Ergonomics

ISSN: 0014-0139 (Print) 1366-5847 (Online) Journal homepage: www.tandfonline.com/journals/terg20

Standing at a kiosk: Effects of key size and spacing on touch screen numeric keypad performance and user preference

Herbert A Colle & Keith J Hiszem

To cite this article: Herbert A Colle & Keith J Hiszem (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, DOI: 10.1080/00140130410001724228

To link to this article: https://doi.org/10.1080/00140130410001724228

Published online: 20 Feb 2007.

ERGONOMICS, 22 OCTOBER, 2004, VOL. 47, NO. 13, 1406–1423

Taylor & Francis
healthsciences

Standing at a kiosk: Effects of key size and spacing on touch screen numeric keypad performance and user preference

HERBERT A. COLLE* and KEITH J. HISZEM

Department of Psychology, Wright State University, Dayton, Ohio 45435, USA

Keywords: Touch screens; Keypads; Key size; Key spacing; Kiosks; Keyboards; Human-computer interaction.

Touch screen input keys compete with other information for limited screen space. The present study estimated the smallest key size that would not degrade performance or user satisfaction. Twenty participants used finger touches to enter one, four or 10 digits in a numeric keypad displayed on a capacitive touch screen, while standing in front of a touch screen kiosk. Key size (10, 15, 20, 25 mm square) and edge-to-edge key spacing (1, 3 mm) were factorially combined. Performance was evaluated with response time and errors, and user preferences were obtained. Spacing had no measurable effects. Entry times were longer and errors were higher for smaller key sizes, but no significant differences were found between key sizes of 20 and 25 mm. Participants also preferred 20 mm keys to smaller keys, and they were indifferent between 20 and 25 mm keys. Therefore, a key size of 20 mm was found to be sufficiently large for land-on key entry.

1. Introduction

In the last 10 years, touch screen displays have become an established interface design option, not just an experimental test bed. Touch screen kiosks are being used increasingly by the public as information sources in locations such as museums, hotels and hospitals, and for self-service in locations such as retail establishments and airports. Users walk up to the unit and stand while they interact with it to obtain information or to conclude a transaction. Most kiosks are durable self-contained units with a small footprint in a public area, making finger touches to the screen an attractive mode of input. Our focus is on kiosks or other fixed-location devices, such as copying machines that use fingers for touch input rather than stylus input used with hand-held devices.

An advantageous design feature of touch screen keys is that, unlike physical keys, their parameters, such as size, spacing and location on the screen, can be easily changed via software. With few menu choices and low density of screen information, response key parameters are not critical. However, when screens are filled with information and many response choices are available, a touch menu or touch numeric keypad competes for screen space with the simultaneous display of other product information. Alternatively, two matrices of keys may compete with each other, say, by using a 6×2 menu of months along with a numeric keypad to enter month-day-year dates (e.g. Gould *et al.* 1990). Thus, when screen space is at a

*Author for correspondence. e-mail: herbert.colle@wright.edu

Ergonomics ISSN 0014-0139 print/ISSN 1366-5847 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/00140130410001724228

premium, it is desirable to keep the dimensions of a touch screen menu or keypad as small as possible without degrading usability. In their early review of ergonomic issues in touch screen use, Pfauth and Priest (1981) identified key size as an important factor. They recommended that keys should have a minimum size of 22 mm, but the recommendation was based on a single unpublished study that had used a hierarchical menu display in a cockpit simulation. Since then, two different approaches have been used to determine key size recommendations.

1.1. *Response location variability*

One approach to determining the minimum size needed for touch input has been to measure response location variability when people attempt to touch a spatial target location (Hall *et al*. 1988, Beringer 1990, Leahy and Hix 1990, Sears 1991). The size of a square envelope needed to capture a given percentage of first touch responses can be estimated from these data without having to experimentally manipulate the size of touch recognition fields. This capture envelope procedure was most clearly described by Hall *et al*. (1988) who found that a tactile recognition field should be 26 mm square to capture 99% of biconditional $x-y$ responses made while standing at a display with minimal parallax. The criterion of 99% was chosen because this was the performance level that reached a statistical asymptote. For touch displays with parallax, a capture envelope of 30 mm was estimated for standing entry and 26 mm for seated entry. Sears (1991) also estimated an envelope of 26 mm for seated entry using a 100% capture criterion, but he argued that a region as small as 22.7 mm could be used if the tactual recognition field was not centred on the visual target and a criterion of 99% was used. Beringer's (1990) study produced the smallest estimated square envelope of 20 mm, possibly because not only were his users seated but they also were provided with touch accuracy feedback on half of the trials. He argued that variability depends on the instructional emphasis on accuracy. In contrast, Leahy and Hix (1990) found the largest capture envelopes in a field study in which people stood for a short testing time at a kiosk in a grocery store. They estimated square envelopes of 36 and 52 mm for the top left and bottom right quadrants of the display screen, respectively.

1.2. *Key size manipulations*

A second approach for evaluating required key sizes has been to manipulate key size experimentally (Whitfield *et al*. 1983, Beaton and Weiman 1984, Martin 1988, Sears *et al*. 1993, Scott and Conzola 1997, Wilson *et al*. 1995, Bender 1999). We refer to key size as the size of the tactual recognition field, not the size of the visually displayed key, although they are congruent in most studies. By systematically manipulating key size, performance on menu, keypad, or keyboard tasks was evaluated directly, instead of using response location variability to estimate key size from a capture envelope. Our focus is on studies that used a land-on response criterion. A land-on criterion is similar to a first touch criterion except that only a touch to one of the tactile recognition fields (keys) is accepted by the system. Responses outside any of these fields are ignored. Although a lift-off response criterion has been found to allow accurate entry with smaller key sizes (Sears *et al*. 1993, Scott and Conzola 1997, Whitfield *et al*. 1983), it is an atypical action for most keying devices and has not been used extensively for touch input.

Surprisingly, although response variability studies recommended square tactile field envelopes that ranged from 20 to 52 mm, only one study that manipulated key size reported data for a key size greater than 20 mm. Bender (1999) found better performance for 30 mm than 10 mm square keys when users entered four digits in a numeric keypad while standing. Wilson *et al.* (1995) used two smaller key sizes. They found better performance with 20 mm than 14 mm keys for users who stood while entering nine digits in a keypad. Martin (1988) used even smaller keys. She found better performance for a 13×13 mm key size than for 6×13 or 13×6 mm key sizes when users entered one to three digits in a keypad. In contrast to these numeric keypad studies, Beaton and Weiman (1984) used menu entry, which only required a single touch. A 4×3 matrix of keys labelled with CVCs had four sizes, 5×10, 5×20, 10×10, and 10×20 mm. Performance was best with the 10×20 mm keys. They also obtained user preferences, which favoured the largest 10×20 mm keys. Their study was most similar to response variability studies because only one touch entry was made without feedback, and only errors were recorded. The other keypad studies measured both errors and entry time.

In summary, studies have agreed that larger key sizes led to better performance. These results were consistent with capture envelope studies based on response variability, because the larger key sizes that were used were closer to the size recommended by the capture envelope studies. However, an independently estimated minimum size for optimal performance could not be obtained from these data.

1.3. *Key spacing*

Matrices of keys also have another parameter that must be considered along with key size. Keys may have spaces between adjacent keys. With physical keyboards, key spacing increases the effective target width because a finger that lands primarily in the space between keys is still likely to activate a target key that it just nicked (Drury and Hoffmann 1992, Hoffmann 1994, Hoffmann *et al.* 1995). However, with a land-on criterion for touch entry, 'space' between two keys does not increase the target width because touch screen keys operate on a different principle than physical keys. The system first determines the location of the touch as a point estimate and then compares this point estimate with the tactile recognition fields. A touch that is primarily to the inter-key space only nicking the target key is unlikely to be registered because the point estimate is outside the key's tactile recognition field. Thus, space may prevent non-target keys from being falsely activated but it is not likely to increase the effective target width. A more complete discussion of this issue can be found in section 4.

Both Martin (1988) and Beaton and Weiman (1984) manipulated key spacing, defining it as the edge-to-edge (not centre-to-centre) distance between adjacent keys. Martin found that a spacing of 6 mm produced better performance than a spacing of 13 mm. Beaton and Weiman used edge-to-edge spacing of 0, 5, 10, 15, and 20 mm factorially combined on both the vertical and horizontal axes. There was a statistically significant three-way interaction of vertical spacing, key height and key width. For the larger 10 mm high keys, errors increased with greater vertical spacing. However, errors appeared to be relatively constant in the range 0–10 mm of vertical spacing. For the small 5 mm high keys, a different pattern was found. Larger spacing reduced errors. Users preferred vertical spacing of 5 or 10 mm and horizontal spacing of 10 mm.

1.4. *Experimental approach*

In general, previous studies have found that larger key sizes and smaller edge-to-edge spacing led to better performance. However, we undertook the present study to obtain guidelines for key size that are more definitive, along with additional information about key spacing especially at smaller intervals. Our research approach used the strategy defined by Hall *et al.* (1988) of looking for asymptotic performance, but we did not use their capture envelope procedure. Instead, we manipulated key size to determine where the dependent variables of entry time, errors and preference ranking reached estimated asymptotes. Key sizes were concentrated at the low end so we did not expect to be able to distinguish an asymptote from a relatively flat minimum, but the criteria are functionally equivalent when the objective is to determine the smallest easily usable array. In addition, edge-to-edge spacing at the lower range of distances were explored using two levels of key spacing.

Data entry in a numeric keypad was used as a task because it has practical significance and it allowed a natural manipulation of the number of key entries. Participants entered string lengths of one, four and 10 digits. Entering a single digit is most similar to selecting an alternative from a menu. Entering four digits successively is more complex because it may entail more extensive planning or programming of the motor sequences (Sternberg *et al.* 1978, Willingham 1998, Klapp 2003) as well as visual search and movement, but the digit string may be held in working memory while digits are entered. With 10 digits, however, users will not be able to enter strings correctly (Nordby *et al.* 2002), unless they look back at least once to review the string presentation. Theoretically, long strings stress the capacity of motor programming, and practically they provide generality by encompassing the range of successive entries that might be made on a kiosk. Two additional dependent variables, first transition time and initiation time (also called reaction time), were examined because they are of theoretical interest for both motor programming and Fitts' law models (e.g. Mohagheghi and Anson 2002).

2. Method

2.1. *Participants*

Twenty students (11 male and 9 female) from introductory psychology courses at Wright State University received credit for participating in this experiment. All participants were right handed and reported having no motor impairments and normal or corrected to normal vision. Mean age was 19.5 years, ranging from 18 to 22. Participants' touch screen experience ($M = 2.60$, SD $= 1.27$) ranged from 1 (little experience) to 5 (very much experience). Index finger width was measured twice ($r = 0.93$) at a distance of 1 cm from the fingertip. Finger width is an anthropometric characterisation of the participants. Garrett (1971) reported that the mean width of the index finger at the distal joint was 18.3 mm and 15.5 mm for male and female participants, respectively. Width of the fingertip pad, which is only 7 – 12 mm, is more important for physical keypad use (Drury & Hoffmann 1992, Hoffmann *et al.* 1995). Section 4 provides more information about the role of fingertip width. Finger width was greater for male ($M = 18$ mm) than for female ($M = 15$ mm) participants ($F(1, 18) = 6.43$, $MSE = 0.712$, $p < 0.05$).

2.2. *Kiosk interface*

A NCR 7401 Web Kiosk (Pentium 266 MHz) was used for testing. It had a 12.1-inch (24.8 cm $\times$ 18.4 cm) SVGA TFT active matrix LCD visual display with a resolution of 800 $\times$ 600 pixels. To input touch responses, the kiosk used 3M ClearTek$^{\mathrm{TM}}$ capacitive touch technology, which had addressable co-ordinates of 16 k $\times$ 16 k with touch co-ordinates reported within 1% of true position. A 3 ms touch contact was required. The touch screen was tilted back so that it was 60° from the horizontal and the centre of the screen was at a fixed height of 1.22 m from the floor for all participants. Its base and the platform it sat on were both flush with the edge of a tabletop.

Interface displays were created using Visual Basic in a Windows NT operating system. Figure 1 shows the layout of the kiosk display. The keypad, which had the 10 digits configured in a telephone layout, was located at the bottom right of the screen. All keys in the keypad, except the Enter key, were square. The vertical height of the Enter key was always twice the height of the other keys, and it was located at the bottom right hand side of the keypad. A Backspace key was above the Enter key. Keys acted as momentary contact switches using a land-on criterion for registering a touch entry. When touched, keys displayed a depressed state to provide visual feedback to users. Auditory feedback was not used. The keypad included a register located just above the keys, which displayed the digits as they were entered. The register and the size of the digits in the register remained constant with changes in key size and spacing. Each key was labelled with a digit or function in 12-point font (3 mm high), which was held constant for all key sizes and spacing. The keypad was located in the bottom right hand corner of the screen. The keypad's lower right corner was an anchor point so that larger displays expanded to the left and upwards for larger key size and key spacing. The keypad graphics we used are available in Visual Basic. A picture of the visual characteristics of the keypad can be found in figure 1 of Scott and Conzola (1997). The kiosk not only displayed a keypad, but it controlled testing by visually

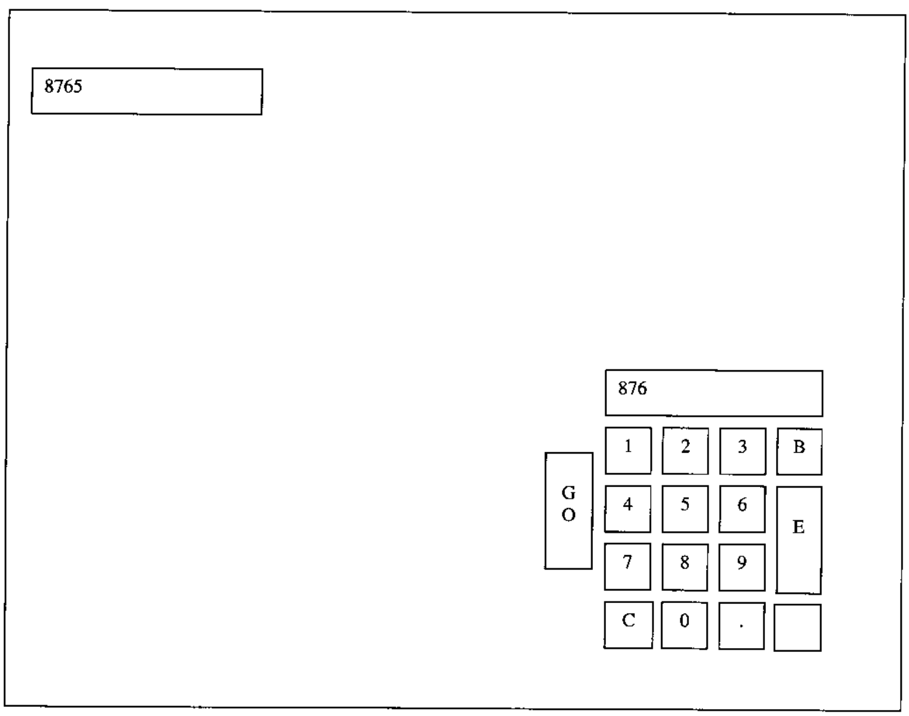

Figure 1. Schematic layout of the kiosk display.

presenting the digit string to be entered in the upper left-hand corner of the screen when a Go key was touched to start each trial.

2.3. *Procedure*

Key size, key spacing and string length were combined factorially in a $4 \times 2 \times 3$ design. Key sizes were 10, 15, 20 and 25 mm on each side. Key spacings were 1 or 3 mm edge-to-edge distances between adjacent keys. String lengths were one, four, or 10 digits.

Testing was conducted in eight blocks of 54 trials per block. Size and spacing were constant within a block while string lengths were presented randomly across trials within each block so that each string length was presented 18 times. Block order was determined by using Latin squares to balance the order of testing each combination of size and spacing across participants. The first 16 participants were randomly assigned to an order determined by two independent 8×8 Latin squares. The last four participants were randomly assigned to four of the rows of a third independently generated Latin square. Each participant completed 18 test trials on each of the 24 experimental conditions, and received nine practice trials prior to the start of each block. A complete session, including instruction, practice and a rest break, was completed in 1–1.5 h.

Participants started a trial by touching the Go key, which presented a randomly-generated string of the digits 0–9 in the upper left-hand corner of the kiosk screen. As digits were entered, they were displayed in the keypad's register above the keys and the Backspace key could be used to correct a mistake. A trial ended when the Enter key was touched, which also removed the displayed digits and those in the keypad register. Participants were not restricted from using more than one finger to operate the keypad, and 14 participants did so some of the time. However, number of fingers (1, > 1) and interactions of it with other factors were not statistically significant.

Participants stood in front of the kiosk to enter digit strings on the displayed keypad. A participant's standing distance from the tip of his or her shoe to the edge of the table was unobtrusively observed by the experimenter twice from marks placed on the floor, once at the start of testing and once at the end. Each foot was measured separately. Mean standing distance was 14.5 cm at the start of testing which decreased to 11.9 cm at the end of testing ($F(1, 18) = 5.76$, $MSE = 120.7$, $p < 0.05$). There were no significant main effect of foot, test time or gender and no interactions among these factors. At the completion of all kiosk trials, participants were given actual size paper copies of the eight keypads created by the four key size and two key spacing combinations. They indicated their display preference by rank ordering the eight displays ('1' = best, '8' = worst).

3. Results

3.1. *Total time*

Total time, which was defined as the time from the onset of the Go key to the onset of the Enter key, was analysed using a $4 \times 2 \times 3$ repeated-measures ANOVA with factors of key size (10, 15, 20, 25 mm), key spacing (1, 3 mm) and string length (one, four, 10 digits). Trials on which an error was made, either corrected or not, were eliminated from the data set and the mean of the remaining trials was computed.

Figure 2 presents the mean total time for key size, spacing and string length with 95% confidence intervals about each data point. The confidence intervals used for graphic display were computed using the MSE from the three two-way interactions

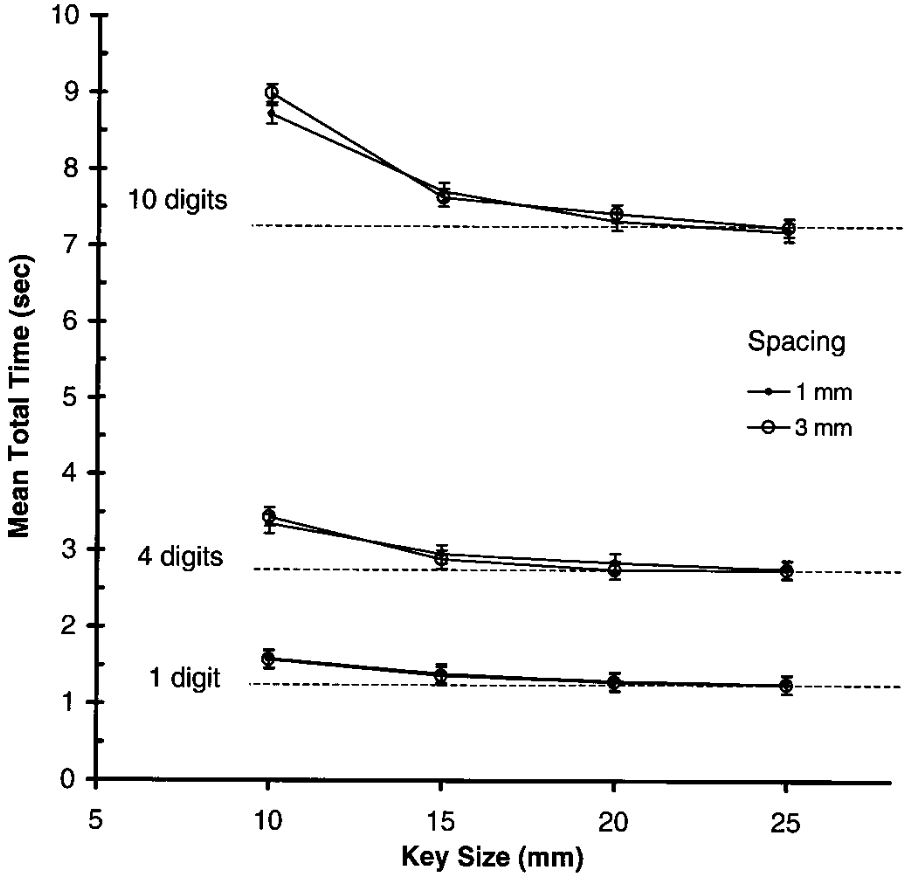

Figure 2. Mean total time to enter all digits for key size, key spacing and string length. Error bars represent 95% confidence intervals. The dashed horizontal lines were set equal to the mean time for the largest key size.

because there were substantial differences in the seven error terms. These three error terms were most relevant to the experimental questions and they were approximately equal in size (Loftus and Masson 1994). However, individual best-estimate error terms were used for the reported statistical tests. A 5% level of confidence after Greenhouse-Geisser correction was used to determine statistical significance.

As would be expected, longer digit strings, which required more touch entries, took more total time. Mean total time to respond was 1.38 s, 2.97 s, and 7.78 s for string lengths of one, four, and 10 digits, respectively ($F(2, 38) = 437$, $MSE = 4.07$, $p < 0.001$). More importantly, total time decreased as key size increased. The main effect of key size was statistically significant ($F(3, 57) = 29.1$, $MSE = 0.633$, $p < 0.001$). Dashed horizontal lines were drawn in figure 2 at the mean time of the largest key size (25 mm) to facilitate comparisons with smaller key sizes. Total time appeared to reach an asymptote at a key size of 20 mm, which was also supported by step down ANOVAs that had comparable power to detect differences. Total time differences could be discriminated until only key sizes of 20 and 25 mm remained. These two conditions were not statistically significant at the 5% level of confidence.

The effect of key size interacted with string length ($F(6, 114) = 19.7$, $MSE = 0.197$, $p < 0.001$). As figure 2 shows, although key size had a larger effect on longer digit strings, the effect of key size was still present for the shorter strings. Total times appeared to reach the same asymptote of 20 mm for the three string lengths. Follow-up ANOVAs supported this description.

Key spacing had no measurable effect on total time. The main effect of spacing was not statistically significant ($F(1, 19) < 1.0$, $MSE = 0.383$), and spacing did not interact with either key size or string length ($F(3, 57) < 1.0$, $MSE = 0.252$, and $F(2, 38) < 1.0$, $MSE = 0.155$, respectively). The three way interaction also was not statistically significant ($F(6, 114) < 1.0$, $MSE = 0.093$). As figure 2 shows, performance with 1 mm or 3 mm spacing was similar at all combinations of key size and string length, except perhaps for a small difference using the smallest key size and the longest digit string.

Gender effects were explored using an additional $2 \times 4 \times 2 \times 3$ ANOVA with gender added as a between-subjects factor. The main effect of gender was not statistically significant ($F(1, 18) = 2.25$, $MSE = 11.8$, $p > 0.05$). Mean total times for male and female participants were 4.25 s and 3.78 s, respectively. Gender also did not interact with any of the other factors. The gender by key size and gender by spacing interactions were $F(3, 54) < 1.0$, $MSE = 0.651$, and $F(1, 18) = 1.16$, $MSE = 0.380$, $p > 0.05$, respectively. The gender by string length was significant when probability was uncorrected ($F(2, 36) = 3.97$, $MSE = 3.52$, $p = 0.028$), but it did not reach the 0.05 level with a Greenhouse-Geisser correction ($p = 0.059$). All other interactions had F ratios of less than 1.

3.2. *Keying accuracy*

A digit string was considered to be correct only if no errors, corrected or not corrected, were made. Number of correct strings were transformed to percent error for presentation, and were analysed using a $4 \times 2 \times 3$ repeated-measures ANOVA with factors of key size, key spacing and string length. Key size, string length and their interaction were all statistically significant ($F(3, 57) = 12.0$, $MSE = 61.9$, $p < 0.001$, $F(2, 38) = 76.5$, $MSE = 89.7$, $p < 0.001$ and $F(6, 114) = 6.02$, $MSE = 35.2$, $p < 0.001$, respectively). As figure 3 shows, when only one digit was entered, mean percent error ($M = 2.13\%$) did not depend on key size. However, errors did decline as key size increased for string lengths of four and 10 digits. Although the 4- and 10-digit curves in figure 3 appear to reach an asymptote at a 20 mm key size, key sizes of 15, 20 and 25 mm could not be discriminated statistically.

The main effect of key spacing again was not significant ($F(1, 19) < 1.0$, $MSE = 53.6$). Spacing did not interact with size, string length nor with the combination of both effects ($F(6, 114) < 1.0$, $MSE = 62.0$, $F(2, 38) < 1.0$, $MSE = 47.8$ and $F(6, 114) = 1.56$, $MSE = 41.0$, $p > 0.05$, respectively). A follow-up analysis of gender found no main effect ($F(1, 18) < 1.0$, $MSE = 241$), and only the 4-way interaction was significant ($F(6, 108) = 2.64$, $MSE = 37.7$, $p < 0.05$). It was not readily interpretable.

3.3. *Other time measures: Initiation and first transition times*

The time to initiate the first finger movement was defined as the time from the onset touch of the Go key to start a trial to its release. These data were also analysed using a $4 \times 2 \times 3$ repeated-measures ANOVA with factors of key size,

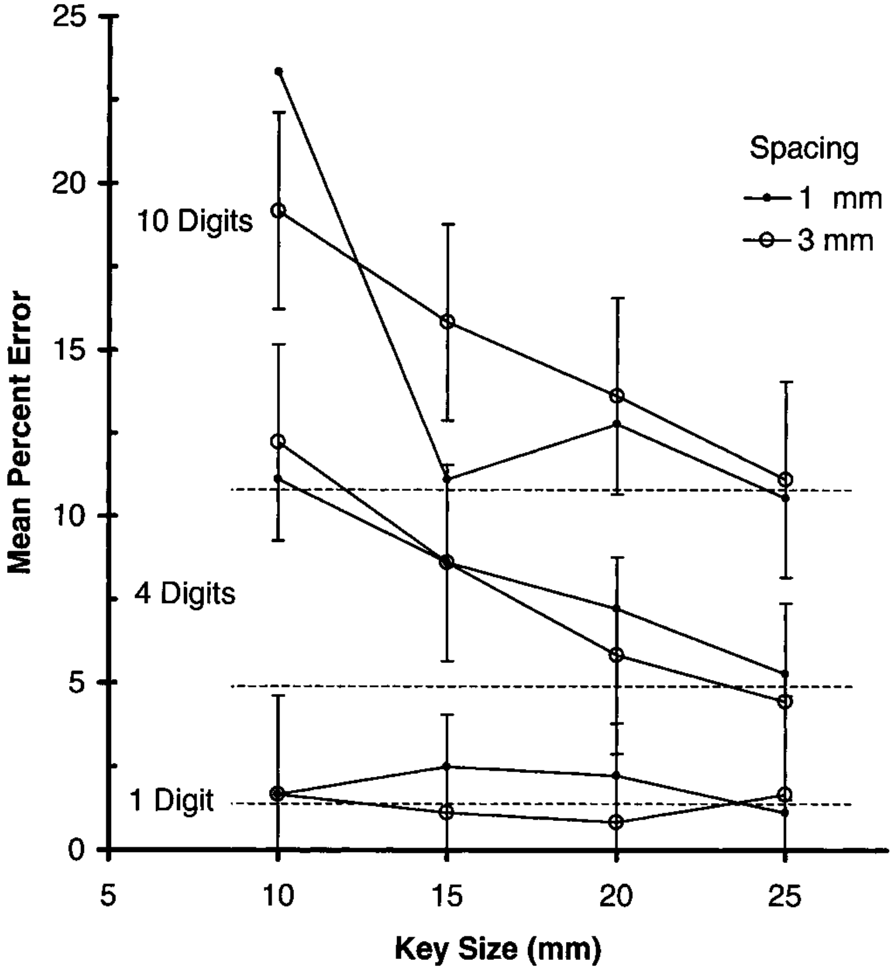

Figure 3. Mean percentage of digit strings with at least one error for key size, key spacing, and string length. To reduce clutter, 95% confidence intervals are shown only for 3 mm spacing. The dashed horizontal lines were set equal to the mean percent error for the largest key size.

key spacing and string length. Initiation times were considerably smaller than total times, but the main effect of key size was significant ($F(3, 57) = 24.9$, $MSE = 0.001$, $p < 0.001$). Mean initiation times from small to large key size were 0.140, 0.116, 0.109, 0.103 s, which again appeared to reach asymptote at a key size of 20 mm (supported by follow-up ANOVAS). Neither string length nor the size by string length interactions were statistically significant ($F(2, 38) < 1.0$, $MSE = 0.00008$ and $F(2, 38) < 1.0$, $MSE = 0.00005$, respectively). However, the main effect of key spacing was statistically, but not practically, significant ($F(1, 19) = 4.84$, $MSE = 0.0005$, $p < 0.05$). Mean initiation time was 0.120 s for 1 mm and 0.115 s for 3 mm spacing. All other interactions were not significant. A follow-up ANOVA adding gender found no main effect of gender and no interactions of gender with other effects.

First transition time was defined as the time from the release of the Go key to a land-on touch of the first digit key. A $4 \times 2 \times 3$ repeated-measures ANOVA with factors of key size, key spacing and string length was used to analyse the data. Only the main effects of key size and string length were statistically significant ($F(3, 57) = 12.8$, $MSE = 0.068$, $p < 0.001$ and $F(2, 38) = 24.6$, $MSE = 0.227$, $p < 0.001$, respectively). As figure 4 shows, first transition time appeared to reach an asymptote at a key size of 15, which was supported by follow-up analyses. The main effect of spacing was $F(1, 19) < 1.0$, $MSE = 0.062$. All interactions with spacing had F ratios of 1 or less.

Gender was evaluated by using $2 \times 4 \times 2 \times 3$ mixed effects ANOVA with gender as a between-subjects factor. The main effect of gender was statistically significant ($F(1, 18) = 4.46$, $MSE = 0.882$, $p < 0.05$). Mean first transition times were 1.14 s for males and 0.95 s for females. However, gender did not interact with key size nor with key spacing ($F(3, 54) < 1.0$, $MSE = 0.070$ and $F(1, 18) < 1.0$, $MSE = 0.065$, respectively). Gender $\times$ string length was significant when probability was un-

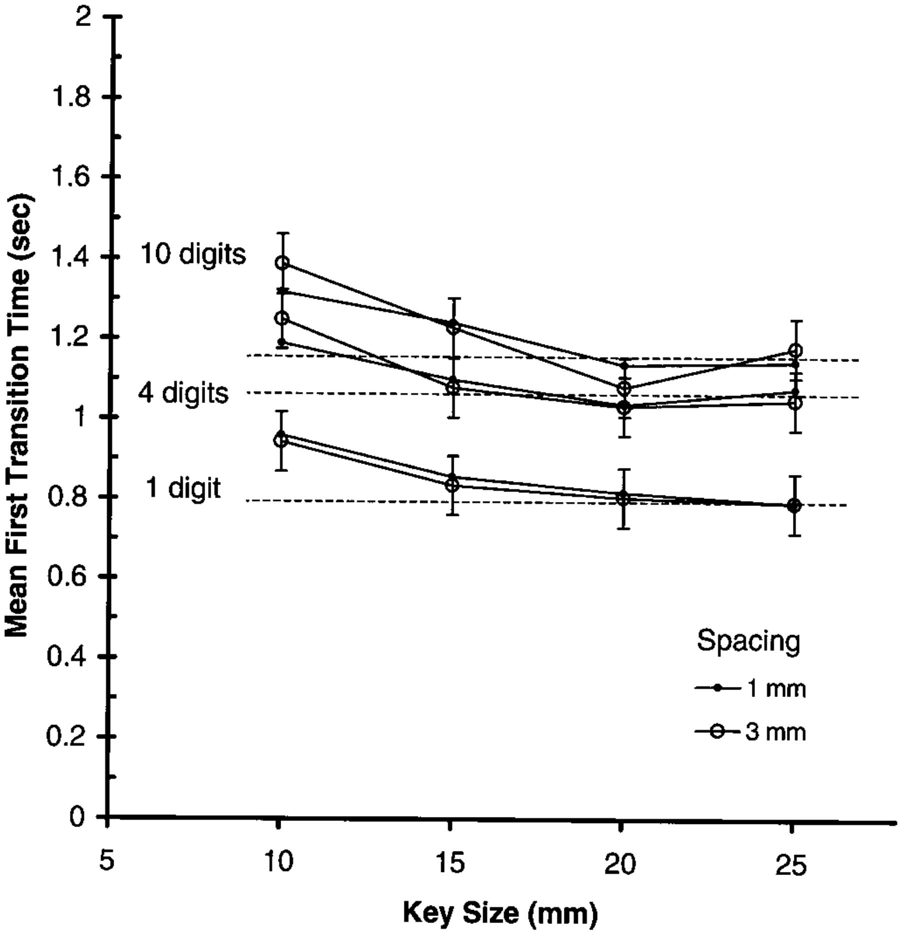

Figure 4. Mean first transition time for key size, key spacing and string length. To reduce clutter, 95% confidence intervals are shown only for 3 mm spacing. The dashed horizontal lines were set equal to the mean time for the largest key size.

corrected ($F(2, 36) = 3.97$, $MSE = 0.196$, $p = 0.028$), but it did not reach the 0.05 level with a Greenhouse-Geisser correction ($p = 0.056$). All other interactions were not significant.

3.4. *Keypad preferences*

Participants rank-ordered the eight size-by-spacing keypad displays ('1' = best, '8' = worst) to indicate their preferences. Rank orders of the eight conditions were analysed using a Friedman test for matched groups, which was statistically significant ($F_r(7) = 53.4$, $p < 0.001$). Table 1 shows the mean ranks. In order to test for key size preferences, the original rank orders were used to generate a rank order for the four key sizes by collapsing over spacing (mean rank for both spacing conditions for each key size). These derived rank orders were also analysed using a Friedman test, which was statistically significant ($F_r(3) = 22.4$, $p < 0.001$). Mean preference ranks (possible range = 1 to 4) were 3.58, 2.60, 1.85, and 1.98, for key sizes 10, 15, 20 and 25 mm, respectively. These results suggest that participants were indifferent to key sizes 20 vs. 25, but preferred them to the two smaller key sizes. Follow-up analyses were consistent with this interpretation. A comparable analysis was performed on key spacing, using the original rank order to derive a rank order for just the two spacing conditions. A Friedman test of these ranks was not statistically significant ($F_r(1) = 1.80$, $p > 0.05$). Mean preference ranks (possible range = 1 to 2) were 1.65 and 1.35 for 1 and 3 mm spacing, respectively.

4. Movement models

Several models have been developed to describe movement time performance on physical keyboards, which were tested using the Fitts movement time paradigm (Drury and Hoffmann 1992, Gan and Hoffman 1988). These models make predictions about the effects of key size and spacing. Therefore, it is instructive to consider how these models relate to the present experiment. It should be noted, however, that the present task was not a Fitts movement time paradigm because, in addition to movement, it required choice and motor programming processes, activities that are common when people use kiosk keypads. Therefore, the present experiment can not be used for a critical evaluation of movement models.

4.1. *Visually guided movement model*

Drury and Hoffmann's (1992) model was based on Fitts' law, which assumed that movements were visually guided. The model also assumed: (a) that keys are all the same size, (b) that keys are square, (c) that inter-key spacings are constant, and (d) movements are made in horizontal or vertical directions. Assumption (d) can be reasonably extended to diagonal movements, as Hoffmann *et al*. (1995) suggested it

Table 1. Mean preference ranks for the eight kiosk keypad displays

	Key Size (mm)				
Key Spacing (mm)	10	15	20	25	*M*
1	7.0	5.1	3.0	3.4	4.6
3	6.4	4.2	3.2	3.6	4.4
M	*6.7*	*4.6*	*3.1*	*3.5*	

Rank of '1' was best display (range 1 to 8).

should be. The following description of Drury and Hoffmann's (1992) model uses their nomenclature. According to Fitts' law, movement time, MT, can be described by the equation,

$$\mathrm{MT} = \mathrm{a} + \mathrm{b(ID)}, \text{ where } \mathrm{ID} = \log_2(2\mathrm{A}/\mathrm{W}). \tag{1}$$

ID is the Index of Difficulty (a and b are constants), and A and W are the amplitude of the movement and target size, respectively. In applying Fitts' law to physical keyboards, Drury and Hoffmann's model included the width of the probe, usually a finger for keypads, as a component of the effective target size, W. Finger width, F, is an important factor for physical keyboards because target size, W, depends on both key size, B, and the width of the finger. Finger width extends the range of activation of a key because a key is activated if the edge of a finger overlaps a portion of a key and there is sufficient space between keys so that it does not depress another key. Two cases were described. If the edge-to-edge separation, C, is greater than finger width, then $W = B + F$, and if the edge-to-edge separation is less than finger width, then $W = 2S - B - F$, where S refers to centre-to-centre spacing, $S = B + C$.

In the present experiment, finger width was clearly less than the edge-to-edge separation. The largest edge-to-edge separation was 3 mm compared to an estimated width of the finger pad of at least 7 mm (Drury and Hoffmann 1992, Hoffmann *et al.* 1995). However, touch screen keys have a different mode of operation than physical keys. Touch screens operate by first computing a centroid for the touch, producing a point estimate of an x–y coordinate on the screen. This coordinate is then compared with the tactual recognition field coordinates for all defined keys. If it falls within one of the fields, that key is activated. If it falls outside all fields, no key is activated. Importantly, if a finger pad's width overlaps a tactual recognition field but the centroid of the touch is outside it, the key will not be activated. Finger width does not extend the range of activation of touch screen keys. Therefore, in order to apply the Drury-Hoffmann model to touch keypads, a third case is needed in which effective target size equals key size, $W = B$.

First transition time data in figure 4 is the simplest data set to consider. It is the mean time for a single movement from the Go key to one of the digit keys, key i. Each horizontal or vertical key amplitude, A_i, has the form $A_i = k_i(B + C)$ for a keypad with square keys (k_i is an integer indicating the number of keys between the target key and the starting position). For example, the amplitude of a movement from the Go key to the six key would be $3(B + C)$ mm with $k_6 = 3$. This relationship also holds for keys that may not be reached with a horizontal or vertical movement, if it is assumed that the amplitude of a movement is the diagonal distance. In this case, k_i may be a real number. For example, the diagonal distance from the Go key to the 'nine' key can be computed from the Pythagorean theorem based on the two legs. The horizontal leg across would be $3(B + C)$ and the vertical leg down would be $1(B + C)$. The computed distance would be $[9(B + C)^2 + 1(B + C)^2]^{1/2} = 10^{1/2}$ $(B + C)$ mm, so that $k_9 = 10^{1/2}$. Therefore, the ID in equation (1) for a movement to key i becomes

$$\mathrm{ID_i} = \log_2[2\mathrm{k_i}(\mathrm{B} + \mathrm{C})/\mathrm{B}] = 1 + \log_2 \mathrm{k_i} + \log_2[1 + \mathrm{C}/\mathrm{B}]. \tag{2}$$

The means in figure 4 include all 10 movement amplitudes to all 10 target keys instead of presenting data for each key separately, as is typically done for Fitts' law analyses. To compare these data with Fitts' law, a mean ID, M(ID), for all key amplitudes must be derived. Mean ID is

$$M(\mathrm{ID}) = (1/\mathrm{N})\sum[\mathrm{ID_i}], \tag{3}$$

where the summation is across the number of key amplitudes, N = 10 in this case. Using equation (2), equation (3) can be shown algebraically to be

$$M(\mathrm{ID}) = 1 + (1/\mathrm{N})\sum \log_2 \mathrm{k_i} + \log_2[1 + \mathrm{C/B}]. \tag{4}$$

Note that the mean ID depends directly on $\log_2[1 + C/B]$ as it does for each individual ID_i in equation (2). Averaging across amplitudes only affects the constant that depends on the set of k_i. In the present experiment, key size, B, and edge-to-edge spacing, C, were experimentally manipulated. The values $\log_2[1 + C/B]$ for all eight combinations of B and C are shown as the top entries in table 2. Mean ID can be found by adding the constant $1 + (1/N)\Sigma\log_2 k_i = 2.073$ to each value in table 2. Because movement time is a linear function of ID according to Fitts' law, movement time should be a linear function of the values in the top entries of table 2. Therefore, as table 2 shows, edge-to-edge spacing should have had at least as great an effect as key size on movement time. However, spacing had no measurable effect on first transition time, although the effect of key size was statistically significant.

It should be noted that the prediction about the relative effects of key size and key spacing also holds for some other definitions of the ID ratio (e.g. MacKenzie 1992). For example, ID can has also been defined as a signal-to-noise ratio, $ID_{SN} = \log_2[(W + A)/W)] = \log_2[1 + A/W]$ so that $ID_{SNi} = \log_2[k_i(1 + 1 + C/B)] = \log_2 k_i + \log_2[2 + C/B]$. Using this version, equation (3) would become $M(ID_{SN}) = (1/N)\Sigma\log_2 k_i + \log_2[2 + C/B]$. These values follow the same pattern as the top entries in table 2, indicating that key spacing again should have had at least as great an effect as key size. Adjusting target width based on accuracy also yielded similar results. Percent error was 2.13% for the first transition and it did not change significantly across all experimental conditions. Estimating effective target width for a nominal 4% error using the technique described by MacKenzie (1992) changed the effective target width, W_e, in the equations to 0.89W, or $B_e = 0.89B$ for all experimental conditions, which did not change any conclusions.

4.2. *Ballistic movement model*

One possible reason why Fitts' law does not describe these data is that the model assumes that movements are visually guided, which they may not have been in the present experiment. Movement amplitudes were small, the mean movement amplitude was 63 mm for the keypad with the largest amplitude, and therefore the movements may have been ballistic, not visually-guided. Gan and Hoffmann (1988) presented data that indicated that when ID $\leqslant 3$, movements were very fast (below 200 ms) and were made ballistically, not visually guided. Note that the largest value in table 2 is 0.378 for a key size of 10 mm and a spacing of 3 mm, so that the largest M(ID) = 2.45, and for this keypad, the longest move (to the digit nine) had an $ID_9 = 3.04$. Therefore, all mean IDs were about three or less, and Fitts' law would not be expected to apply. Gan and Hoffmann (1988) found support for a model for ballistic movements in which movement time was described by

Table 2. Values of $\log_2[1 + C/B]$ and $(B + C)^{1/2}$ for all combinations of key size, B, and edge-to-edge key spacing, C

	Key Size (mm)				
Key Spacing (mm)	10	15	20	25	*M*
1	0.138	0.093	0.070	0.057	0.089
	3.32	4.00	4.58	5.10	4.25
3	0.378	0.263	0.202	0.164	0.252
	3.61	4.24	4.80	5.29	4.48
M	0.258	0.178	0.136	0.110	
	3.46	4.12	4.69	5.20	

Top entries in the table are values of $\log_2[1 + C/B]$ for the Drury and Hoffmann (1992) visually-guided movement model, and bottom entries are values of $(B + C)^{1/2}$ for the Gan and Hoffmann (1988) ballistic movement model. Mean ID for transition time data in Figure 4 can be obtained by adding 2.073 to the top entries. Mean $A^{1/2}$ can be obtained by multiplying the bottom entries by 1.475. Movement time is a linear function of the tabled entries for each model.

$$\mathrm{MT} = \mathrm{a} + \mathrm{bA}^{1/2}. \tag{5}$$

In the present experiment the amplitude to a particular key is $A_i = [k_i(B + C)]$ so that $A_i^{1/2} = [k_i(B + C)]^{1/2} = k_i^{1/2}(B + C)^{1/2}$ and the mean amplitude

$$M(A^{1/2}) = (1/N)\sum[k_i(B + C)]^{1/2} = [(\sum k_i^{1/2})/N](B + C)^{1/2}. \tag{6}$$

Again, note that mean amplitude depends directly on $(B + C)^{1/2}$ as it does for each individual amplitude. Averaging across amplitudes only affects the constant $(\Sigma k_i^{1/2})/N$ that depends on the set of k_i. The bottom entries in table 2 show the values of $(B + C)^{1/2}$ for the eight experimental conditions. Movement times should again be a linear function of the values in table 2. The values show that the effect for spacing should not be greater than the difference between 20 and 25 mm key size, consistent with the data in figure 4. Importantly, however, the values for key size indicate that movement time should have increased as key size increased because larger key sizes increase movement amplitude. Instead, the data indicated that movement time decreased with larger keys.

4.3. *Reaction time and movement time*

It should be noted that although the IDs were three or less, mean transition times in the current experiment were all greater than 800 ms, considerably greater than the 200 ms movement times reported by Gan and Hoffmann (1988). Typically, movement time experiments are designed to minimize the effects of response selection and motor programming, or the experiments are designed to isolate these effects from the effects of movement time. The present paradigm was most similar to the discrete movement version of the Fitts paradigm (Mohagheghi and Anson 2002). In this paradigm, participants hold a probe on the starting position until an imperative cue is presented, then they move to either a target on the left or a target on the right. The time from the onset of the imperative cue to when the probe is lifted from the table, the initiation time, is considered to be an index of reaction time, and the first (and only) transition time is considered to be an index of movement time. Thus, it is assumed that two processing stages can be isolated by using these two dependent variables. However, it is unlikely that this assumption was tenable in the present experiment. First transition time probably included some choice time and motor programming time because mean initiation times in the current experiment were very short, ranging from 103 – 140 ms. These were too short for a 3.2 bit choice reaction time. In the present experiment trials were not initiated with a finger resting on the starting position (Go key). Apparently, the Go key was pressed and released rather rapidly, a common action in key pressing.

Thus, first transition times may not have been exclusively measures of movement times. The keypad entry task required two other types of processes, which could have influenced the data. Both a relatively difficult response selection and motor programming for multiple responses were required because a choice paradigm was used in which participants did not know which response movement was to be executed until the imperative cue (the list of digits displayed) was presented (Klapp 1996, 2003). Response selection would have added considerable time to the first transition time. In order to select the first movement in the present keypad task, participants had to make a choice from among ten equiprobable keys (3.2 bits of uncertainty). Therefore, it is not surprising that first transition times were longer than the movement times reported by Gan and Hoffmann (1988), which were made without choice uncertainty. Response selection time, however, should have been a constant because first choice uncertainty was constant for all experimental conditions (Klapp 1996, 2003), which should have just added a constant to the intercept of equation (4) or equation (6). It should not have been influenced by key size or spacing nor by string length.

If reaction time was not isolated from first transition time, then the motor programming of multiple responses also should have contributed to first transition times. When the execution of alternative response movements depends upon information in an imperative cue, responses must be programmed during the interval following the cue, and multiple or complex responses should take more motor programming time (Klapp 1996, 2003). The data shown in figure 4 were consistent with this assumption. First transition times depended on how many digits were to be entered later in the sequence (string length). For example, the single movement from the Go key to a digit took more time when four entries were to be made than when only one entry was to be made. This can be contrasted with a typical movement time paradigm in which the movement to be executed is known in advance and consists of a single movement or reciprocal tapping between two known locations. Klapp (1996, 2003) identified two different processes (referred to as SEQ and INT) that affected the time to respond. These processes depended on how many motor chunks were programmed (SEQ) and on the complexity (INT) of each motor chunk. Both processes are likely to have been active in the present experiment. For example, participants may have processed the digit sequence 3059 as three, zero, five, nine or as thirty, fifty-nine, creating motor programmes for four simple motor chunks or two more complex ones. Importantly, although the results have not been consistent, estimates of programming time have been found to decrease as target size increased (Klapp 1996, Mohagheghi and Anson 2002). Therefore, motor programming, not just movement execution, may have affected the results of the present experiment, and overwhelmed the effects of movement time.

4.4. *Modelling total entry times*

Understanding the full complexity of digit entry on kiosk keypads would require a model of total entry times (see figure 2), as well as first transition times. Models for total times are more difficult to construct. Movement time components alone would have to be considered for three phases (Go key to first digit key, one or more digit key presses, and then last digit key to enter key), as Hoffmann (1994) did for multi-element visually-controlled tasks. Each phase would have different sets of amplitudes. The mean ID or mean $A^{1/2}$ for each phase would use equation (4) or (6), respectively, except that each phase would have its own set of k_i. For example, in the middle phase $1 + (1/N)\Sigma \log_2 k_i = 1.700$ and $(1/N)\Sigma k_i^{1/2} = 1.298$ for the 90 possible non-zero movement transitions. A further complexity, however, is that a key press of a digit could be repeated (digit strings were generated by sampling with replacement) because this is a common occurrence on kiosk keypad entry. Therefore, another process specifying the times for the other 10 zero-amplitude movements (the diagonal of the transition matrix) would need to be specified. Of course, response selection and motor programming also are likely components of phase two. Predictive models for keypad entry of complex tasks incorporating some of these processes, along with some others, have been attempted (e.g. MacKenzie 2003, Pavlovych and Stuerzlinger 2004).

5. Discussion and recommendations

Although it is not clear how to model the data, the empirical results were relatively straightforward. All three major dependent variables generated a relatively consistent pattern of results. A key size of 20 mm always was sufficient for optimal performance or for user satisfaction. A 20 mm key size was a statistical asymptote in the range of sizes used for both total entry time and for user preference ranks. Error data presented a noisier and slightly more complicated picture. When a single digit was entered, percent error did not depend on key size. For strings of 4 or 10 digits, errors decreased with key size but differences among 15, 20 and 25 mm were not statistically significant. However, error bars for three of the four 15 mm conditions did not include the dashed line at the 25 mm level. Given the total time and preference data, a 15 mm key size would not be recommended.

Although the data from the present experiment suggested that performance and preference reached an asymptote at a key size of 20 mm, the true asymptotic level could be in the interval between 15 and 20, because we did not evaluate key sizes in this interval. Key sizes of 17–18 mm might be as usable as 20 mm. Additional research would be needed to narrow the estimated size range.

These results extended the results of previous research that manipulated key size using numeric keypads and menus (Beaton and Weiman 1984, Martin 1988, Wilson *et al*. 1995, Bender 1999). The results of these studies indicated that larger key sizes were better, but we found evidence in all three dependent variables that 20 mm was sufficiently large. The results can also be compared with the conclusions of the response variability studies (Hall *et al*. 1988, Beringer 1990, Leahy and Hix 1990, Sears 1991). Most of these studies recommended square capture envelopes greater than 20 mm. Only Beringer (1990) estimated a square envelope of 20 mm. Response variability studies measured accuracy for single touches. In the present study, one-digit entries were the most comparable condition, and accuracy was 98% correct for all key sizes, including the smallest key size of 10 mm. However, we would not recommend 10 or 15 mm key sizes for single touch entries based on the entry time and preference data. The reasons that higher key size estimates were obtained with response variability procedures are unclear.

Scott and Conzola (1997) found that performance on keypad entry using a lift-off criterion depended on gender or finger size, which were correlated. We found that gender did not interact with key size or spacing, although female participants had smaller fingers. Gender also did not affect response variability in the results of Hall *et al*. (1988).

There were no measurable effects of 1 vs. 3 mm edge-to-edge key spacing. This is important because the results were obtained with a wide range of key sizes and string lengths. Previous studies found that key spacing above 10 mm degraded performance for key sizes close to the 20-mm recommendation (Beaton and Weiman 1984, Martin 1988). Beaton and Weiman's data suggested that there were no performance differences for 0, 5 and 10 mm spacing for 10 $\times$ 20 mm keys. Thus, the data for all studies suggested that an edge-to-edge spacing as small as 0 – 1 mm should not degrade performance. This result is inconsistent with the results found with physical keypads (Drury and Hoffmann 1992, Hoffmann *et al*. 1995). However, it is not surprising because spacing acts differently in touch screens, only increasing centre-to-centre spacing without also increasing effective key size. Beaton and Weiman also found that users expressed a preference for a 5 or 10 mm key spacing. We found no preferences for 1 vs. 3 mm for a wide range of key sizes and keying entries. Testing context, including tasks, feedback and the range of key spacings, could influence differences in subjective preferences.

In summary, based on these data we recommend that key size no smaller than 20 mm with 1 mm edge-to-edge spacing should be used if sufficient space is available. However, if space is very limited, then by interpolation our data allow the possibility that a 17 mm key size might be acceptable, and Beaton and Weiman's (1984) data along with ours suggest that an edge-to-edge spacing as small as 0 mm might be acceptable.

Acknowledgements

We thank Mark Hoffman, CTO, Jackie Huffman and Sally Cohen for suggestions and comments, and NCR Corp. for providing the kiosk. We thank Mark Lee for sharing his keyboard software and for his advice about it and the kiosk, and Brian Porter for writing the new kiosk programmes.

References

Bender, G. T. 1999, Touch screen performance as a function of the duration of auditory feedback and target size, Ph. D. dissertation Wichita State University, Kansas, USA.

Beaton, R. J. and Weiman, N. 1984, *Effects of touch key size and separation on menu-selection accuracy*, Tektronix Technical Report TR 500-01, Tektronix Corporate Library 50-510, Portland, Oregon, USA

Beringer, D. B. 1990, Target size, location, sampling point and instructional set: more effects on touch panel operation, *Proceedings of the Human Factors Society 34th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 375 – 379.

Drury, C. G. and Hoffmann, E. R. 1992, A model for movement time on data-entry keyboards, *Ergonomics*, **35,** 129 – 147.

Gan, K. C. and Hoffmann, E. R. 1988, Geometrical conditions for ballistic and visually controlled movements, *Ergonomics*, **31,** 829 – 839.

Garrett, J. W. 1971, The adult human hand: Some anthropometric and biomechanical considerations, *Human Factors*, **13,** 117 – 131.

Gould, J. D., Greene, S. L., Bois, S. J., Meluson, A. and Rasamny, M. 1990, Using a touchscreen for simple tasks, *Interacting with Computers*, **2,** 59 – 74.

Hall, A. D., Cunningham, J. B., Roache, R. P. and Cox, J. W. 1988, Factors affecting performance using-touch entry systems: Tactual recognition fields and system accuracy, *Journal of Applied Psychology*, **73,** 711–720.

Hoffmann, E. R. 1994, Optimum layout of an array of controls, *International Journal of Industrial Ergonomics*, **14,** 251–261.

Hoffmann, E. R., Tsang, K. K. and Mu, A. 1995, Data-entry keyboard geometry and keying movement times, *Ergonomics*, **38,** 940–950.

Klapp, S. T. 1996, Reaction time analysis of central motor control, in H. N. Zelaznik (ed), *Advances in motor learning and control* (Champaign, IL: Human Kinetics Publishers), 13–36.

Klapp, S. T. 2003, Reaction time analysis of two types of motor preparation for speech articulation: Action as a sequence of chunks, *Journal of Motor Behavior*, **35,** 135–150.

Leahy, M. and Hix, D. 1990, Effect of touch screen target location on user accuracy, *Proceedings of the Human Factors Society 34th annual meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 370–374.

Loftus, G. R. and Masson, M. E. J. 1994, Using confidence intervals in within-subject designs, *Psychonomic Bulletin & Review*, **1,** 476–490.

Martin, G. L. 1988, Configuring a numeric keypad for a touch screen, *Ergonomics*, **31,** 945–953.

MacKenzie, I. S. 1992, Fitts' law as a research and design tool in human-computer interaction, *Human-Computer Interaction*, **7,** 91–139.

MacKenzie, I. S. 2003, Motor behaviour models for human-computer interaction, in J. M. Carroll (ed.), *Toward a multidisciplinary science of human-computer interaction* (San Francisco: Morgan Kaufmann), 27–54.

Mohagheghi, A. A. and Anson, J. G. 2002, Amplitude and target diameter in motor programming of discrete, rapid aimed movements: Fitts and Peterson (1964) and Klapp (1975) revisited, *Acta Psychologica*, **109,** 113–136.

Nordby, K., Raanaas, R. K. and Magnussen, S. 2002, The expanding telephone number Part 1: Keying briefly presented multiple-digit number, *Behaviour & Information Technology*, **21,** 27–38.

Pavlovych, A. and Stuerzlinger, W. 2004, Model for non-expert text entry speed on 12-button phone keypads, *Proceedings of CHI2004*, (New York: Association of Computing Machinery), 351–358.

Pfauth, M. and Priest, J. 1981, Person-computer interface using touch screen devices, *Proceedings of the Human Factors Society 25th Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 500–504.

Sears, A. 1991, Improving touchscreen keyboards: Design issues and a comparison with other devices, *Interacting with Computers*, **3,** 253–269.

Sears, A., Revis, D., Swatski, J., Crittenden, R. and Shneiderman, B. 1993, Investigating touchscreen typing: The effect of keyboard size on typing speed, *Behaviour & Information Technology*, **12,** 17–22.

Scott, B. and Conzola, M. S. 1997, Designing touch screen numeric keypads: Effects of finger size, key size, and key spacing, *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 360–364.

Sternberg, S., Monsell, S., Knoll, R. L. and Wright, C. E. 1978, The latency and duration of rapid movement sequences: Comparisons of speech and typewriting, in G. E. Stelmach (ed.), *Information processing in motor control and learning* (New York: Academic Press), 117–152.

Wilson, K. S., Inderrieden, M. and Liu, S. 1995, A comparison of five user interface devices designed for point-of-sale in the retail industry, *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, (Santa Monica, CA: Human Factors and Ergonomics Society), 273–277.

Whitfield, D., Ball, R. G. and Bird, J. M. 1983, Some comparisons of on-display touch input devices for interaction with computer generated displays, *Ergonomics*, **26,** 1033–1053.

Willingham, D. B. 1998, A neuropsychological theory of motor skill learning, *Psychological Review*, **105,** 558–584.