EFFECTS OF KEY SIZE AND SPACING ON THE COMPLETION TIME AND ACCURACY OF INPUT TASKS ON SOFT KEYPADS USING TRACKBALL AND TOUCH INPUT

Martin Schedlbauer
Department of Computer Science
University of Massachusetts, Lowell, MA

Simulated keyboards on touch screens are becoming the norm for data entry on mobile and kiosk systems. Since onscreen keyboards compete with other user interface elements for limited screen space, it is essential that soft keyboard designs are optimally laid out. This paper describes an experiment in which the performance and accuracy of data input on soft keyboards with square key of two different widths (10 and 15*mm*) and two inter-key gap distances (1.5 and 4.5*mm*) were evaluated. Three methods of input were studied: finger, stylus, and trackball. Entry times were the shortest and most accurate for stylus touch, although trackball input was the most accurate for the smallest key size. The spacing between keys did not exhibit a significant effect regardless of key size and input method. A key size of 15*mm* appears to be sufficiently large to provide acceptable accuracy for touch input, although a key size of 10*mm* was equally acceptable for trackball input.

INTRODUCTION

Touch screen displays have become an established input method for mobile and kiosk devices. Examples of such devices include self-serve terminals, navigation systems, and personal digital assistants (PDA). Among the most common interactions, besides command selection, is the input of alphanumeric strings, such as account numbers, GPS coordinates, names, and addresses. Touch screen displays offer the possibility of displaying keyboards directly on the screen. Unlike physical keyboards, *soft keyboards* displayed on touch screens have the advantage that designers can manipulate the size, spacing, and location of individual keys. However, due to the limited screen size and resolution of touch displays, designers must optimize the geometry of soft keyboards.

Several studies have been conducted that systematically varied key size and edge-to-edge spacing (*gap*) between keys (Bender, 1999; Colle & Hiszem, 2004; Scott & Conzola, 1997; Sears & Zha, 2003). Bender found that 30*mm* square keys had fewer errors and it took less time to enter a four digit sequence compared to 10*mm* square keys. The more recent study by Colle & Hiszem determined that the gap between keys did not have a significant effect on performance and accuracy and that task performance improved as the key size increased. In their experiments, only finger touch input was evaluated. Neither their experiment, nor the experiments by Bender and Scott & Conzola, compared the relative performance of finger touch versus stylus and trackball selection. Sears & Zha found that the error rate for stylus input on very small soft keyboards, such as those on PDAs, was invariant to key (or button) size.

The study described in this paper augments the work of Colle & Hiszem by evaluating and comparing finger touch to stylus and trackball input. A numeric keypad design was chosen because the entry of numeric strings has practical significance. For example, navigation systems require the input of latitude and longitude coordinates which are principally numeric strings.

METHODS

The experiment by Colle & Hiszem used four variations of key sizes (10, 15, 20, and 25*mm*) and gap sizes of 1 and 3*mm*. They did not discover a significant difference in terms of performance and accuracy between the larger sizes (20 and 25*mm*) and the 15*mm* size, so the present study was restricted to two sizes (10 and 15*mm*). Similar to Colle & Hiszem's method, the visual area of each key was equal to the tactual recognition field and a land-on tapping strategy was employed. Scott & Conzola found that a lift-off response criterion was more accurate for smaller target sizes, but most touch screens report input to the system as soon as the probe lands on the target and not once the probe lifts off.

Gap was measured as the edge-to-edge spacing between adjacent keys.

Subjects

Eight participants (7 male, 1 female) were recruited via e-mail and flyers from the researcher's department with the only restriction that participants must be right-handed because the test apparatus was not reconfigurable for left-handed use. The age of the participants ranged from 22 to 57 years. All participants had extensive experience using computers and had normal or corrected-to-normal vision and reported no other physical impairments. The participants received compensation in the form of a gift certificate.

Apparatus

The experiment was carried out on a Gateway M275 PC (1.6GHz CPU, 512MB RAM) running Windows XP and Java 5. A 15” (305×230*mm*) LCD touch screen manufactured by Elo Touch Systems (Model ETL151C-8SWA) using surface-acoustic technology was attached to the Gateway M275 and was used as the primary display. For the stylus trials, a special Elo stylus was utilized. A Kensington Turboball (Model 64227) was connected to the computer via USB. Figure 1 shows the experimental setup in which the touch screen was mounted on a fixed surface tilted at an angle of 55° to allow for use while standing up. The trials were presented using a research workbench developed by the author (Schedlbauer *et al.*, 2005).[1]

Figure 1. Experimental apparatus in which the targets were displayed on the touch screen to the right. The left and top screens were unused in this experiment.

Procedure

Each participant was presented with 4 blocks of 8 different strings each containing 9 characters. The strings were all in the format “D XXX YY ZZZ” where D was either ‘N’ or ‘S’ and X, Y, and Z were digits simulating part of a GPS coordinate. The white space was only shown to improve readability. It was not part of the string and was not entered during the experiments. Each trial block displayed a soft keypad with a fixed key size and a fixed gap (see Figure 2).

Timing started as soon as the keypad was displayed so that initial reaction and choice discrimination time could be recorded. Auditory feedback confirmed successful entry of a character. Errors were also signaled audibly, but with a different sound. Any tap outside the target region was recorded but otherwise ignored.

To reduce the effects of learning, participants were given several warm-up trials.

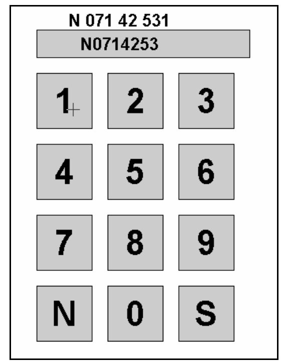

Figure 2. A soft keypad on which a string is being entered (not rendered to scale). The string that must be entered is displayed above the keypad and the characters are echoed as they are selected. The cursor is shown over key ‘1’. The key size is 15*mm* wide with a gap of 4.5*mm*.

Experimental Design

The experiment varied key size and gap (spacing between the edges of the keys), and input method as independent within-subject variables. The dependent variables were task completion time (*TT*) and error rate (*ER*). Two differently sized square keys were tested (*width* = 10 and 15*mm*) at two gap distances (1.5 and 4.5*mm*). All 4 blocks were repeated with finger, stylus, and trackball input while standing up.

The strings shown in Table 1 were chosen since they maximize the movement distance between the keys while varying the string enough from trial to trial so as to not cause a significant learning effect. The presentation order of the strings was randomized.

Table 1. List of strings entered by subjects.

N 173 01 937	N 167 34 029	S 016 27 394	N 037 29 160
N 060 19 730	S 173 49 102	N 037 19 052	S 130 37 915

RESULTS

Q-Q plots and Shapiro-Wilk tests of the observed values for *TT* showed strong normality for both input methods. Therefore, parametric tests were used in the analysis. No outliers were removed from any of the collected data sets.

[1] The research workbench is available at `http://www.cs.uml.edu/~mschedlb/mte`.

To attenuate the influence of outlying values and arrive at mean performance parameters, the measured task completion times for each condition (key width and gap) were averaged across all subjects resulting in 32 samples (2 key widths × 2 gap × 8 strings).

Task Completion Time Analysis

The mean task completion time for finger touch across all eight numeric strings was 5920*ms* (*sd* = 561) resulting in an average time of 658*ms* per digit. The average value for *TT* for *width* = 10*mm* was 6287*ms*, whereas for *width* = 15*mm* it was 5552*ms*, a statistically significant decrease of 11.9% [$F(1,30) = 23.857$, $p < 0.001$]. Gap was not found to have a significant effect on mean task completion time [$F(1,30) = 0.852$, $p = 0.363$].

For stylus input, the mean task completion time across all eight numeric strings was 5614*ms* (*sd* = 623) resulting in an average time of 624*ms* per digit. A paired-samples *t*-test showed that the decrease of 5.2% compared to finger touch was significant ($t = 3.94$, $p < 0.001$). The average value of *TT* for *width* = 10*mm* was 5829*ms*, whereas for *width* = 15*mm* it was 5399*ms*, a marginally significant decrease of 7.4% [$F(1,30) = 4.169$, $p = 0.05$]. As with finger touch, gap size was not found to have a significant effect on mean task completion time [$F(1,30)=2.589$, $p = 0.118$].

For trackball input, the mean task completion time across all eight numeric strings was 9231*ms* (*sd* = 551) resulting in an average time of 1026*ms* per digit. Paired-samples *t*-tests showed that the increases of 55.9% compared to finger touch and 64.4% compared to stylus were significant ($t = 39.94$, $p < 0.0001$ and $t = 35.55$, $p < 0.0001$). The average value of *TT* for *width* = 10*mm* was 9631*ms*, whereas for *width* = 15*mm* it was 8830*ms*, a decrease of 8.3% that is significant [$F(1,30) = 36.143$, $p < 0.0001$]. As with the touch modes of input, gap size did not have a significant effect on mean task completion time [$F(1,30)=1.407$, $p = 0.245$].

Figure 3 summarizes the task completion times for each target size by method of input.

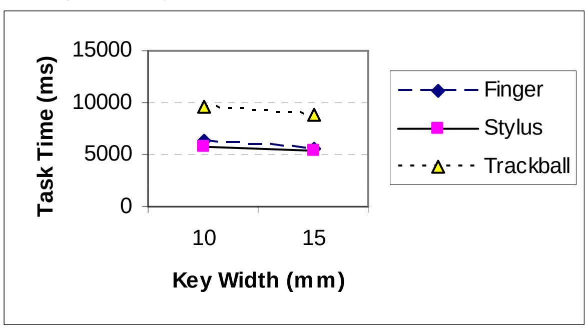

Figure 3. Task completion time for each input method by key width.

Accuracy Analysis

The accuracy analysis used in this paper is analogous to the metric applied by Colle & Hiszem. For each target size, the number of missed hits was recorded and then averaged over the total number of hits.

The mean error rate using finger input was 3.5% (*sd* = 0.218). The keypad with a key width of 10*mm* had a significantly higher error rate of 5.1% compared to an error rate of 1.9% for the keypad with a key width of 15*mm* ($t = 3.533$, $p < 0.001$, Bonferroni corrected). Although gap size did not by itself exhibit a significant interaction with error rate ($p = 0.118$), there was a significant two-way interaction with width and gap as factors ($p = 0.0119$). A one-way ANOVA with the input string as a factor did not reveal any significant differences in error rates between the eight numeric strings.

The mean error rate for stylus was 2.4% (*sd* = 0.194). The increase in accuracy of stylus touch compared to finger touch was found to be significant via a paired-samples *t*-test ($t = 1.936$, $p = 0.05$). Once again, the keypad with a key width of 10*mm* had a significantly higher error rate of 3.4% compared to an error rate of 1.5% for the keypad with the larger keys ($t = 2.365$, $p < 0.05$). Like finger touch, the gap size did not by itself exhibit a significant interaction with error rate ($p = 0.054$). Unlike finger touch, no significant two-way interaction between error rate and width and gap as factors was observed [$p = 0.0852$], suggesting that stylus input is less sensitive to the configuration of the keypad. A one-way ANOVA with the input string as a factor did not reveal any significant differences in error rates between the eight numeric strings.

The mean error rate for trackball was 2.2% (*sd* = 0.167). A paired-samples *t*-test showed the differences between finger touch and trackball input to be significant ($t = 2.343$, $p = 0.019$), although the difference between trackball input and stylus touch was not significant ($t = 0.486$, $p = 0.626$). Interestingly, the differences in the error rates based on key width were not significant ($t = 0.496$, $p = 0.620$). Generally, the error rates were comparable to those of stylus touch. Furthermore, a one-way ANOVA with width and gap as a factors did not show a significant interaction ($p = 0.457$). Like stylus touch, the trackball was not sensitive to key widths and gaps. Once again, a one-way ANOVA with the input string as a factor did not reveal any significant differences in error rates between the eight numeric strings.

Figure 4 provides a graph of the mean error rates for each key width by method of input.

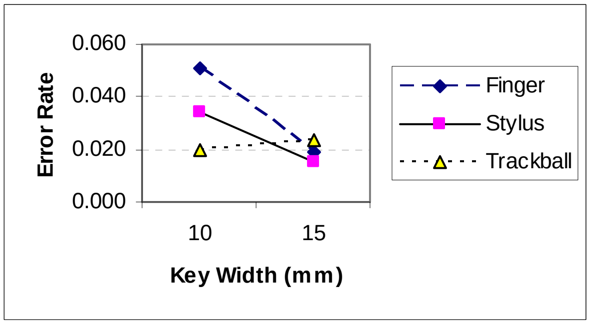

Figure 4. Error rate for each input method by key width.

An alternative view of accuracy is the dispersion of the selection end points. The dispersion is calculated as the mean standard deviation of the selection end points for each target. Figure 5 (on last page) shows the selection end points for each key width-gap combination and input method. The mean standard deviations for each key size and input method are presented in Table 2. Interestingly, as target size gets larger, spatial variability increases as well, indicating that subjects work faster but with less precision. The differences in the spatial variability between the small and large key sizes are significant for every input method as shown by *t*-tests using Bonferroni corrections ($p < 0.01$). The differences between the input methods are significant only for the larger key size of 15*mm* ($p < 0.001$, Bonferroni corrected). Overall, stylus selection had the least variability.

Table 2. Mean standard deviation of selection end points by key width and input method.

	Mean SD		
Key Width	*Finger*	*Stylus*	*Trackball*
10*mm*	8.93	8.29	8.64
15mm	10.33	9.28	11.75

DISCUSSION

Colle and Hiszem report a much higher mean error rate of about 18% for 10*mm* keys and about 16% for 15*mm* keys, but their experiments were conducted using four-wire capacitive touch technology versus the more accurate and more responsive surface acoustic technology employed in this experiment. In addition, their keypads did not provide audible feedback when an incorrect key was pressed. One can conclude that auditory feedback and touch technology are important design elements in soft keypad implementations. Furthermore, their experiments only evaluated finger touch.

Stylus input is more accurate for smaller targets compared to finger input, making that method of input more appropriate for devices where screen size is limited. Overall, stylus was the most efficient and accurate input method, although trackball was most accurate for the smaller key size. Because the error rate for trackball is invariant to key size and (unlike stylus and finger touch input) does not exhibit a reduced error rate for the larger target size, it is a more useful input device for applications where screen size is small and an indirect input device is usable, such as maritime or in-vehicle navigation systems.

CONCLUSIONS

Overall, the most effective input method in terms of speed was stylus touch, while the trackball was the most accurate. Gap size did not have any significant effect, and therefore keys can be spaced closer together when screen real estate is at a premium. For all input methods, the error rate decreased significantly as the target size increased. Extrapolating from these observations, trackball input should be used when screen size is at a premium and accuracy is paramount. If completion time is most critical, stylus based touch is most desirable. The increased error rate can be ameliorated by increasing target size above an extrapolated width of 12*mm* and providing appropriate error correction mechanisms as well as auditory and visual feedback.

REFERENCES

Bender, G. (1999). *Touch screen performance as a function of the duration of auditory feedback and target size*. PhD. Dissertation, Wichita State University, Kansas, USA.

Colle, H. A., and Hiszem, K. J. (2004). Standing at a kiosk: Effects of key size and spacing on touch screen numeric keypad performance and user preference. *Ergonomics*, *47*(13), 1406-1423.

Schedlbauer, M., Pastel, R., and Heines, J. (2005). An extensible and interactive research platform for exploring Fitts' law. *Technical Report 2005-014, Department of Computer Science*, University of Massachusetts Lowell. Available online at http://teaching.cs.uml.edu/~heines/techrpts/details.jsp?Year=2005&SeqNo=015.

Scott, B., and Conzola, V. (1997). Designing touch screen numeric keypads: effects of finger size, key size, and key spacing. In *Proceedings of the Human Factors and Ergonomics Society 41th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 360-364.

Sears, A. and Zha, Y. (2003). Data entry for mobile devices using soft keyboards: understanding the effects of keyboard size and user tasks. *International Journal of Human-Computer Interaction*, 16(2), 163-184.

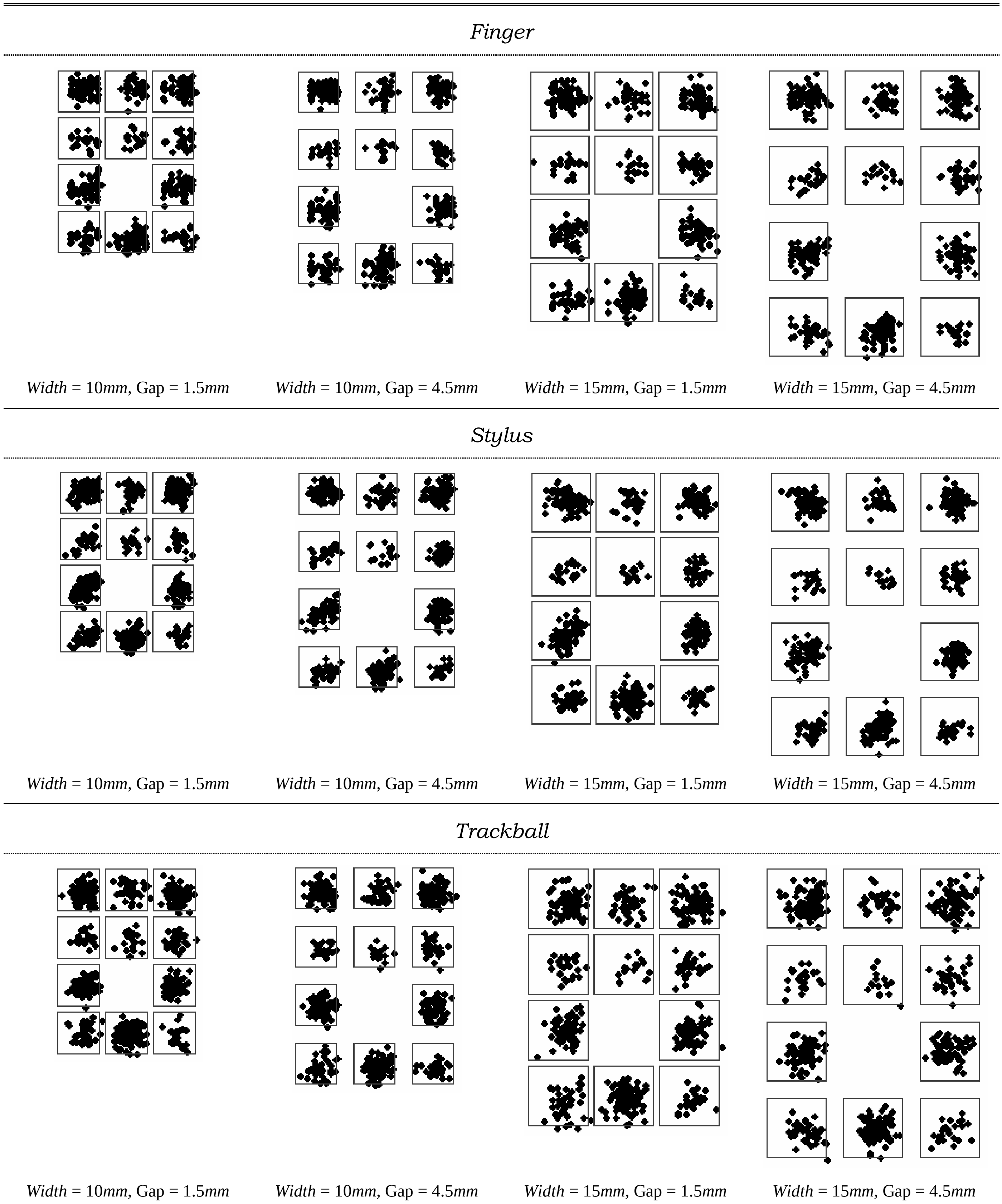

Figure 5. Actual selection end points for each key width-gap combination arranged by input method. Visually, the spatial dispersion of the end points is more pronounced for trackball compared to finger and stylus input. Selections outside the target areas (errors) are not shown.