Touch Behavior with Different Postures on Soft Smartphone Keyboards

Shiri Azenkot
Computer Science & Engineering | DUB Group
University of Washington
Seattle, WA 98195 USA
shiri@cs.washington.edu

Shumin Zhai
Google Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
zhai@google.com

ABSTRACT

Text entry on smartphones is far slower and more error-prone than on traditional desktop keyboards, despite sophisticated detection and auto-correct algorithms. To strengthen the empirical and modeling foundation of smartphone text input improvements, we explore touch behavior on soft QWERTY keyboards when used with two thumbs, an index finger, and one thumb. We collected text entry data from 32 participants in a lab study and describe touch accuracy and precision for different keys. We found that distinct patterns exist for input among the three hand postures, suggesting that keyboards should adapt to different postures. We also discovered that participants' touch precision was relatively high given typical key dimensions, but there were pronounced and consistent touch offsets that can be leveraged by keyboard algorithms to correct errors. We identify patterns in our empirical findings and discuss implications for design and improvements of soft keyboards.

Author Keywords

Mobile text entry, touch interfaces.

ACM Classification Keywords

H.5.2. [Information interfaces and presentation]: User interfaces—input devices and strategies.

INTRODUCTION

Text entry on small devices remains slower and more error-prone than on traditional computer keyboards [10]. In recent years, the soft QWERTY keyboard (Figure 1) has emerged as a standard smartphone text entry method, offered as the default keyboard on millions of Android phones, iPhones, and Windows phones. These keyboards are commonly designed for finger (as opposed to stylus) input, and have sophisticated algorithms to mitigate user error.

Understanding touch behavior is critical for improving keyboard algorithms. While prior research on touch keyboards has made progress in understanding users' speed and aggregate error rate [2,3,18,19,21], our work focuses on *where* users touch when targeting specific keys. Patterns in touch locations can help automatically prevent and correct errors. For example, if we find that users commonly hit below target keys, a keyboard algorithm can compensate by adding the typical pixel offset to the input touch's screen coordinates, preventing errors caused by this tendency. We identify such patterns with the goal of providing a foundation for improving soft keyboard algorithms.

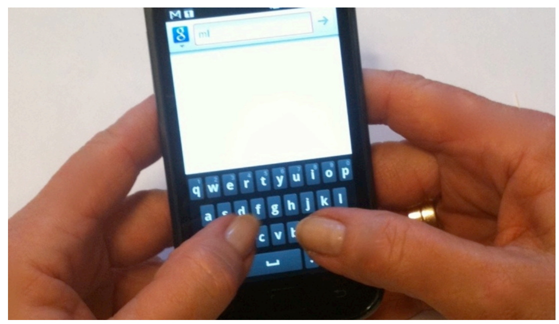

Figure 1. A user prepares to enter text on a smartphone.

Conceivably, users' touch patterns may vary according to hand posture (i.e., how they hold the mobile phone and type on it). We observed that people have different preferences in hand posture and may change hand posture depending on the use situation. When holding a cup of coffee, for example, one may enter text with a single thumb since only one hand is available. In other situations, one may use an index finger or two thumbs to type. We therefore compare three common hand postures in our analysis, showing distinct trends among them. To our knowledge, this is the first systematic investigation of touch patterns in relation to these three postures.

In this work, we analyze text entry data we collected from 32 participants who entered text using two thumbs, an index finger, and a thumb. We report the traditional text entry measures of words per minute (WPM) and error rates, but focus our discussion on touch patterns particularly in terms of offsets and spreads for different keys. An offset is the distance between the user's touch point and the target key center. The spread of touch points refers to their distribution for a specific key. Large spreads and offsets both increase the chance to touch outside key boundaries, causing errors by the conventional definition. It is important to understand what the natural tendencies of these parameters are when

MobileHCI'12, September 21–24, 2012, San Francisco, CA, USA.

the user is unencumbered by current error correction algorithms and how these parameters change with different hand postures.

Our contributions include (1) the analysis of touch behavior for text entry with three common postures, and (2) a discussion of how future keyboards can be improved based on the patterns we identify. We hope this work will serve as a foundation for improving touch keyboards.

RELATED WORK

The current work is situated in the growing body of literature of touch screen interfaces and text input. We can only briefly highlight three basic categories of work most related to the current study: investigations of touch performance in general, efforts to model and understand text entry performance, and improvements in key detection and auto-correction algorithms.

Understanding Touch Performance

Several papers, reviewed in this section, have explored human touch performance on capacitive touch screens, but not in the context of text entry. Continuous text entry has higher cognitive and motor demands than pointing tasks with a single target, so text entry behavior may be different.

Lee and Zhai [16] investigated soft button user performance in comparison to hard buttons and as a function of implement (finger vs. stylus), feedback (auditory and tactile-vibrato), and button size.

Henze et al. [8] analyzed a wealth of touch data collected from a smartphone game that was deployed in the field. Game players hit targets across the screen and Henze et al. identifyed patterns in their touch offsets. We draw inspiration from their analysis and visualizations, and compare their results to ours. While we collected a smaller amount of data through lab studies rather than field deployment, we had more control and were able to compare differences among hand postures.

Like Henze et al., Holz and Baudisch [11] also found that users touched a surface with consistent offsets from a target. Their subsequent studies show that a user's perceived contact point was approximated by the center of the fingernail, which was above the actual contact point along the finger's axis [12].

Wang and Ren [23] compared touch offsets among all five fingers, reporting that the thumb and pinky finger were less accurate than the others.

Understanding Text Entry Performance

Several papers aim to better understand touch performance during text entry. Findlater et al. [4] presented an analysis of 10-finger typing on large touch surfaces that is parallel to our work on smartphones. Other papers have focused on modeling time rather than understanding causes of errors. MacKenzie and Zhang [19] present a model based on Fitts' Law for predicting expert speed on pen-based soft keyboards. In subsequent work, MacKenzie and Soukoreff [18] presented a model for two-thumb entry which is empirically validated and adjusted by Clarkson et al. [2]. These models predict the words per minute (WPM) of expert typists and do not consider error patterns as we do.

Improving Key Detection and Auto-Correct Algorithms

Our work provides a foundation for improving key detection and auto-correct algorithms, so we also consider work that aims to improve these techniques.

Soft keyboards can in effect dynamically change the underlying key size according to context in order to correct touch error. Goodman et al. [6] used character probabilities along with observed touch point distributions for keys to correct user error. They made several general observations about touch point distributions. Gunawardana et al. [7] proposed a less aggressive approach to key-target resizing that does not prevent users from typing their desired text. In later work, key-target resizing algorithms were personalized through the Text Text Revolution game [21], which collected users' touch data to train the key-target resizing algorithm.

Kristensson and Zhai [13] used geometric pattern matching to do word level error correction. Both Goodman et al. [6] and Kristensson and Zhai [13] focused on stylus input on early generations of touch screen products.

Other work identified errors on physical mini-QWERTY keyboards using timing [3], or on number pad keyboards with nine soft buttons [22]. Gong et al. [5] aimed to improve auto-correct systems by rearranging the order of corrected word candidates that were displayed to the user.

A relatively new error tolerant method of text entry is the gesture keyboard that first appeared in the academic literature in 2003 [14,27,28]. We limit our investigation to point touch only, although previous work has shown gesture keyboards and touch keyboards can be linked in their underlying algorithms [13].

USE OF DIFFERENT HAND POSTURES IN TEXT ENTRY

To base our touch behavior study on empirical facts with regard to hand postures, we conducted a survey to determine how often people used two thumbs, one thumb, and one finger to enter text on touch screen phones.

Method

Seventy-five participants completed our survey, all of whom had touch screen phones. There were 58 (77%) males and 17 (23%) females, with an average age of 32 (*SD* = 9). Participants were Google employees and were recruited via an internal mailing list.

The survey included three questions about the frequency in which participants used two thumbs, one thumb, or one finger to enter text. An example of this type of question was:

> How often do you enter text on your phone with two thumbs?

- Almost always
- Often
- Sometimes
- Almost never

Participants were able to select one response per question.

Results

Participants reported relatively high proficiency levels with smartphone text entry. The mean level of proficiency was 5.05 (*SD* = 1.27), on a 7-point scale where 1 indicated the user had no experience and 7 indicated the user was an expert at entering text on soft smartphones keyboards.

As shown in Figure 2, each of the three postures was used at least "sometimes" by about 60% of the participants. On the other hand, no single method was used by all participants. Two thumbs, one thumb, and one finger were "almost never" used by 37%, 35% and 41% of the participants, respectively.

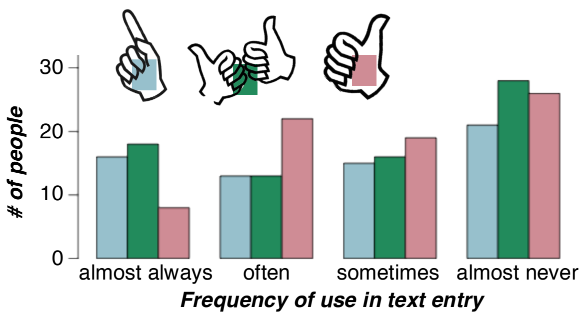

Figure 2. Use of an index finger, two thumbs, and one thumb during text entry on smartphones.

A deeper exploration of the contexts in which each posture is preferred would provide further insight into our survey results. Yet these preliminary findings do suggest that touch behavior with all three postures should be considered when designing improved soft keyboards.

COLLECTING TOUCH POINT DATA

We conducted a study to collect the touch point data that we explore in this work. Our goal was to collect and analyze data that reflected fundamental human behavior, with minimal bias towards a particular keyboard design.

Participants

We recruited 32 participants (15 females and 17 males) with an average age of 27 (*SD* = 6). Participants were required to have experience with text entry on a touch screen phone. Only two participants were left-handed and the rest were right-handed. All participants were Google employees and were recruited through an internal mailing list. The same recruitment process was used for this study and the survey that was described in the previous section.

Apparatus

Participants completed the study by typing phrases on a custom Android application, which we call the Keyboard Touch Collector (KTC). This application displayed a series of phrases and a minimal keyboard as shown in Figure 3. We designed the KTC to achieve two goals:

1. To capture fundamental user behavior on soft keyboards that is unaffected by the design (and potential shortcomings) of current keyboards.
2. To align touch points with target letters. We needed to reliably label touch points with the key the user targeted. If we did not know which key the user intended to hit, a touch point provided little information about the user's performance.

The KTC keyboard dimensions were the same as those of the default Android keyboard. We chose a minimal design to prevent more complex visual elements from affecting the user's behavior either positively or negatively. In other words, we wanted to collect *baseline* touch patterns and performance measures.

Figure 3. The Keyboard Touch Collector application used for collecting touch data in our study.

The application logged the *x* and *y* pixel coordinates of each keyboard touch along with the aligned character in the presented phrase. Touch points were logged on "touch down" events.

When a user touched the KTC keyboard, an asterisk appeared on the screen. We chose to provide limited feedback because displaying the input character would involve using a particular character detection algorithm that may affect the user's behavior. Even a naïve algorithm, where the selected key is the one that visually bounds the touch (without key-target resizing), may cause the user to slow down her input rate to prevent errors. Another naïve algorithm, where the "right" character is displayed regardless of the touch input, may give the user a false sense of confidence and cause unusually erroneous input. Thus, we do not inform the user whether her touch was “correct” to observe natural performance that is independent of keyboard algorithms.

The KTC did not have a BACKSPACE key and users were unable to correct errors. Correcting errors takes time: the user must enter the BACKSPACE key in addition to reentering the deleted character. An ideal keyboard, therefore, would tolerate imprecision and automatically correct the user’s errors instead of requiring her to manually correct them. We wanted to capture the natural variance and biases in touch behavior without the assistance or limitations of the current generation of soft keyboards. We thus decided that analyzing human performance without error correction is essential for guiding improved keyboard design.

Users were asked to input short and shortened phrases that were randomly selected from the phrase set designed by MacKenzie and Soukoreff (M&S) [17,24]. There are a number of studies on phrase set selection for text input testing, e.g., [15,20]. Some of these studies question whether the M&S phrase set represents mobile text input with new technologies that use language models in predictive text input. However, at this time there is still no universally agreed standard set that considers all the factors involved in text input testing, such as memorability, mutual-information between test and application, and fun factor for the participants. For the purpose of the current study that focuses on touch patterns around individual characters, the M&S phrase set seemed appropriate since it has a good correlation with common English text at a letter frequency level [17].

We used a Samsung Galaxy S phone running Android 3.2 for all sessions, with a screen size of 60 by 100 mm with a resolution of 480 by 800 pixels. The keyboard dimensions were 60 by 37.5 mm.

Study Procedure

Each participant completed one 30 minute session where he or she answered a short questionnaire and then entered phrases on the KTC. The first ten phrases served as a warm-up and participants were able to take short breaks after each set of 10 phrases. We instructed them to type “as accurately and as naturally” as possible.

Participants completed the study while seated in an office environment. We instructed each participant to enter text with “two thumbs,” “one index finger,” or “one thumb.” For the latter two postures, we asked participants to choose one hand and use it throughout the study. No further specification was given in order to capture participants’ natural hand posture inclinations.

Experiment Design and Analysis

Our experiment was designed with the between-subjects factor Posture with three levels: Two-thumbs, One-thumb, and One-finger. Postures were randomly assigned to each participant and balanced, such that 11 participants entered text with two thumbs, 11 entered text with one thumb, and 10 entered text with one finger. While a 32 participants study is not large for a between-subject design, it is on par with or larger than most perceptual-motor skill tests.

Statistical results are reported with a significance level of $\alpha = 0.05$.

Amount of Data Collected

We collected a total of 86,888 labeled touch points, not including warm-up phrases; 35,124 labeled touch points were entered with two thumbs, 27,628 were entered with one thumb, and 24,126 were entered with one finger. On average, each participant entered 117 phrases.

Labeling Touch Points

Reliably labeling touch points was important in most of our analysis, yet our data included some points that seemed to be mislabeled, lying far from their labeled keys. Such points were probably a result of omission, insertion, or transposition of characters. We removed points that were more than 1.5 times the key height away from their labeled key center (see Figure 4), resulting in the removal of 3.19% of the data.

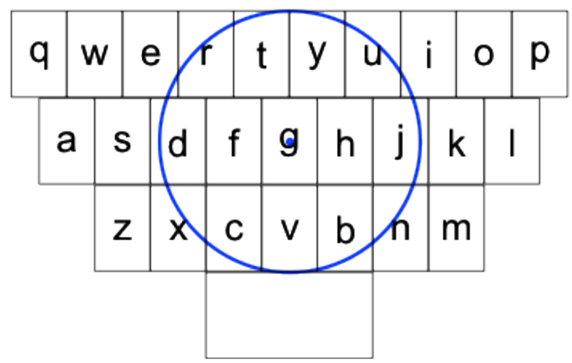

Figure 4. The blue circle marks the region that defines outliers for the “g” key. All points outside the blue circle were not included in parts of our analysis that required labeled points.

WORDS PER MINUTE

Our goal in this section was to compare the text entry speeds for the three hand postures. The KTC offered a baseline upper bound for smartphone keyboards because entry rates were unhindered by error correction. Such an upper bound baseline simulates the performance of a “perfect” smart touch keyboard whose algorithms would correct all of a user’s errors automatically, enabling her to achieve unhindered entry rates. To our knowledge, our results are the first empirical estimates of upper bound smartphone touch tapping performance.

The average entry speed for all participants was 41.01 WPM (*SD* = 13.43), similar to the higher speeds achieved by Rudchenko et al. [21]. When calculating WPM, every 5 characters (including the SPACE character) were counted as one word. As seen in Figure 5, entry with two thumbs was fastest, at 50.03 WPM (*SD* = 13.57). One-finger input rates were far lower, averaging only 36.34 WPM (*SD* = 11.28).

Entry rates with one thumb were 33.78 WPM (*SD* = 6.87), slightly lower than with one finger.

We found that the differences in speed among postures were statistically significant. WPM measures for phrases were not normally distributed (Kolmogorov-Smirnov D = 0.111, $p < 0.0001$), so we ran a Kruskal Wallis test that revealed a significant effect of Posture on WPM ($x^2_{(2)}$ = 1101.4; $p < 0.001$). A post-hoc test using Mann-Whitney tests with Bonferroni correction showed that there were significant differences between all pairs of postures.

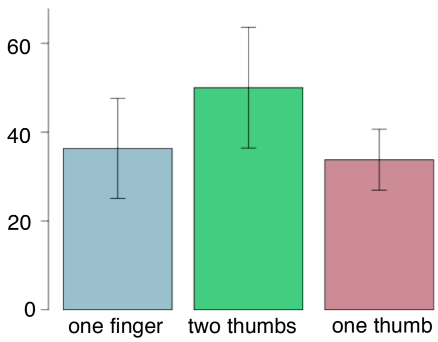

Figure 5. Words per minute vs. hand posture. Entering text with two thumbs was significantly faster than entering text with one finger or one thumb.

We expected two-thumb entry to be faster than the other postures. Since each thumb only moves within one half of the keyboard, less time is required to move from one key to the next. Furthermore, a user can move one thumb towards its next target while the other thumb is active, parallelizing movement from key to key. Frequent left and right alternation, a property of the QWERTY layout, was originally designed to minimize mechanical jamming but later became a major advantage for two-handed typing [26]. Such a property still gives a time advantage to two-thumb typing as is shown in our results.

It is surprising that one-thumb entry was only slightly slower than one-finger entry. The index finger seems far more dexterous than the thumb, and one-thumb entry was not as commonly used as one-finger entry according to our survey results. Entering text with one thumb offers the advantage of being one-handed, since using an index finger requires one to hold the phone with the other hand. It may also have the advantage of being more stable since the thumb movement is relative to the same hand that holds the phone.

ERROR RATE

Along with words per minute, the mean phrase error rate is traditionally reported in text entry studies. Our goal in this section was to compute baseline error rates for the three postures that may help keyboard designers assess keyboard improvements that aim to reduce the overall error rate.

For this purpose we use a narrow definition for error that follows from the hard key analogy, where a touch that lies outside of the visual boundary of its target key is considered an error. Reasonable key detection or auto-correct algorithms would use probabilistic inference to decide which key a user targeted given a touch point and its context (*e.g.*, a language model, a user's past behavior) and the visual boundaries would not necessarily matter. Our error rates thus provide an upper bound that can serve as a point of comparison for better keyboard algorithms.

The error rates among postures reflected a speed-accuracy trade-off, as seen in Figure 6. Overall, the mean error rate (number of errors in a phrase divided by the total phrase length) was 8.87% (*SD* = 10.51). The highest error rate was obtained with the two-thumb posture at 10.80% (*SD* = 12.46). The error rate for one-finger input was lower, at 8.17% (*SD* = 9.57), and the error rate for one-thumb input was the lowest, at 7.00% (*SD* = 7.84).

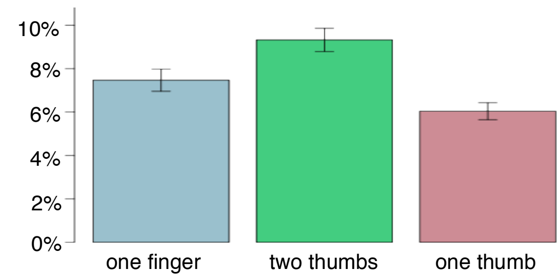

Figure 6. Error rate vs. hand posture. The lines and whiskers mark the 95% confidence intervals for the mean error rates.

Differences between postures were mostly statistically significant. As is typical, the distributions of error rates per phrase were not normal (Kolmogorov-Smirnov D = 0.200, $p < 0.0001$), so we ran a Kruskal-Wallis test that showed there was a significant effect of Posture on Error Rate ($x^2_{(2)}$ = 65.1, $p < 0.001$). A post-hoc test using Mann Whitney tests with Bonferroni correction showed that there were significant differences between error rates with two thumbs and one thumb ($p < 0.001$) and between input with two thumbs and one finger ($p < 0.001$). There was no significant difference between error rates with one-thumb and one-finger input.

We expected the error rate for one-thumb input to be higher than for one-finger input. Prior work found a thumb to be less accurate in touch screen pointing [23], but it seemed digit width did not yield more errors in our context of soft keyboarding. Note that in this context the one-thumb condition differed from the one-finger condition in two aspects: different touching digit and different holding hand. With one thumb keyboarding, the same hand holds the device and activates the key presses.

In the following sections, we explore the distributions of touches for keys, which provide far more information about the kinds of errors that users make than traditional text entry measures.

TOUCH POINT DISTRIBUTIONS

To better understand touch performance, we explore touch point patterns for keys that are input using the three hand postures. Previously, we presented measures that were computed for phrases entered by the user. We now take a different approach, where we group input touches by target key rather than phrase, examining touch locations relative to their target key areas.

As expected, the distribution of touch points per key was roughly a bivariate Gaussian, confirming previous findings [6,23]. The keyboards on the left in Figure 7, Figure 8, and Figure 9 show the collected touch points for each of the three postures over an outline of the keys (see Figure 4 for key labels). There is overlap between keys, however, so many of the points are occluded. Also, some keys have substantially more data points than others, so it is difficult to detect trends in these plots.

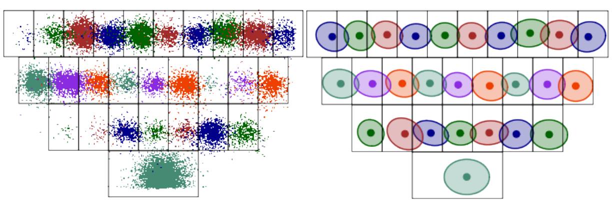

Figure 7. Touch points (left) and 95% confidence ellipses (right) for text entered with one finger.

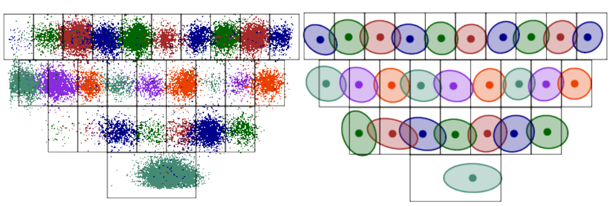

Figure 8. Touch points (left) and 95% confidence ellipses (right) for text entered with two thumbs.

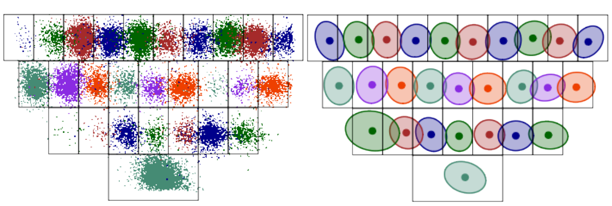

Figure 9. Touch points (left) and 95% confidence ellipses (right) for text entered with one thumb.

For a more informative visualization, we plotted the 95% confidence ellipses for the points of each key, shown on the right of Figure 7, Figure 8, and Figure 9. The centers of the ellipses mark the mean of the touches for a key, showing how accurate they were. The closer the ellipse center is to its key center, the higher the accuracy for that key. The size of the ellipses reflect the collective spread of the points, showing how precise they were. The smaller the ellipse, the more precise the touch points were for the key.

In addition to accuracy and precision, the ellipses also show the shape of the spreads. While the keys are tall rectangles, the touch spreads are ellipses that are often wider than they are tall. Some ellipses are stretched along the diagonal. The orientation of the major axis of each ellipse corresponds to the covariance, or correlation, between the x and y coordinates of the points. Early work on stylus keyboards also found that covariance varies among keys [6], but no further discussion was provided.

We believe the covariance is determined by the context of touches for the key: the direction whence the digit came and possibly also the direction in which the digit is headed. Plotting the touches of particular bigrams or trigrams may be useful for understanding how the shapes of the ellipses form. We defer this analysis to future work.

VERTICAL TOUCH POINT OFFSETS

In this section, we explore the vertical offsets of touch points for keys in different postures. The vertical offset of a point is the difference between the y coordinate of a touch point and the y coordinate of the key center. The mean offsets indicate whether there is a tendency to hit below (a positive offset) or above (a negative offset) the key.

There is an overall trend that touches tended to land below key centers. Figure 10 shows the vertical offset for each key in every posture. There are only a handful of keys with positive vertical offsets, all of which are negligible. The patterns and magnitudes of vertical offsets varied among postures. The mean vertical offset for keys in the one-finger posture is largest, at 4.69 pixels (SD = 2.92), the mean for keys in the two-thumb posture is 2.16 pixels (SD = 2.06), and the mean for keys in the one-thumb posture is 3.25 pixels (SD = 2.37). Note that the height of each alphabetical key is 75 pixels.

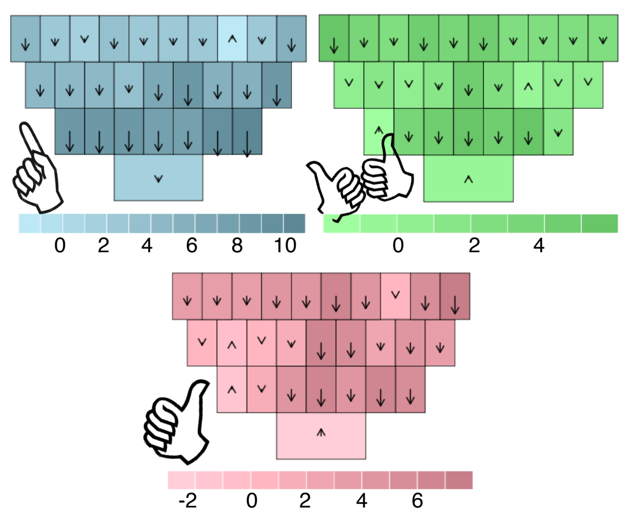

Figure 10. Vertical offsets for touch points in all three postures. The color gradient of each key indicates the value of the offset. The arrows show relative offset sizes among keys, but are not drawn in the scale of the keyboard outline.

Our findings are generally in agreement with findings in the context of touch pointing at single targets. Holz and Baudisch [12] show that the actual contact point lies below the user's perceived contact point for a touch along the fingertip. However, Henze et al. [8] found that targets that lie roughly in the bottom 10% of the screen induced touch points with high negative vertical offsets (i.e., the contact points were above the targets). Henze et al. collected hit data for targets that appeared across the entire screen, however, and our participants hit targets that were only within the bottom portion of the screen. On average, it seems the touch point data from Henze et al.'s experiments had a positive vertical offset, similar to our findings, despite the different use context (keyboarding vs. single target pointing).

As the top left keyboard in Figure 10 shows, one-finger touch data exhibits a striking trend across the key rows. Offsets of keys nearly double from one row to the next. The mean vertical offset for keys in the top row is 2.13 pixels (*SD* = 1.80); for the second row from the top the mean vertical offset is 4.89 pixels (*SD* = 1.87), and for the third row, it is 8.10 pixels (*SD* = 1.20), which was more than 10% of the key height.

It seems that the thumb is not as likely to hit below the target as the index finger, as the two-thumb and one-thumb postures do not exhibit such strong trends in vertical offsets. Keys in the center of the keyboard have slightly higher vertical offsets for two-thumb entry while keys on the right side of the one-thumb keyboard tend to have slightly higher vertical offsets than keys on the left side.

HORIZONTAL TOUCH POINT OFFSETS

We explore horizontal offsets, or the difference in the *x* coordinates of touch points and their target key centers, in this section. A positive offset indicates that the touch lies to the right of the key center. Horizontal offsets revealed more surprising trends than vertical offsets.

Horizontal offsets were more pronounced than vertical offsets. The horizontal locations of the ellipses in Figure 7, Figure 8, and Figure 9 clearly show that the touch points are not centered within their target keys along the *x* axis. In some cases, such as the "q" key in the one-thumb posture, the center of touch points is almost at the right key boundary. Figure 11 shows the mean horizontal offsets for keys in all three postures.

The SPACE key in all three postures has large positive horizontal offsets. We discuss this in a later section and focus on trends among the alphabetical keys in this section.

For all three postures, offsets were greater in magnitude on the left side of the keyboard than on the right. For two-thumb entry, it seems that participants were more accurate with their right thumbs. Recall that we only had two left-handed participants, so participants appear to be more accurate with their dominant hand. For one-finger and one-thumb entry, it seems participants were less accurate when reaching for keys across the keyboard.

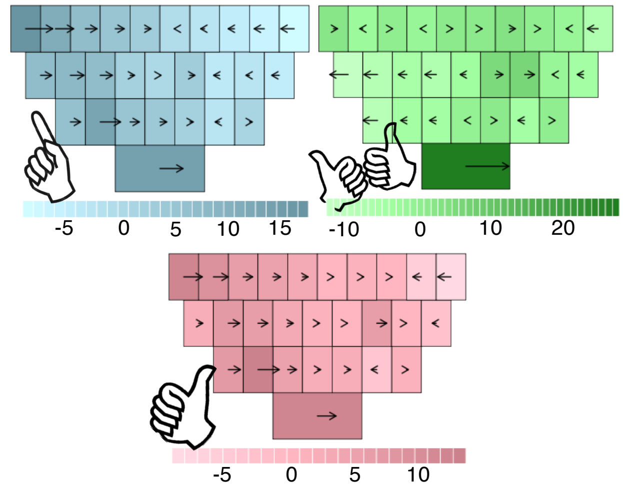

Figure 11. Horizontal offsets for touch points in all three postures. As in Figure 10, the color gradient of each key reflects the value of the offset. The arrows show relative offset sizes among keys, but are not drawn to the scale of the keyboard outline.

The trends in horizontal offsets on the left side of the keyboard were strikingly different among postures. Keys in this region had negative offsets with two-thumb entry and positive offsets with one-thumb and one-finger entry. This suggests participants try to minimize the distance their thumb or finger must travel and in many cases, they did not reach far enough across the keyboard.

As with vertical offsets in our data, horizontal offset patterns for the single-digit postures support prior work [6,8]. Henze et al. [8] also found (1) larger horizontal offsets on the left side of the screen (although their study involved discrete target pointing) and (2) smaller negative horizontal offsets on the right side of the screen. While Henze et al. conducted unsupervised field deployments where hand posture was not known, we conjecture that many participants used one finger or one thumb to play their game.

TOUCH POINT SPREADS

We observe the sizes of touch spreads, comparing the precision of user hits for different keys. To measure the spread size, we compute the standard deviation of touch points, which is the expected distance of an observed touch from the mean.

Figure 12 shows average spreads for the collective data and for the individual participants (the spread of points per key per participant). A large collective spread can emerge in two ways. Participants may have large spreads of touch points, showing high imprecision when targeting keys. Alternatively, participants may be fairly precise, but have largely differing offsets which result in large spreads of the collective data. Both cases are important to explore. Individual trends have implications on the design of adaptive or personalized keyboards, while spreads in the collective data have implications for keyboards that are designed to support all users.

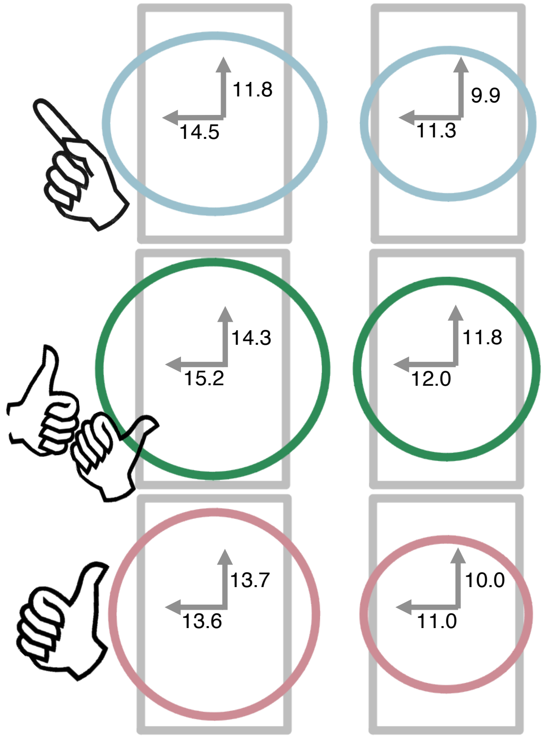

Figure 12. Collective (left) and average participant (right) touch spreads, marked by 95% confidence ellipses relative to a key outline. The ellipses are drawn to scale. The standard deviation (SD) is reported in pixels for each dimension, indicated by an arrow.

For two-thumb and one-finger entry, touch point spreads for keys are larger in the x dimension than in the y dimension, even though keys are much taller than they are wide. The differences between dimensions are more noticeable in the one-finger posture, as seen in Figure 12. Current keyboard dimensions are probably a result of the key height being less constrained than key width, which is limited to 1/10 of the total screen width. Our data suggests that the key height may be reduced if it is desirable to have the keyboard occupy a smaller portion of the screen. Further research is needed to ensure there is no coupling effect between the vertical and horizontal dimensions. More constrained vertical dimension may lead to overall more cautious user behavior.

Data collected during one-thumb entry showed a distinct pattern in spread height across the keyboard. The spreads for keys in the top left of the keyboard were taller than they are wide, unlike the spreads in the other two postures. Surprisingly, we did not observe these trends within the points of individuals; it seems the large overall spreads were a result of varying offsets among participants, possibly due to different thumb sizes. For example, the standard deviation in the y dimension of all one-thumb entry points for the 'q' key was 14.26 pixels. The mean standard deviation for the points of individual participants, however, was just 8.88 pixels ($SD = 4.66$).

Individual participants were more precise than the collective data suggested, with spreads that were only slightly larger than the key widths. Small keys thus do not appear to be a major cause of error; different offsets among participants produced large overall spreads and seem to be a more compelling source of errors.

THE SPACE KEY

We explore touch points for the SPACE key separately, since this key is different than the alphabetical keys in form and function. While all keys have the same height, the width of the SPACE key is typically at least three times the width of an alphabetical key on mainstream smartphone keyboards, as shown in Figure 13. The SPACE key is functionally different because it delimits words and triggers auto-correct modifications.

Figure 13. The iPhone's keyboard (top-left) SPACE key spans five alphabetical keys, the Windows Phone 7 keyboard (top-right) SPACE key spans four alphabetical keys, and the Android keyboard (bottom) SPACE key spans three keys.

As seen in Figure 7, Figure 8, and Figure 9, the spreads of touch points for the SPACE key mostly lay to the right of the key center. Horizontal offsets were large, especially for the two-thumb posture. It seems that participants mostly hit the SPACE key with their right thumb. Findlater et al. [10] also found that participants often hit the space key with their right thumbs when typing with 10 fingers. Our one left-handed participant who entered text with two thumbs in the study tended to hit to the right of the SPACE key center as well. It would be interesting to gather data from other left-handed people to see if they used their right thumbs.

In addition to being shifted, the spreads of points on the SPACE key were small compared with the width of the key. The spread in the two-thumb data were largest among the postures, and yet the 95% confidence ellipse only covered two-thirds of he key's width. This suggests that the space key may not need to be as wide (but perhaps moved to the right), yet it is even wider on other common keyboards.

DISCUSSION AND IMPLICATIONS FOR DESIGN

Our empirical findings show distinct patterns among three common hand postures used to enter text on mobile phones.

This suggests that researchers and practitioners should consider hand posture when designing and evaluating keyboards. We argue that keyboards should therefore detect, automatically or with user input, which posture is being used to support a large number of people in different mobile computing contexts. We distill our findings and discuss their implications for design, following the idea that hand posture should be considered.

Touch points tend to land below key centers. There should be a small (several pixels) vertical correction subtracted from each touch point.

One-finger input has the highest vertical offsets, which increases along the top three rows of the keyboard. Vertical corrections should be larger in one-finger input, and should depend on the key row.

Touch points for one-finger and one-thumb input tend to hit towards the center of the right half of the keyboard. Horizontal corrections should be added to touch points accordingly, effectively shifting the centers of the keys in the left side of the keyboard to the right.

Touch points on the left side of the keyboard are largely offset to the left for two-thumb entry. The affected key centers should be shifted to the left to accommodate. Also, automatic corrections in general should be more aggressive for keys in this region and posture because their high negative offsets cause many points to land outside of key boundaries.

Spreads of points are wider than they are long for one-finger and two-thumb input. Keys may not need to be as tall as those we tested, since the vertical space is not utilized.

Spreads of points are relatively small among participants, but collective spreads are larger, and they vary by key and posture. Participants were relatively precise, but their offsets differed. This suggests that adaptable keyboards that adjust for key offsets according to the individual (personalization) and the hand posture (adaptation) would perform well.

The spreads of points on the space key were shifted to the right and did not span the full width of the key. The space key can be made smaller and moved to the right of the bottom row.

These implications for design should be viewed as suggestions for future experiments. It is likely that any of the changes we recommend will affect user behavior. For example, if a vertical correction was added for each key, participants may adapt to this change and produce even greater vertical offsets in their touches. Since participants received minimal feedback from the Keyboard Touch Collector in our study, we believe our data does show raw, intrinsic, and in some ways optimal, performance for a keyboard in the given layout. More importantly, the value of the touch patterns we presented is not limited to the outlined design implications. Designers may draw their own conclusions according to their design goals and constraints from our well controlled empirical findings.

LIMITATIONS AND FUTURE WORK

Our analysis has several limitations, which we view as opportunities for future work.

First, we plan to determine the effect of visual design elements on touch behavior. We collected data on a basic keyboard which served as a baseline. It would be interesting to compare our data to data collected on keyboards with varying key gutters, color gradients, label sizes, etc. We tested and collected our data on one device size. A larger or smaller physical size of the keyboard may impact the touch patterns significantly.

Second, we hope to collect additional data and explore the speed-accuracy trade-off. By observing touch data for specific bigrams, we can control the distance moved and compare accuracy for different movement speeds.

Third, it would be interesting to analyze error across the dimension of time and order. We have focused on touch location, i.e., imprecision and inaccuracy, as a cause of error and our criteria (the same number of taps as characters in the target phrase) eliminated most other types of errors. However, transposition of inputs, as found in prior work on physical keyboards [3], and other timing patterns may account for and help us correct errors as well.

Fourth, in our study we instructed participants to enter data "as accurately and as naturally" as possible. This could have encouraged them to be unusually cautious. It is an open question if participant behavior would significantly change if the goal was to type as fast as possible.

Finally, we have only studied touch tapping as a way of entering text on soft keyboards. Understanding the fundamental motor control behavior and patterns in gesture keyboards [13,14,28] is another future research direction.

CONCLUSION

We have explored touch performance on soft QWERTY keyboards that are used by millions of people on their smartphones. We found that an index finger, two thumbs, and one thumb are all common hand postures used for entering text, and they induce different touch patterns and overall performance. Our empirical findings and implications for design may serve as a foundation for keyboard researchers, designers, and developers to improve keyboard layouts and algorithms. Understanding touch performance is an important step in advancing soft keyboards on small screens so they can ultimately out-perform traditional hard keyboards, continuing to expand the potential of mobile computing.

ACKNOWLEDGMENTS

This research was conducted at Google Research where the first author was a graduate intern. We thank Zhiwei Guan, Pernilla Qvarfordt, and Jacob O. Wobbrock for sharing their experimental research and data analysis experience with us.

REFERENCES

1. Bao, P., Pierce, P., Whittaker, S., and Zhai, S. (2011). Smart phone use by non-mobile business users. *Proc. MobileHCI '11*, 445-454.
2. Clarkson, E., Lyons, K., Clawson, J., and Starner, T. (2007). Revisiting and validating a model of two-thumb text entry. *Proc. CHI '07*, 163-166.
3. Clawson, J., Lyons, K., Rudnick, A., Iannucci, Jr., R., and Starner, T. (2008). Automatic whiteout++: correcting mini-QWERTY typing errors using keypress timing. *Proc. CHI '08*, 573-582.
4. Findlater, L., Wobbrock, J. O., and Wigdor, D. (2011). Typing on flat glass: examining ten-finger expert typing patterns on touch surfaces. *Proc. CHI '11*, 2453-2462.
5. Gong, J., Tarasewich, P., and MacKenzie, I. S. (2008). Improved word list ordering for text entry on ambiguous keypads. *Proc. NordiCHI '08*, 152-161.
6. Goodman, J., Venolia, G., Steury, K., Parker, C. 2002. Language modeling for soft keyboards. *Proc. IUI '02,* 194-195
7. Gunawardana, A., Paek, T., and Meek, C. (2010). Usability guided key-target resizing for soft keyboards. *Proc. IUI '10*, 111-118.
8. Henze, N., Rukzio, E., and Boll, S. (2011). 100,000,000 taps: analysis and improvement of touch performance in the large. *Proc. MobileHCI '11*, 133-142.
9. Himberg, J., Häkkilä, J., Kangas, P., and Mäntyjärvi, J. (2003). On-line personalization of a touch screen based keyboard. *Proc. IUI '03*, 77-84.
10. Hoggan, E., Brewster, S., & Johnston, J. (2008). Investigating the effectiveness of tactile feedback for mobile touchscreens. *Proc. of CHI '08,* 1573-1582.
11. Holz, C. and Baudisch, P. (2010). The Generalized Perceived Input Point Model and How to Double Touch Accuracy by Extracting Fingerprints. *Proc. CHI '10*, 581-590.
12. Holz, C. and Baudisch, P. (2011). Understanding Touch. *Proc. CHI '11*, 2501-2510.
13. Kristensson, P-O., Zhai, S. (2005). Relaxing Stylus Typing Precision by Geometric Pattern Matching. *Proc. IUI '05*, 151-158.
14. Kristensson, P.O. and Zhai, S. (2004). SHARK2: A Large Vocabulary Shorthand Writing System for Pen-based Computers. *Proc. UIST '04*, 43-52.
15. Kristensson, P. O., and Vertanen, K. (2012). Performance comparisons of phrase sets and presentation styles for text entry evaluations. *Proc. IUI '12*, 29-32.
16. Lee, S., Zhai, S. (2009). The performance of touch screen soft buttons. *Proc. CHI '09*, 309-318.
17. MacKenzie, I. S. and Soukoreff, R. W. (2003). Phrase Sets for Evaluating Text Entry Techniques. *Proc. EACHI '03*. 754–755
18. MacKenzie, I. S. and Soukoreff, R. W. (2002). A Model of Two–Thumb Text Entry. *Proc. GI '02*, 117–124
19. MacKenzie, I. S., Zhang, S. X. (1999) The design and evaluation of a high-performance soft keyboard. *Proc. CHI '99*, 25-31.
20. Paek, T., and Hsu, B.-J. P. (2011). Sampling representative phrase sets for text entry experiments: a procedure and public resource. *Proc. CHI '11*, 2477-2480.
21. Rudchenko, D., Paek, T., and Badger, E. (2011). Text Text Revolution: A game that improves text entry on mobile touchscreen keyboards. *Proc. Pervasive '11*, 206-213
22. Kumiko Tanaka-ishii. (2007). Word-based predictive text entry using adaptive language models. Nat. Lang. Eng. 13, 1 51-74.
23. Wang, F. and Ren, X. (2009). Empirical evaluation for finger input properties in multi-touch interaction. *Proc. CHI '09*, 1063-1072.
24. Wobbrock, J. O., and Myers, B.A. (2006). Analyzing the input stream for character-level errors in unconstrained text entry evaluations. *ACM TOCHI* 13 (4), pp. 458-489.
25. Wobbrock, J. O., Myers, B. A., and Aung, H. H. (2008) The performance of hand postures in front- and back-of-device interaction for mobile computing. Int. J. Hum.-Comput. Stud. 66, 12, 857-875.
26. Yamada, H. (1980). A historical study of typewriters and typing methods: from the position of planning Japanese parallels. Journal of Information Processing, 1980. 2(4): 175-20.
27. Zhai, S. and Kristensson, P.-O. (2003). Shorthand Writing on Stylus Keyboard. *Proc. CHI '03,* 97-104.
28. Zhai, S. and Kristensson, P.O. (2006). Introduction to Shape Writing. IBM Research Report RJ10393 (also as Chapter 7 of I. S. MacKenzie and K. Tanaka-Ishii (eds), Text Entry Systems: Mobility, Accessibility, Universality, Morgan Kaufmann Publishers, 139-158.