining the user experience and design of the multimodal signals, and combinations thereof, we continually adjusted the visual and auditory parameters of some signals based on user feedback. As such, it is difficult to accurately interpret raw statistics. However, the numbers do convey general trends so we report them here.

### *4.2.1 Unexpected-key feedback*

7/11 participants listed that they would leave the unexpected-key feedback semaphore turned on by default, even when we used the earlier "distracting" versions shown in Figures 1(a-b). 5/11 listed it as their top choice stating that they "depended" on it. In particular, they noted how it would alert them to the candidates area where they would almost certainly find their desired word as a choice. Many participants in fact claimed that this was their "strategy" for typing as quickly and as accurately as possible. Interestingly, after we locked down the visual design, half of the participants said they did not perceive the visual signal and only relied on the auditory signal, and half said that they would turn off the auditory signal, as they considered the visual feedback informative enough.

### *4.2.2 Auto-correction feedback*

All 11/11 participants loved the signal. Some of the participants even stated that they "depended" on knowing when words were being auto-corrected. Furthermore, several participants noted that they could predict when certain words would be auto-corrected (e.g., "obama"), which helped them prepare to select their replaced word in the candidate area, as shown in Figure 2(c). Many participants asked to have the replaced word automatically added to the dictionary when selected from the candidate area. We indeed did have this feature, but left it out of the study to explore how users would leverage the auto-correction feedback.

### *4.2.3 Key-trail feedback*

Only 5/11 participants stated that they would leave this signal on by default with 2/5 claiming that they "depended" on it. In probing about what they meant by "depend", the participants stated that the trail or afterglow effect served to reassure them that they were typing the right keys so that they could continue to focus on hitting the small rendered keys. Informally, we observed that those 2 participants had very slow baseline typing speeds. For the other participants (3/5), the afterglow effect was just "cool eye candy" that enhanced their user experience. Of the 6 who did not choose to leave the signal on by default, 2 really disliked it (though this may be due to the fading rate, which we later adjusted).

### *4.2.4 Key-prediction guidance*

Only 5/11 participants stated that they would leave the signal on by default with 3/5 listing it as their top choice. Of the 5, only 1 claimed he "depended" on it, particularly for 1-handed thumb use. As we hypothesized, 4/5 participants who preferred the signal stated that they would leave it on "just in case [they] were unsure of how to spell a long word".

### 4.2.5 Combinations

Overall, we seemed to have succeeded in creating multimodal signals that could be combined effectively with each other. Initially, 2 participants disliked the combination of the key-prediction feedback with the key-trail feedback, but that was when the buttons were entirely colored, as in Figure 4(a). Once the blue was applied only to the letters, we did not receive any further complaints. According to the participants, no other combinations of the multimodal signals seemed bothersome, not even when all four of the signals were concurrently in use.

## 4.3 Discussion

Throughout the usability study, participants seemed to recognize the difficulties of typing on a small touchscreen keyboard and appreciated how the multimodal signals allowed them to continue focusing on the keyboard area, making comments to that effect. The only signal that received all positive responses was the auto-correction feedback, which is not surprising given the frustration participants felts having a correctly spelled word be mistakenly auto-corrected. Participants also felt that the auto-correction feedback worked well in combination with other signals.

The majority of participants also liked the unexpected-key feedback with many of them commenting on how it effectively made the candidates area more useful – that is, by alerting them to disambiguation candidates that corrected their text. The combination of the unexpected-key feedback with the auto-correction feedback worked well to direct users to the candidates areas, enhancing their overall typing experience.

With respect to the key-prediction feedback, although we expected some participants to utilize the signal for spelling purposes, we did not expect any users to find it useful for reducing the visual search space so as to "depend" on it. Indeed, for 2-handed typing, QWERTY users should be able to transfer their muscle memory and open-loop recall to the smaller form factor. However, we forgot that this is not necessarily the case for 1-handed thumb use. We revisit the issue of handedness in the user experiment.

Both the key-prediction feedback and the key-trail feedback had the same number of participants who would leave the signal on by default. However, based on the comments, participants seemed to think the key-trail feedback had less user value. Participants who were not the slowest baseline typists did not think it was all that useful to be informed that they are on the right track, even though that is part of what users do when they monitor their text. This may have to do with the bipolar key frequency distribution of the QWERTY layout, where keys and their next key tend to be on opposite sides of the keyboard (prompting many researchers to create optimized key layouts [23]). Indeed, if the next key is on the other side of the keyboard, users will find it more difficult to keep track of where they have been.

### 4.3.1 Implications

In attempting to reconcile contradictory comments for a multimodal signal, we observed that, in general, participants with slower baseline typing speed seemed to find the signal more useful than those accustomed to typing quickly on touchscreen keyboards. Furthermore, the slower typists seemed to appreciate visual highlighting more than the faster typists, who by and large preferred only an auditory signal. Given this observation and the diversity of user responses for all signals (except for the auto-correction feedback), it seems reasonable to recommend that in designing a touchscreen keyboard users should be given the option of turning signals on and off in a control panel. Because some users, such as fast typists, seemed to be more tuned to auditory feedback than visual, giving users fine-grained control over which modality is on in the control panel would be extremely valuable.

As an aside, we should point out that, as with all usability studies, some participants may have been influenced by the novelty of interacting with a new touchscreen device as well as the need to reciprocate for having received a gratuity. In general, for our study, participants did not seem to hesitate in pointing out negative aspects.

# 5. Experiment

From the usability study, we learned that the auto-correction feedback enhanced user experience and seemed to work well in combination with other multimodal signals. The next multimodal signal that garnered a majority of positive responses was the unexpected-key feedback. For this signal, usability participants stated that they thought it was beneficial for text entry and even developed strategies for using it in conjunction with the candidates area. However, we wanted to see if in fact it could improve text entry beyond perceived performance. Likewise, because the key-prediction guidance seemed to help some usability participants with the spelling of words and even 1-handed thumb input, we decided to put it to the test. Finally, we decided not to further explore the key-trail feedback given its lukewarm reception and seemingly low user value.

Overall, we conducted a controlled text entry experiment comparing:

1. Unexpected-key feedback combined with auto-correction feedback
2. Key-prediction guidance combined with auto-correction feedback
3. Auto-correction feedback alone as a baseline

For (1) and (2), we tested the combination of auto-correction feedback with either the unexpected-key feedback or the key-prediction guidance because we wanted to gauge how well the signals interacted with each other for timed text entry. In retrospect, we should have had no multimodal signals as a baseline because having (3) as a baseline could make it harder to find significant differences for (1) and (2). This means that any differences we do find would be impressive.

## 5.1 Methods

### 5.1.1 Participants

We recruited 18 participants (9 males and 9 females) between the ages of 21-39 using the same professional contracting service as before. The average age was 29.7. Participants again came from a wide variety of occupational backgrounds. All participants were compensated for their time. 5 owned touchscreen phones sometime in their life, 9 owned QWERTY phones sometime in their life, and 7 owned 12-key numeric phones only. In our statistical analysis, the type of phone participants owned did not turn out to be significant predictor. 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.

### *5.1.2 Stimuli*

We again utilized MacKenzie and Soukoreff's [21] phrase set except this time, in order to make sure the 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. We did not include the supplemental phrases with words not found in our dictionary as we did in the usability study so that we could examine how close our results came to other text entry research that only utilized the same MacKenzie and Soukoreff stimuli. In this case, the auto-correction feedback never occurred for words that were correctly spelled but mistakenly auto-corrected by the system. For each condition, subjects received 8 practice and 20 stimuli items.

### *5.1.3 Procedure*

All participants were first taught the basics of using the touchscreen keyboard on the same mobile device as mentioned previously in the usability study. Users 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. They had as much time as they needed to memorize the phrase. 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". We left the phrase on the computer screen because in previous experiments with the same task, some participants experienced difficulties with memorization under timed conditions. 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.

Because text entry experiments often exhibit a learning effect, we first introduced participants to each of the conditions using sample phrases. Users had as much time as they needed to familiarize themselves with the condition and the task. Once they felt "ready", we had them practice the timed task on 8 phrases. After the practice phase, we asked participants if they wanted to continue practicing. No participant needed extra practice.

### *5.1.4 Experimental Design*

Our primary independent variable was *SignalType* consisting of **Baseline**, Unexpected-key Feedback combined with Auto-correction Feedback (**UnexpectedKey**+), and Key-prediction Guidance combined with Auto-correction Feedback (**KeyPredict+**). Because the way in which a user holds the mobile device can affect typing performance, we had as another independent variable, *Handedness*, consisting of **OneHand** and **TwoHands**. For **OneHand**, participants held the device with 1 hand and used their finger on the other hand to enter text. For **Two Hands**, participants held the device with both hands and entered text using their thumbs. Instead of having participants try both **OneHand** and **TwoHands** with each of the different *SignalType* conditions within the 2 hour period, we instead had half of the participants (9/18) use **OneHand** and other half use **TwoHands**. Note that we carefully counter-balanced the order of the 3 *SignalType* conditions within each of the 2 *Handedness* conditions.

As our dependent variables, we examined time to enter text, accuracy, and the efficiency measure *keystrokes-per-character* (KSPC) [20]. As an additional measure, we also looked at 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, due to the fact that not all participants were proficient at this targeting task, pressing backspace was the only way users could correct text. Hence, the number of backspaces is a proxy for the number of corrections.

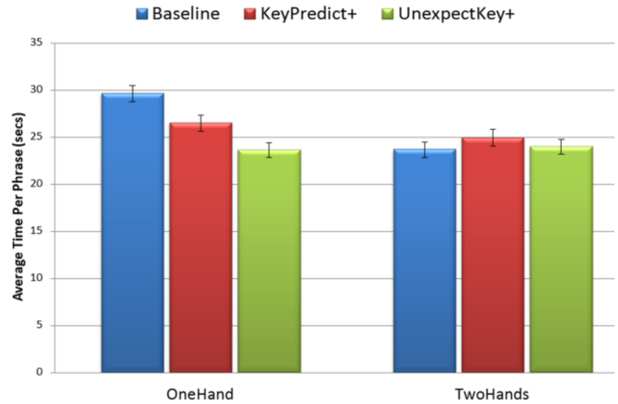


**Figure 5. Average time per phrase in seconds grouped by *Handedness* and then by *SignalType*. Error bars represent standard errors of the mean.**

Overall, we conducted a mixed factorial design experiment where *Handedness* (2 conditions) was a between-subjects factor and *SignalType* (3 conditions) was a within-subjects factor.

## 5.2 Results

### *5.2.1 Time to Enter Text*

For typing speed, we measured the average time participants spent entering in each phrase. Because usability participants mentioned how they developed a strategy whereby they would type quickly and use the unexpected-key feedback to alert them to candidates in the candidates area, we hypothesized that the **UnexpectedKey**+ condition would significantly reduce average typing time. Because key-prediction guidance improved typing speed by 25% in the BigKey soft keyboard [2], and our own usability participants found it to be useful for spelling purposes and 1-handed thumb use, we also hypothesized that **KeyPredict**+ would reduce average typing time. Using a mixed design ANOVA, we found a significant main effect for *SignalType* ($F_{2,712} = 8.57$, $p < .001$) as well as an interaction effect with *Handedness* ($F_{2,712} = 10.57$, $p < .001$). Figure 5 displays a chart breaking down average typing time by *Handedness* and *SignalType*. As shown in the chart, when participants used **TwoHands**, the signals all performed more or less the same. However, when participants used **OneHand**, the signals varied greatly in speed. Indeed, post-hoc comparisons revealed that **UnexpectedKey+** ($\mu_{secs} = 23.81$) took significantly less time than **KeyPredict**+ ($\mu_{secs} = 25.73$; $p < .01$) and less time than the **Baseline** ($\mu_{secs} = 26.65$; $p < .001$). The **KeyPredict**+ condition was not, however, significantly different than the **Baseline**.

### *5.2.2 Accuracy*

We measured accuracy in terms of whether or not the participant ultimately typed in the correct phrase, or simply *phrase accuracy*. We also considered finer-grained measures, such as the Minimal String Distance Error Rate [20], which computes the distance between two strings in terms of the lowest number of edit operations required to turn one string into the other. However, we found that we could not use these measures due to the fact that some of the errors made by users were those related to pressing a

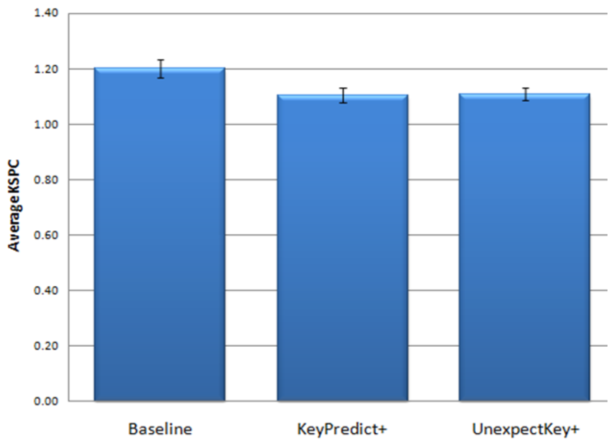


**Figure 6. Average keystrokes-per-character (KSPC) for *SignalType* conditions with standard errors of the mean.**

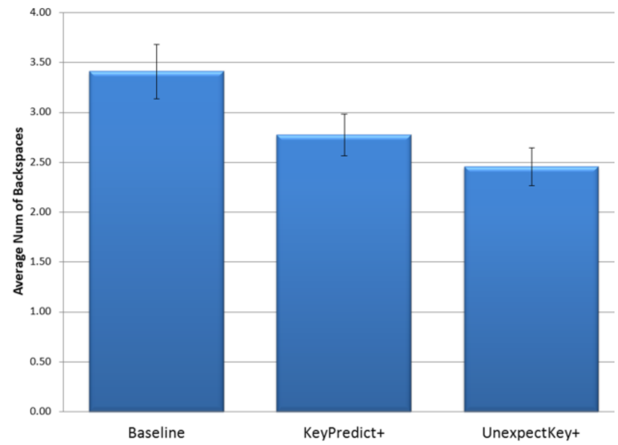


**Figure 7. Average number of backspaces for the *SignalType* conditions with standard errors of the mean.**

word prediction or disambiguation candidate, which swaps the typed text completely. In this case, one keystroke can engender an entire word error.

In terms of phrase accuracy, we hypothesized that both **UnexpectedKey+** and **KeyPredict**+ would exhibit higher accuracies than **Baseline** for the same reasons as before in discussing our hypotheses for time to enter text. Contrary to our hypothesis, we did not find a main effect for *SignalType*, nor did we find an interaction effect with *Handedness*. This may have been due to the fact that there were very few errors in general.

### *5.2.3 KSPC*

In terms of text efficiency, 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). We did not find an interaction effect with *Handedness*. As expected, post-hoc comparisons revealed that **UnexpectedKey+** (μ = 1.11) had significantly lower KSPC than the **Baseline** (μ = 1.20; $p$ < .01) and so did **KeyPredict**+ (μ = 1.10; $p$ < .01). However, the two were not statistically different from each other.

Figure 6 shows the average KSPC for the *SignalType* conditions. If we assume that users accurately hit only the key they are supposed to and not leverage word prediction candidates, the KSPC on a QWERTY layout should be 1 [20]. Note that all *SignalType* conditions were above 1, indicating that typing on a small touchscreen keyboard with fat fingers is challenging, even with additional feedback and guidance.

### *5.2.4 Number of Backspaces*

We measured the number of backspaces as a way to separate intentionally corrected mistakes from KSPC, which encompasses those as well as keystrokes for selecting word prediction candidates. 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, in post-hoc comparisons, the only significant difference was between **UnexpectedKey+** (μ = 2.46) and the **Baseline** (μ = 3.41; $p$ < .01). Figure 7 shows the average number of backspaces for the *SignalType* conditions. We did not find any interaction effect with *Handedness*.

## 5.3 Final Questionnaire

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. We did this to confirm our qualitative findings from the usability study. 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, it was still too disruptive for them. This again re-affirms the utility of having a control panel for the multimodal signals, as was discussed with the usability study.

## 5.4 Experiment Discussion

Despite the fact that the **Baseline** condition for *SignalType* included the Auto-correction feedback, which should make 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 these 3 dependent variables, **UnexpectedKey+** emerged as the best combination of signals. In summary, **UnexpectedKey+** reduced average time per phrase by 10.5%, 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 [17], 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. Another limitation is that our stimuli, as mentioned previously, did not contain any words that were not in our typing intelligence dictionary. This may have boosted **UnexpectedKey+** and **KeyPredict**+ because they operate best on known words. In practice, this limitation is not likely to diminish our findings because our dictionary is already fairly comprehensive at 78K+ words. Furthermore, once users commit to any text (e.g., by undoing replaced words with the auto-correction feedback) we can add those words to the dictionary.

Having utilized the same stimuli as Hoggan et al. [9], we compared at a gross level the performance of **UnexpectedKey+** to that of using a mobile device with tactile feedback added to the touchscreen keyboard. In terms of average time per phrase, **UnexpectedKey+** was at about 24 seconds whereas the tactile keyboard was at about 20 seconds in a laboratory setting (as reported in [9]). In terms of KSPC, **UnexpectedKey+** was at about 1.1 whereas the tactile keyboard was at about 1.3 in a laboratory setting. Because the experiments utilized different hardware, software, experimental procedures and probably different stimuli (given that neither experiment used all of the MacKenzie and Soukoreff 500 phrases [21]), it is impossible to draw any conclusions. A head-to-head comparison needs to be conducted on the same device. We thought it might be interesting though to see if our numbers were in the same range as previous research. The point we make here is that if in fact the numbers are comparable, that would mean that using a combination of visual and auditory signals can be as good as using tactile signals, but at a much lower cost (since no new hardware is involved). A promising avenue for future research is to explore how all the modalities can be combined together to make touchscreen keyboards much more usable, and hopefully closer in performance to that of physical keyboards.

In terms of other directions for future research, although we found statistically significant difference in a laboratory setting, we did not have users attempt to interact with multimodal signals in a mobile setting – i.e., as users were walking or riding a train (e.g., as in [9]). This could provide valuable in-situ data which could help refine our design. Another direction is 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 as users generally become faster and more accurate using a mobile touchscreen keyboard.

## 6. Design Implications

From having conducted a usability study to first identify combinations of visual and auditory signals that would provide useful feedback and guidance in the keyboard area, and then a user experiment to investigate whether these combinations could actually improve text performance, we collected several design implications which we plan to take advantage of ourselves for deploying a touchscreen keyboard.

First, we learned that creating multimodal signals in the keyboard area that convey the same kind of information users would discern in monitoring their typed output can not only enhance user experience but also improve text performance. There are probably other multimodal signals we have not yet considered that also fulfill the same objective.

Second, participants differed in terms of their preference for visual versus auditory signals. As such, visual signals should be subtle, so that those who want to use it can exploit it, and those who do not, can ignore it. We strove to design such subtle visual signals during the usability experiment. However, this leaves open a discoverability problem: if the signals are too subtle, first-time users may not discover and effectively utilize them.

Third, providing a control panel from which users can turn on and off, and even adjust modal parameters would greatly benefit users. We say this for many reasons. We found that users not only express different preferences for modalities, but the different modalities themselves seem to benefit different users depending on their baseline typing speed and whether they are using 1 versus 2 hands. Slow typists (including 1-handed typists) seemed to greatly appreciate the visual signals whereas the fast typists did not. Furthermore, different situations often call for one modality over another. For example, too much ambient noise can render an audio signal useless, so having a control panel from which users can toggle modalities provides tremendous utility. As an aside, it might be interesting to see how users would respond to a system that automatically detects the general typing speed of the user as well as whether they are using 1 or 2 hands, and to adjust the different types of multimodal signals it employs.

## 7. 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 four 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?”, key-trail feedback answers “Am I still on the right track?”, and key-prediction guidance answers “Where is my next key?” The first three signals provide feedback about events users would normally discern when they are monitoring the text area. The fourth signal provides guidance about how to avoid typing mistakes. We also discussed the results of a usability study we conducted to assess and refine the user experience of these signals, and combinations thereof. Because our primary goal is to deploy a commercial product with exceptional user experience, even if it is just perceived user experience, we could have stopped at the usability study and deployed just the auto-correction feedback (which received nothing but positive responses). However, we decided to conduct a user experiment because increased text performance can also enhance user experience. So, 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 improved typing speed by 11%, reduced keystrokes-per-character by 8%, and reduced backspaces by 28%. Finally, we summarized everything we learned about designing multimodal signals with our discussion of design implications.

Besides all the directions for future research that we discussed in the paper, we have not yet looked at whether these multimodal signals and combination thereof could be useful for non-QWERTY keyboard layouts as well as on larger touchscreen devices such as tablets tabletops. We hope that the research we presented here will encourage others to pursue these directions.

## 8. REFERENCES


[1] Anonymous.

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

[3] Baudisch, P., Tan, D., Collomb, M., Robbins, D., Hinckley, K., Agrawala, M., Zhao, S., & Ramos, G. 2006. Phosphor: explaining transitions in the user interface using afterglow effects. In *Proc. UIST*, 169-178.

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

[5] Brown, L. & Brewster, S. 2008. Multidimensional tactons for non-visual information display in mobile devices. In *Proc. of MobileHCI*, 231 - 238.

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

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

[8] Hoffmann, A., Spelmezan, D. & Borchers, J. 2009. TypeRight: A keyboard with tactile error prevention. In *Proc. of CHI*, 2265-2268.

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

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

[11] http://windows.microsoft.com/en-US/windows7/Type-without-using-the-keyboard-On-Screen-Keyboard

[12] Kaaresoja, T., Brown, L. & Linjama, J. 2006. Snap-Crackle-Pop: Tactile feedback for mobile touch screens. In *Proc. of Eurohaptics*, 565-566.

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

[14] Kristensson, P. 2007. Discrete and continuous shape writing for text entry and control. Ph.D. thesis. Linköping University. Sweden.

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

[16] Lee, J., Dietz, P., Leigh, D., Yerazunis, W. & Hudson, S. 2004. Haptic Pen: A tactile feedback stylus for touch screens. In *Proc. of UIST*, 291-294.

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

[18] Lewis, C., & Rieman, J. 1993. *Task-Centered User Interface Design: A Practical Introduction*. Distributed via anonymous ftp (ftp.cs.colorado.edu).

[19] Luk, J., Pasquero, J., Little, S. MacLean, K., Levesque, V. & Hayward, V. 2006. A role for haptics in mobile interaction: Initial design using a handheld tactile display prototype. In *Proc. of CHI*, 171-180.

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

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

[22] MacKenzie, I.S., & Zhang, X. (2008). Eye typing using word and letter prediction and a fixation algorithm. In *Proc. of the ACM Symposium on Eye Tracking Research and Applications – ETRA 2008*, 55-58.

[23] MacKenzie, I.S., Zhang, X., & Soukoreff, R. 2003. Text entry using soft keyboards. *Behaviour and Information Technology*, 18(3): 235-244.

[24] Magnien, L., Bouraoui, J. & Vigouroux, N. 2004. Mobile text input with soft keyboards: optimization by means of visual clues. In *Proc. of MobileHCI*, 337-341.

[25] 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.

[26] Norman, D. A. 1988. *The Design of Everyday Things*. Basic Books.

[27] Potter, R.L., Weldon, L.J., & Shneiderman, B. 1988. Improving the accuracy of touch screens: An experimental evaluation of three strategies. *Proc. of CHI*, 27-32.

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

[29] Shannon, C.E. 1951. Prediction and entropy of printed English. *The Bell System Technical Journal*, 30.