Alphanumeric entry on pen-based computers

I. SCOTT MACKENZIE, BLAIR NONNECKE AND STAN RIDDERSMA

Department of Computing & Information Science, University of Guelph, Guelph, Ontario, Canada N1G 2W1

CRAIG MCQUEEN

Computer Science Department, University of Toronto, Toronto, Ontario, Canada M5S 1A4

MALCOLM MELTZ

Architel Systems Corp., 190 Attwell Drive, Toronto, Ontario, Canada M9W 6H8

Two experiments were conducted to compare several methods of numeric and text entry for pen-based computers. For numeric entry, the conditions were hand printing, tapping on a soft keypad, stroking a moving pie menu, and stroking a pie pad. For the pie conditions, strokes are made in the direction that numbers appear on a clock face. For the moving pie menu, strokes were made directly in the application, as with hand printing. For the pie pad, strokes were made on top of one another on a separate pie pad, with the results sent to the application. Based on speed and accuracy, the entry methods from best to worst were soft keypad (30 wpm, 1.2% errors), hand printing (18.5 wpm, 10.4% errors), pie pad (15.1 wpm, 14.6% errors), and moving pie menu (12.4 wpm, 16.4% errors).

For text entry, the conditions were hand printing, tapping on a soft keyboard with a QWERTY layout, and tapping on a soft keyboard with an ABC layout (two rows of sequential characters). Tapping on the soft QWERTY keyboard was the quickest (23 wpm) and most accurate (1.1% errors) entry method. Hand printing was slower (16 wpm) and more error prone (8.1% errors). Tapping on the soft ABC keyboard was very accurate (0.6% errors) but was slower (13 wpm) than the other methods.

These results represent the first empirical tests of entry speed and accuracy using a stylus to tap on a soft keyboard. Although handwriting (with recognition) is touted as the entry method of choice for pen-based computers, the much simpler technique of tapping on a soft keyboard is faster and more accurate.

1. Introduction

Pen-based computers herald a new frontier for interactive systems. Small, light weight, untethered—they promise to reach new application domains through ease-of-use and to extend current practice through increased mobility. The technology of pen-based computers is remarkably mature. High-resolution LCD displays with embedded digitizing input match the input/output capabilities of typical desktop computers. The bigger issue is what to do with the technology; or, rather, how to do it.

The most dramatic paradigm shift is input: there is no keyboard. With natural input from a pen, pen-based computers promise to supplant both the QWERTY keyboard for text entry and the mouse for pointing, dragging, and selecting.

However, meeting the input demands of users remains a challenge. Two broad forms of input are (a) gestures to evoke an appropriate action, and (b) alphanumeric input which is converted to ASCII. This paper is concerned with the latter.

In narrow, highly constrained applications (known as "vertical" applications), text entry is minimized by exploiting on-screen options, menus, etc. Even so, limited alphanumeric entry is usually required. For unconstrained applications, such as word processing or spreadsheets, substantial alphanumeric entry will occur.

This paper explores and empirically compares several methods of text and numeric entry for pen-based computers. These include character recognition, two variations of soft keyboards, and two variations of pie menus. We are motivated by the need for empirical data to guide designers of pen-based systems supporting text and numeric entry.

1.1. HANDWRITING RECOGNITION

Pen-based input is touted as an "empowering" technology, capable of reaching those who shun computer technology. Just as we write, draw and scribble in a notepad, pen-based computers support multi-faceted, natural input. Input recognition can occur at many levels, from simple strokes to cursive handwriting. Recognition "engines" come in many flavours. Some are limited to block-printed characters. Others accept mixed printed and cursive script, boosting performance through context, dictionaries, constrained symbol sets, user profiles, and training.

Ideally, the performance of a recognizer matches or exceeds the performance of users. That is, a perfect recognizer accepts and interprets natural handwriting at a rate controlled by the user. However, as Halfhill (1993) notes, "it'll be a long time before handwriting recognizers are as good as pharmacists at interpreting anybody's sloppy scrawl".

The human performance limits in handwriting are in the range of 12 wpm for block printing (Devoe, 1967) to 33 wpm for fast, cursive writing (Wilkund & Dumas, 1987). A recognizer should accept and recognize input at these rates, with negligible latency. The metric "wpm", for words per minute, is calculated as the entry rate in characters per minute, divided by five. That is, a "word" is five characters, including letters, punctuation and spaces.

Accuracy is a separate issue. In Gibbs' (1993) survey of 13 recognizers from seven vendors, seven quoted "walk-up" accuracy of 92% for character-level recognition. Two cited rates of 85% and 90%. The remaining four cited rates of 85–90% for word-level recognition assisted by a standard dictionary. Accuracy should improve if adaptive techniques are employed; however, walk-up usability is extremely important if pen technology is to entice naive users. Apple acknowledges walk-up error rates as high as 40–50% with their *MessagePad,* for example (Cassleman, 1993). Training a system to recognize an individual's handwriting is possible, but brings other problems, such as the constant need for back-up user profiles on a host system. Gibbs (1993) also notes: "there is no accepted standard for evaluating accuracy. Each vendor assesses their own accuracy as they please" (p. 31).

1.2. SOFT KEYBOARDS

Text entry through key selection on a software-generated keyboard displayed on a CRT is not new. Soft keyboards are common, for example, in current graphical user interfaces (e.g. the calculator in the Apple Macintosh's Control Panel). For pen-based computers, the idea is somewhat analogous to hunt-and-peck typing using one finger and one hand, the difference being that a pen (stylus) is tapped on the surface of an LCD display/digitizer. Empirical tests of typing speeds and error rates for pen-tapping are lacking, however.

A related input scheme is a touchscreen keyboard with text entered using fingers. For touch-entry, Gould, Greene, Boies, Meluson and Rasamny (1990) reported typing rates of 12 wpm; Wilkund and Dumas (1987) found speeds of 14–18 wpm with error rates under 1%. Sears, Revis, Swatski, Crittenden and Shneiderman (1993) varied the size of the touchscreen keyboard and found rates of 9.9–20.3 wpm for novices and rates of 21.1–32.5 wpm for experienced users. In that latter study, subjects used multiple fingers with both hands, so the results are not generalizable to pen-based input.

In another experiment, Sears (1991) found rates of 25.4 wpm for finger typing and 17.1 wpm for mouse typing using the same visual keyboard. This suggests that the directness of touch entry with a finger (and, hence, a stylus) is preferred over spatially transformed input using a mouse.

We should note that a disadvantage of pen-tapping is the lack of kinesthetic feedback and the inability to use a "home row" as a tactile reference point (Wilkund & Dumas, 1987). Hence, a visual connection must be maintained during entry. Many handwriting recognition systems allow input to occur anywhere on the digitizing surface, with text automatically arriving at the application's "insertion point". Depending on the application, this may be critical in choosing the mode of input. As an example, consider a border-crossing guard entering license plate numbers into a pen-based computer. An entry mode that requires on-screen eye fixation would be a serious impediment.

1.3. GESTURES, PIE MENUS AND CLOCK STROKES

Since keypad entry forces eye fixation on the screen and handwriting entry suffers from accuracy problems, the investigation of other entry methods is warranted. For the limited case of numeric entry, we consider another possible input scheme.

One of the most alluring claims for pen-based computing is that natural gestures can form the core repertoire of interaction techniques. Numerous studies have shown the tremendous potential of the pen as a gestural input device (e.g. Buxton, 1986; Wolf & Morrel-Samuels, 1987; Hardock, 1991; Kurtenbach & Buxton, 1991). Although researchers hasten to point out that alphanumeric symbols are germane to pen-based input, we would acknowledge that these are culturally biased and are learned only with considerable practice.

Goldberg and Richardson (1993) presented a system called "unistrokes" for text entry. Their system consisted of symbols that map one-to-one to the regular alphabet. The symbols consisted of a single stroke to make them fast to write and easy to recognize. They reported an average writing rate of 2.8 letters per second
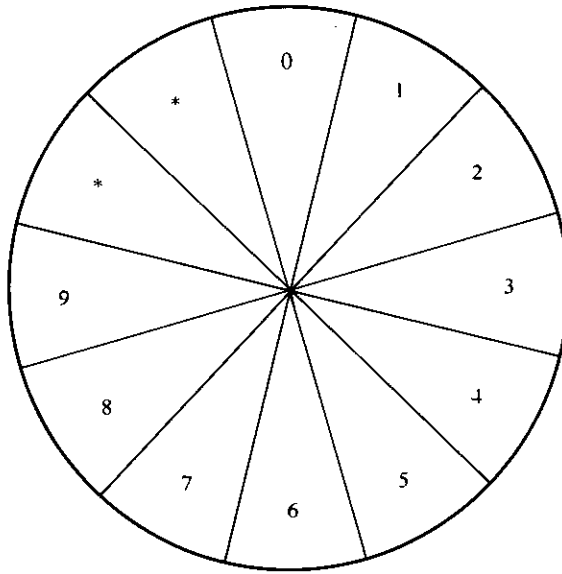
FIGURE 1. The clock metaphor for numeric entry.

(33.6 wpm); however, only three subjects were tested. Perhaps a similar yet simpler entry method would be effective for numeric entry.

Callahan, Hopkins, Weiser and Shneiderman (1988) investigated pie menus for selection. They showed that an appropriate organization within a pie menu improves performance. Kurtenbach, Sellen and Buxton (1993) found that increasing the number of slices in the pie monotonically increased response time, except for pie menus that contained 12 items, similar to a "clock face". One of the authors' suggestions was that the clock metaphor may have reduced visual search time, thus reducing overall response time.

Combining unistrokes and pie menus with a clock metaphor yields a new single-stroke numeric entry method—clock strokes. That is, the user enters a number by stroking from an arbitrary starting point toward where that number occurs on the face of a clock. This is illustrated in Figure 1. A "3" is entered as a stroke to the right, a "6" as a downward stroke, and so on. Note that 0 is at the 12 o'clock position and the 10 and 11 o'clock positions are not used. The advantages are single-stroke entry, a strong metaphor to minimize learning, and the scripting of strokes anywhere at any size.

We investigated clock strokes in two ways. One is to stroke where the digit is entered, similar to handwriting. We call this the *moving pie menu*. This requires on-screen eye fixation. To avoid this, we devised another technique in which the user performs the strokes on a stroking pad with all strokes made on top of one another. We call this a *pie pad*. The resulting digit is automatically sent to the application's insertion point. This method is what Goldberg and Richardson (1993) call "heads-up writing". The advantages of a pie pad are a reduced writing area, eyes-free operation, and less wrist fatigue.

One distinction between the moving pie menu and the pie pad lies in the technique for correcting errors. For the moving pie menu, strokes are made directly

in the application, so a correction requires only a short movement to the point of error, wherein the correct stroke is made. For the pie pad, a mechanism is required to move the cursor or re-establish the insertion point at the point of error. This would most likely entail a larger movement from the pie pad to the application and back to the pie pad.

In the following sections, we describe two experiments that explore human performance in pen-based text and numeric entry tasks using handwriting, soft keyboards, a moving pie menu, and a pie pad. Performance is measured by the speed and accuracy of digit entry.

2. Experiment 1: numeric entry

2.1. METHOD

Subjects
Three female and 13 male volunteer subjects were used in the study. All were university students who used computers on a daily basis.

Apparatus
Software to run the experiment was developed in C using Microsoft's *Pen Windows*. Microsoft's handwriting recognition software (included with *Pen Windows*) was used and was configured to recognize the digit symbols only.

Hardware for the experiment consisted of a 50 MHz PC-486 with a Wacom *PL-100V* tablet for pen entry. The *PL-100V* is both a digitizer for user entry and a 640 × 480 LCD screen. Using the combination of the tablet and host computer allowed the experiment to run without system lag and allowed user entry to also appear on a regular VGA monitor. The monitor was tilted to prevent subjects from seeing it. Digits produced for user entry were generated using the internal random number generator provided by the C compiler.

The clock stroke conditions accepted the digits one to nine as strokes in the same direction as on the face of a clock. Digit zero was assigned the 12 o'clock direction. If the user stroked in the 10 or 11 o'clock directions, it was recorded as not recognized. Each digit had a quantization space of 15° on either side of its ideal angle (0° for "0", 30° for "1", and so on).

Procedure
The task consisted of entering digits provided by the software using one of four conditions. The conditions were (a) handwriting, (b) soft keypad, (c) moving pie, and (d) pie pad, as illustrated in Figure 2. Digits were presented randomly in groups of five. A group of five was called a sequence. Ten sequences made up a block.

Subjects performed all four conditions over two sessions of about 50 min each. The sessions took place over two days. Conditions were counterbalanced using a Latin square to minimize transfer effects.

Subjects were instructed to aim for both speed and accuracy when entering the digits. If a mistake was made, subjects were instructed to ignore it and continue with the sequence. The tablet was set flat on the table and subjects were told to rest their
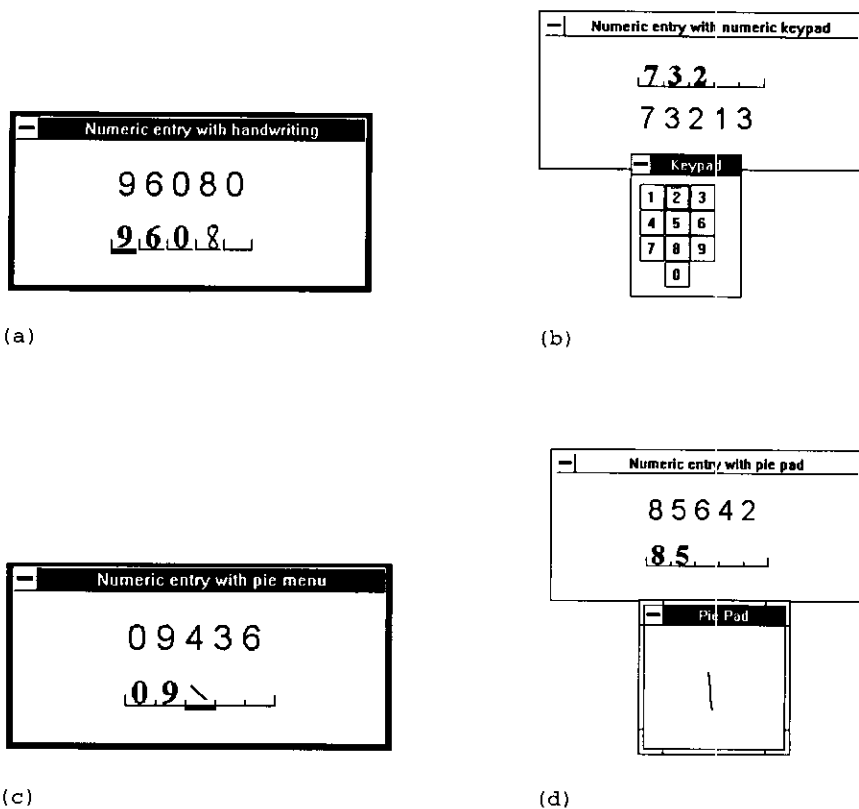
FIGURE 2. The four experimental conditions were (*a*) handwriting in space provided, (*b*) keypad, (*c*) moving pie menu stroking in space provided, and (*d*) pie pad.

hand on the tablet so tablet positioning would be consistent across subjects. In the case of the two clock stroke conditions, the idea was explained to subjects by showing them a picture of where the numbers were located in the pie. While stroking, the pie was not displayed.

The keypad and pie pad were windows that could be re-positioned (dragged). Subjects were allowed to relocate these to any location on the screen. This was particularly important for the left-hand subjects.

Execution of a condition consisted of a brief practice session of 10–15 sequences and then 20 min of recorded entry. To help motivate subjects, summary data for accuracy and speed were displayed at the end of each block. Typically, subjects completed 15–20 blocks for a total of 750–1000 digits within the allocated 20 min. A feedback clock was produced upon the recording of a digit.

Two timing values were recorded for each digit: preparation time and scripting time. Preparation time is the time from the end of the previous digit to the start of the current digit. Scripting time is the amount of time that the pen is in contact with the tablet while forming the digit or pie stroke. Preparation time plus scripting time equalled the total entry time for a digit.

Note that scripting time with the soft keypad is virtually zero due to the brevity of the tap. Note also that the timing value for the first digit in a sequence is

meaningless because there is no start time to reference from. Thus, the data for the first digit in a sequence were not used for the summary statistics.

2.2. RESULTS AND DISCUSSION

Condition effects

Data were summarized for the first, middle and last three blocks for each condition. The data entered in the analysis of variance were from the last three blocks only, to minimize learning effects. This data set contained 120 digits per subject for 16 subjects, which totals 1920 digits for each condition.

There was a significant main effect for condition on entry time ($F(3,45) = 59.3$, $p < 0.0001$) and error rate ($F(3,45) = 22.1$, $p < 0.0001$). The mean values for each condition are illustrated in Figure 3. To facilitate comparisons with other studies, entry speed is shown in words per minute (wpm) in the figure. Note that the keypad is superior to the other methods for both accuracy and entry time. At 30.4 wpm, keypad tapping is comparable to Goldberg and Richardson's (1993) unistrokes or to Sears et al.'s (1993) full-size touchscreen keyboard.

The error rate reported herein for handwriting (10.4%) was higher than the 8% walk-up error rate cited for Microsoft's recognizer in Gibbs' (1993) survey. The poor performance we observed would undoubtedly worsen with an unconstrained symbol set. This is a serious problem with handwriting recognition, and one that is played-out regularly in the popular press. Handwriting recognition has been called
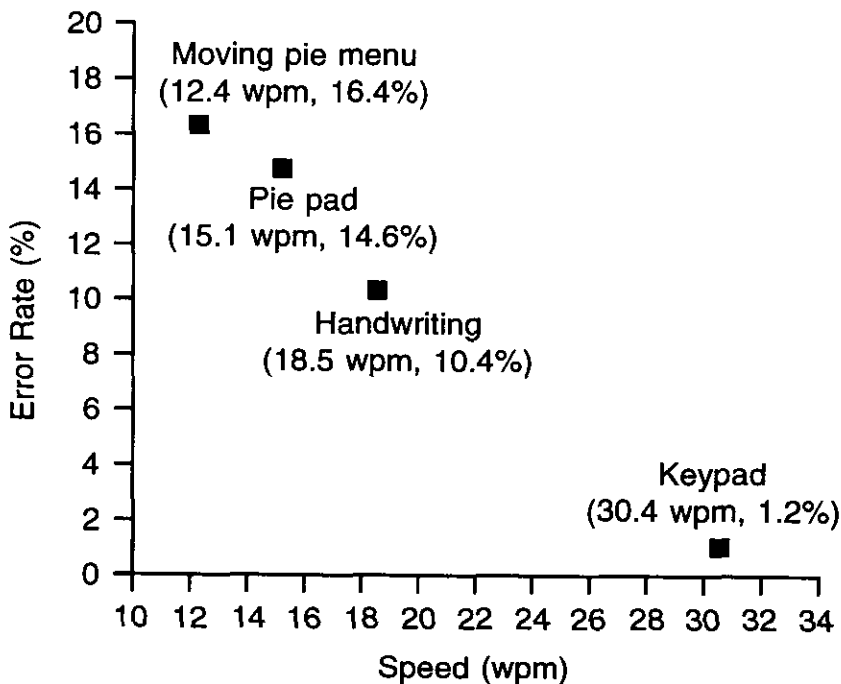

FIGURE 3. Comparison of the four conditions for error rates and entry time.

"impractical" (Eglowstein, 1993), "marginal" (Halfhill, 1993), and "disappointing" (Caruthers, 1993).

Learning effects

The improvement of speed and accuracy across sessions was investigated. Although there was no effect across blocks for accuracy ($F(2,30) = 1.5$), there was a significant main effect across blocks for entry time ($F(2,30) = 14.2$, $p < 0.0001$). There was no noticeable improvement in accuracy for any of the conditions over the session. This is consistent with Bailey's (1989) observation that "in activities were performance is primarily automatic the proportion of errors will remain fairly constant, but the speed with which the activity is performed will increase with practice" (p. 101).

The improvement in speed over the sessions relative to the initial performance was in the following order: keypad (14.1%), pie pad (11.3%), moving pie menu (6.2%), and handwriting (3.7%). The keypad condition was the easiest condition to learn, so a large decrease in entry time is expected. Handwriting is the most natural task, since subjects import substantial skill; thus entry time reduction was not great. The improvements observed in the pie menu conditions would likely continue since these conditions are the most unfamiliar.

Performance by digit

Since the entry technique varies across digits, analyses on a per digit basis are warranted. For handwriting, the digits 3 and 8 accounted for 43.4% of the errors. We attribute this to the recognizer, since the strokes for "3" and "8" are very similar. Indeed, our test of handwriting accuracy is more a test of the recognition ability of the particular product than a test of handwriting accuracy per se. Other recognizers would produce different results and merit investigation on their own.

With the keypad condition, 0 and 5 had the highest error rates: 1.93% and 2.41% respectively compared with a mean of 1.2%. As expected for the clock stroke conditions, off-angle strokes accounted for the majority of errors, while on-axis strokes (0, 3, 6, 9) exhibited the lowest error rates. This effect is shown in Figure 4.

Preparation vs. scripting time

Although the total time to enter a digit was greater for the pie menu conditions than for handwriting, a comparison of preparation and scripting time reveals the potential of pie menus (see Table 1).

Preparation and total time were less for handwriting than for the pie conditions, while the stroking time during handwriting was greater by about 50%. Hence, the pie conditions would be faster than handwriting if preparation time was reduced. Preparation time is higher with the pie tasks because they require a conscious act. Handwriting, on the other hand, is a highly learned motor skill. In fact, subjects commented that the pie methods were fatiguing and demanded a lot of concentration. If practiced over a number of days, the pie menu tasks would become more automatic with a reduced cognitive load. This, no doubt, would reduce preparation time.
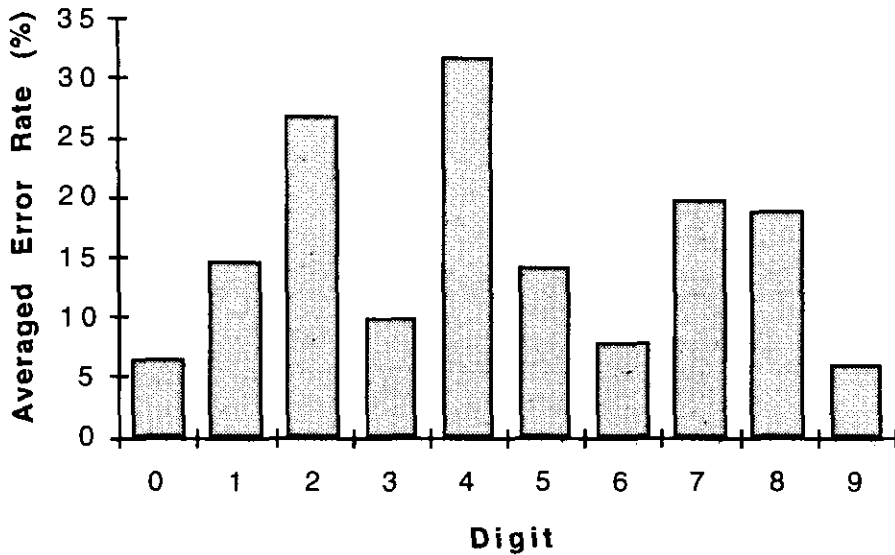
FIGURE 4. Error rates by digit for the pie conditions. The lowest error rates were for the on-axis digits: 0, 3, 6 and 9.

Condition preference

Subjects were asked to rate each condition on a one to five scale. The results are listed in Table 2. The keypad received the highest preference ratings, followed closely by handwriting. The pie pad was preferred over the moving pie.

Many of the subjects realized that the keypad was a better entry method yet they

TABLE 1

Comparison of preparation and scripting times (ms) for experiment 1

Condition	Preparation time	Scripting time	Total
Handwriting	334	315	649
Pie pad	569	226	794
Moving pie	763	207	970

TABLE 2

Subject preferences (frequency) for experiment 1

Condition	Rating†				
	5	4	3	2	1
Handwriting	1	11	4	0	0
Keypad	4	8	4	0	0
Moving pie	0	1	3	7	5
Pie pad	0	2	3	8	3

† 1 = least preferred, 5 = most preferred.

preferred handwriting. They felt it was more natural, easy, and directly corresponded to the task. The implication of this is that if a new user is given the choice of handwriting or keypad entry, the user may choose handwriting even though it is less efficient.

Improving the pie menu
Often a subject's strokes were in the general direction of the number but the subject would curve the line either at the start or beginning. Since the digit assignment is generated by the angle formed from the start and end points, the digit could be quantized incorrectly if the trajectory curved into the wrong "slice". Other ways of determining the desired digit, such as using points 10% of the total distance from each end of the line, might improve the accuracy of the clock stroke methods.

Further research into pie divisions could improve the performance of the pie menu methods. For instance, the digits that have higher error rates could be assigned a larger slice of the pie. Investigation into using 10 instead of 12 sections could be done. This would eliminate the clock metaphor and reduce the method to a pie menu selection. It was found that the mean directions for the digits was not directly at the allocated angles (e.g. the mean stroking angle for 0 was at 5.6 degrees rather than 0 degrees). Perhaps using the actual mean angles as the centre of the slice would improve the accuracy. Also, digits with higher standard deviation values for stroking angle could be allocated larger slices of the pie.

3. Experiment 2: text entry

3.1. METHOD

Subjects
Four female and 11 male volunteer subjects were used in the study. All were university staff or students who used computers on a daily basis.

Apparatus
The equipment was the same as for experiment 1.

Procedure
The task consisted of entering characters provided by the software using one of the three methods. The conditions were (a) hand printing, (b) soft keyboard with a QWERTY layout, and (c) soft keyboard with an ABC layout, as illustrated in Figure 5. No training was provided for the recognizer, as we were interested in the walk-up acceptance and performance for pen-based computers. Phrases containing 22 characters (4 words and 3 blanks) were randomly presented in blocks of three. For the purpose of calculating entry rates, a phrase constitues 4.4 words (1 word = 5 characters, including spaces). The single letter frequency count table of Mayzuer and Tresselt (1965: p. 14) was used to create a character-balanced phrase set. Nine blocks were used for each condition for a total of 594 characters (including blanks). Conditions were counterbalanced using a Latin square to minimize transfer effects.
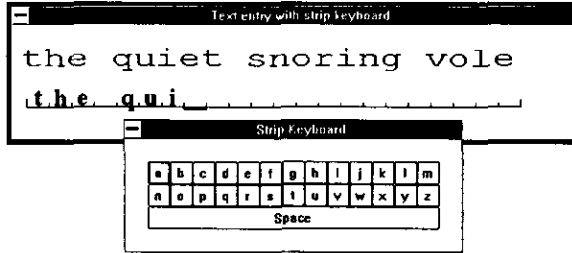
All three conditions were tested in a 1 h session. Subjects were instructed to
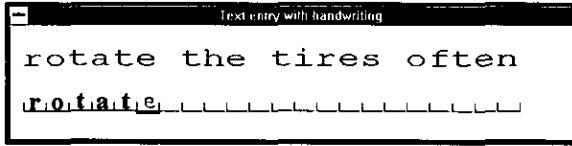
(a)
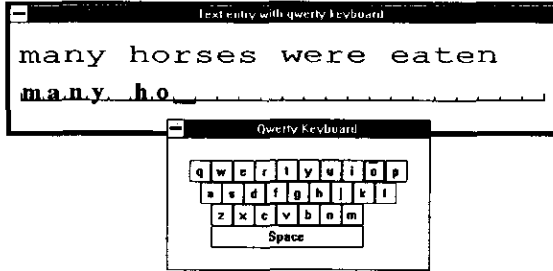
(b)

(c)

FIGURE 5. The three experimental conditions were (a) hand printing, (b) QWERTY tapping and (c) ABC-tapping.

aim for both speed and accuracy when entering the characters. They were also told to ignore mistakes and continue with the rest of the sequence. The tablet was set flat on the table or propped slightly at the back, as was preferred by several subjects.

Execution of a condition consisted of a brief practice session of three phrases and then nine blocks of recorded entry (27 phrases). Subjects memorized and spoke aloud each phrase before entering the text. To help motivate subjects, summary data for accuracy and speed were displayed at the end of each block. A feedback click was produced upon the recording of a character.

For each character, the time from the completion of the previous character to the completion of the current character was recorded. The timing value for the first character in a sequence is meaningless as there is no start time from which to reference. Thus, the data for the first character in a phrase were not used for the summary statistics.
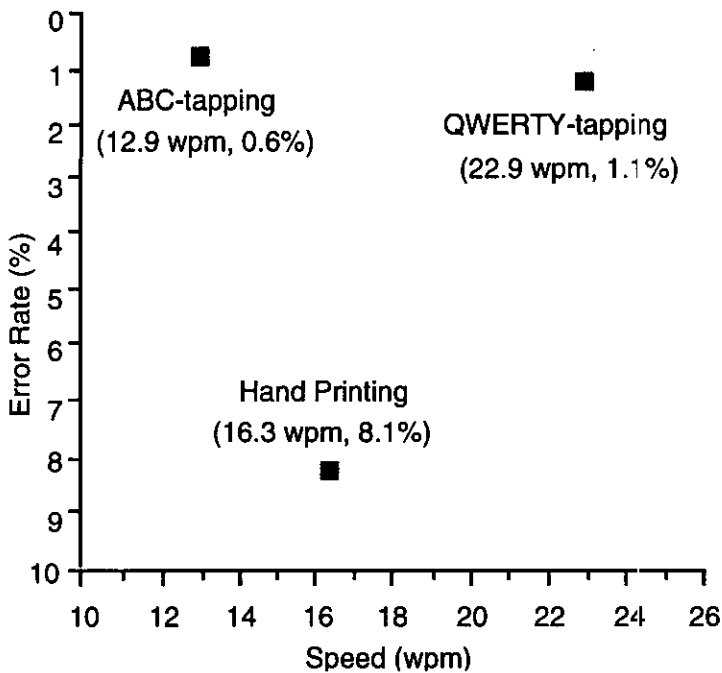
FIGURE 6. Comparison of the three conditions for error rates and entry speed.

3.2. RESULTS AND DISCUSSION

For each condition, data were summarized on a per block basis. The data entered in the analysis of variance were from all blocks. For each condition, the data contained at least 400 characters per subject for each of the 15 subjects.

There was a significant main effect for condition on users' entry time ($F(2,28) = 95.6$, $p < 0.0001$) and error rate ($F(2,28) = 33.6$, $p < 0.0001$). The mean values for each condition are shown in Figure 6. Entry times were converted to words per minute (wpm) for comparison with other studies.

Accuracy for hand printing was more highly varied than the other two conditions, as seen in Figure 7. For hand printing, three subjects had error rates less than 5.0%, while 3 had error rates greater than 15.0%. One subject achieved an error rate for hand printing (3.0%) that bettered the performance of another subject using QWERTY-tapping (3.1%).

Learning
Although there was no effect across blocks for accuracy ($F(8,112) = 74$), there was significant main effect across blocks for entry time ($F(8,112) = 12.9$, $p < 0.0001$). Apparently, subjects did not improve their accuracy with practice, however, they did get faster, as seen in Figure 8. This is consistent with our results from experiment 1.

The QWERTY-tapping and hand printing conditions exhibited the greatest improvement in absolute speed over the nine blocks (increase of 2.8 wpm and 2.9 wpm respectively), however, hand printing had the highest rate of improvement
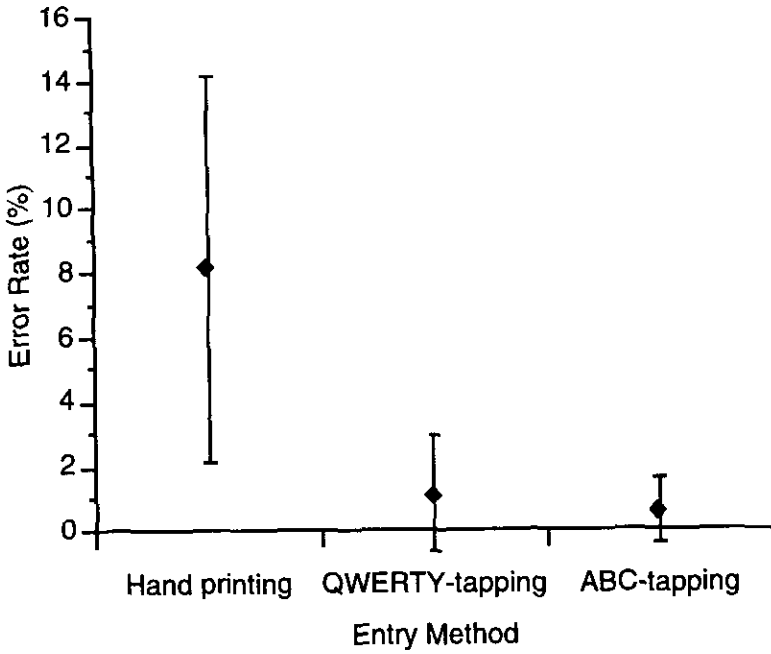
FIGURE 7. Error rate for each condition with standard deviation error bars.
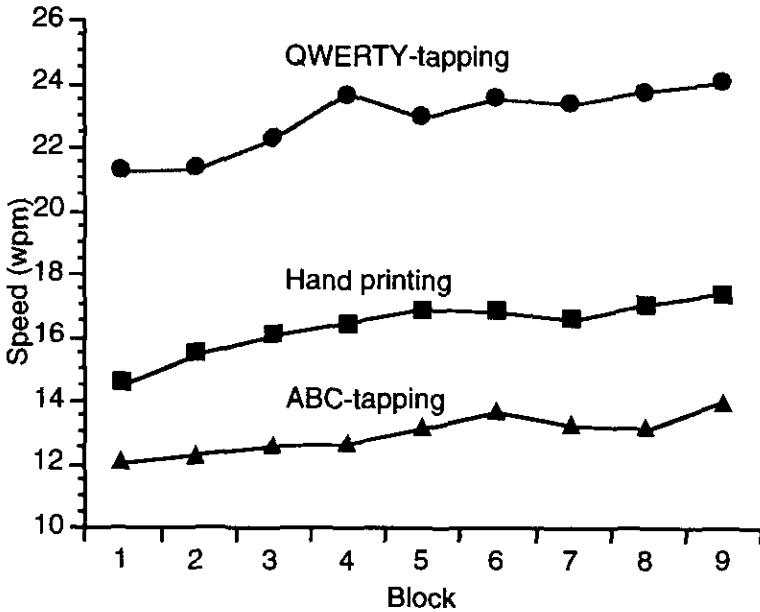

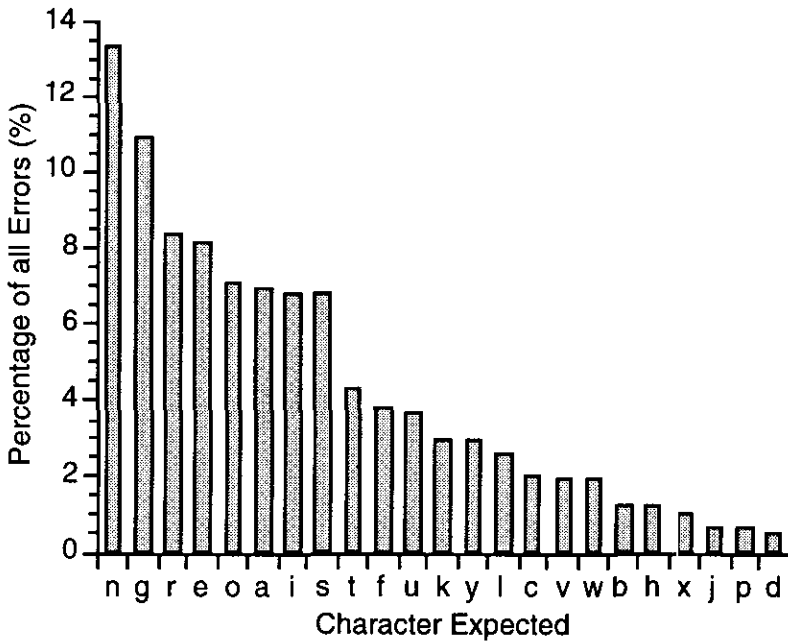FIGURE 8. Learning as increased entry speed.

FIGURE 9. Characters expected by the subjects that were posted as some other character.

(19.9%) due to the lower initial value. The ABC-tapping condition improved the least over the nine blocks (2.0 wpm for a 16.3% increase in speed).

Error rates by character
Error types were examined for each condition. For the QWERTY keyboard, 49% of the errors occurred when subjects tapped a key directly adjacent to the target key. This value rose to 60% for the ABC keyboard.

For the hand printing condition, errors were examined for the frequency of the characters involved. The most frequently misinterpreted character was the letter "n" (13.4% of all errors, as shown in Figure 9). For each character, there are 25 possible mis-interpretations, As shown in Figure 10, the Microsoft recognizer posted the letter "c" most frequently when a recognition error was made (35.9% of all errors).

In all, there were 81 unique error pairs (character expected, character posted). Table 3 shows the 10 most frequent error pairs—characters printed by the subjects (characters expected) are shown in conjunction with the characters posted by the Microsoft recognizer. The letter "n" appears twice in the character expected row, and the letter "c" appears five times in the character posted row.

Preferences
Subjects were asked to rate each condition in order of preference. The results are listed in Table 4. Hand printing and QWERTY-tapping received equally high first
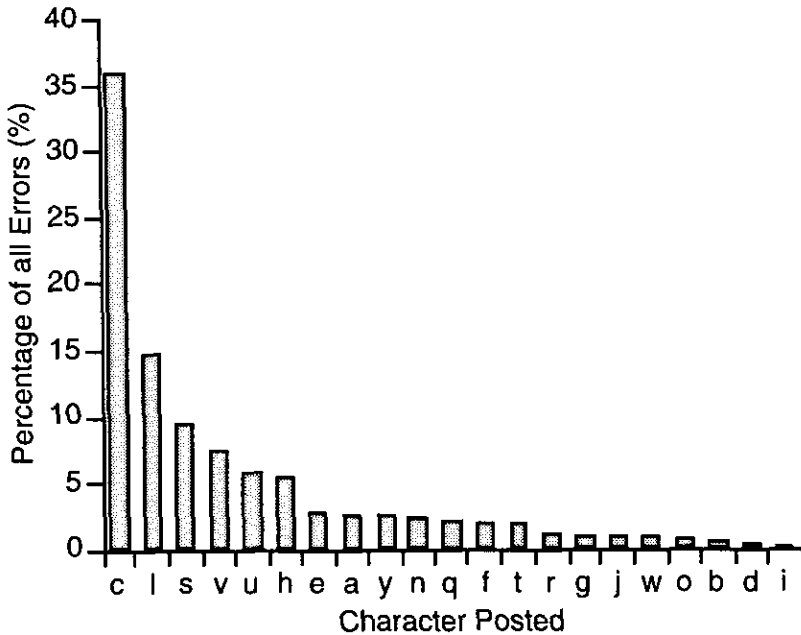
FIGURE 10. Characters posted by the recognizer in error.

choice ratings, each being preferred by seven subjects, while the ABC-tapping was the least preferred, with 12 of the 15 subjects rating it third. However, QWERTY-tapping received a greater number of second choice ratings than did hand printing (8 vs. 5).

TABLE 3

The 10 most frequent translation errors for the hand printing condition in experiment 2

| Character expected | g | a | r | i | n | e | o | s | e | n |
Character posted	s	c	v	l	c	c	c	c	l	h
Proportion of total errors (%)	6.7	5.9	5.9	5.4	5.2	5.0	4.9	4.3	3.0	2.8

TABLE 4

Subject preferences (frequency) for experiment 2

Condition	Rating		
	First	Second	Third
Hand printing	7	5	3
QWERTY-tapping	7	8	0
ABC-tapping	1	2	12

Keyboard layouts

Changing the layout of the keyboard from QWERTY to ABC significantly lowered the entry rate, due to the subjects' unfamiliarity with the ABC keyboard layout. Subjects indicated that they would be able to achieve high entry rates using ABC-tapping given sufficient practice. Also, suggestions were made to improve the performance of ABC-tapping: placing the characters in a 5 × 6 matrix rather than a 2 × 13 matrix, and putting them in one long row or column. Subjects believed that the 5 × 6 matrix would provide a smaller visual scanning area and that the linear keyboard would reduce the confusion caused by the arbitrary break in the ABC keyboard.

None of the users felt that the keys were too small or too large, even though most errors were in hitting adjacent keys. This suggests that the relatively low error rate is balanced by the ease with which the keyboard is tapped; that is, the wrist did not need to be lifted as it would with larger keys. The low error rate for ABC-tapping is likely related to an unfamiliar layout requiring conscious effort. Given enough practice, the slightly lower error rate for ABC-tapping may rise to that of the more familiar QWERTY-tapping.

Accuracy

In contrast, hand printing was significantly slower and more error prone than QWERTY-tapping. The error rate reported here is for a restricted character set (lower case letters) with a similarly restricted character recognizer. This rate is similar to that quoted for unconstrained recognizers (Gibbs, 1993). Given a full character set and an unconstrained recognizer, error rates would be even higher. The observed 8% error rate would be an unacceptable error rate for an optical character recognizer—it is unlikely that users of pen-based systems would be satisfied with even higher error rates.

Subject effects

During the course of a session, subjects tried to adapt to the recognizer, although this adaptation was not significant in terms of error reduction. Adaptation by users with practice, training of the recognizer, user profiles, and aids such as dictionaries will reduce the error rate. It remains to be seen whether walk-up users can be enticed with such poor initial performance.

The character error frequencies indicate that certain letters are more problematic than others. These patterns of misinterpretation could be used to fine tune a character recognizer.

Several subjects had hand printing error rates approaching that of QWERTY-tapping, while others had substantially higher rates. This suggests that some users will have less difficulty with hand printing entry. It also suggests that alternatives, such as QWERTY-tapping, will be the preferred entry method for a group of users. This is substantiated by the nearly equal split in subject preference between hand printing and QWERTY-tapping.

Several subjects commented on the pen skating across the display. Current display technology still lacks the feel of pen and paper. It was also observed that a few of the right-handed subjects cramped their hand as they approached the right side of the display. This cramping resulted in distorted printing, which in turn caused errors

in recognition. Given that pen-based computing is also portable computing, the "comfortable" environment of this experiment is atypical. Vibration, cramped space and unsteady surfaces are but a few environmental factors which will degrade performance for all entry methods.

4. Conclusion

The two experiments demonstrate overwhelmingly that tapping on a soft keyboard is a fast and accurate entry method for pen-based computers. Entry speed is in the range 22–30 wpm with error rates around 1%. Handwriting is slower (16–18 wpm) and, due to limitations in the recognition software, more error prone. Error rates with the Microsoft recognizer were measured at 8–10%, even though the recognizer was constrained only to interpret 10 numeric symbols (experiment 1) or 26 lower case alpha symbols (experiment 2).

For numeric entry, the two pie menu conditions performed worse than hand printing or tapping on a soft keypad. Although the pie strokes were entered more quickly than hand-printed numbers, the pie conditions required more mental preparation time. With enough practice, pie menus would probably become a fast and accurate entry method. This is a moot point, however, since ease-of-use for the novice is an important requirement for pen-based computers.

For text entry using a soft keyboard, the QWERTY layout performed better than an alternate layout using two rows of sequential letters. Although the latter is effective in conserving screen layout, the venerable QWERTY layout remains the logical choice for alphanumeric soft keyboards.

References

BAILEY, R. W. (1989). *Human performance engineering*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

BUXTON, W. (1986). Chunking and phrasing and the design of human–computer dialogues. In H.-J. Kugler, Ed. *Proceedings of the IFIP 10th World Computer Conference—Information Processing '86*, pp. 475–480. Amsterdam: Elsevier.

CALLAHAN, J., HOPKINS, D., WEISER, M. & SHNEIDERMAN, B. (1988). An empirical comparison of pie vs. linear menus. *Proceedings of the CHI '88 Conference on Human Factors in Computing Systems*, pp. 95–100. New York: ACM.

CARUTHERS, F. (1993). Pen input drives new class of personal computers. *Computer Design*, March, 12–16. (OEM supplement.)

CASSLEMAN, G. (1993). Newton here but in short supply. *Computing Canada*, 16 August, 1, 4.

DEVOE, D. B. (1967). Alternative to handprinting in the manual entry of data. *IEEE Transactions on Human Factors in Electronics*, **HFE-8**, 21–32.

EGLOWSTEIN, H. (1993). Applying the power of the pen. *Byte*, July, 132–140.

GIBBS, M. (1993). Handwriting recognition: A comprehensive comparison. *Pen*, March/April, 31–35.

GOLDBERG, D. & RICHARDSON, D. (1993). Touch-typing with a stylus. *Proceedings of the*

INTERCHI'93 Conference on Human Factors in Computing Systems, pp. 80–87. New York: ACM.

GOULD, J., GREENE, S., BOIES, S., MELUSON, A. & RASAMNY, M. (1990). Using a touchscreen for simple tasks. *Interacting with Computers,* **2,** 59–74.

HALFHILL, T. R. (1993). PDAs arrive but aren't quite here yet. *Byte,* October, 66–86.

HARDOCK, G. (1991). Design issues for line-driven text editing/annotation systems. *Proceedings of Graphics Interface '91,* pp. 77–84. Toronto: Canadian Information Processing Society.

KURTENBACH, G. & BUXTON, B. (1991). GEdit: a testbed for editing by contiguous gestures. *SIGCHI Bulletin,* **23**(2), 22–26.

KURTENBACH, G., SELLEN, A. & BUXTON, B. (1993). An empirical evaluation of some articulatory and cognitive aspects of marking menus. *Human–Computer Interaction,* **8,** 1–23.

MAYZUER, M. S. & TRESSELT, M. E. (1965). Tables of single-letter and diagram frequency counts for various word-length letter-position combinations. *Psychonomic Monograph Supplements,* **1**(2), 13–32.

SEARS, A. (1991). Improving touchscreen keyboards: design issues and a comparison with other devices. *Interacting with Computers,* **3,** 252–269.

SEARS, A., REVIS, D., SWATSKI, J., CRITTENDEN, R. & SHNEIDERMAN, B. (1993). Investigation touchscreen typing: the effect of keyboard size on typing speed. *Behaviour & Information Technology,* **12,** 17–22.

WILKUND, M. E. & DUMAS, J. S. (1987). Optimizing a portable terminal, keyboard for combined one-handed and two-handed use. *Proceedings of the 31st Annual Meeting of the Human Factors Society,* pp. 585–589. Santa Monica, CA: Human Factors Society.

WOLF, C. G. & MORREL-SAMUELS, P. (1987). The use of hand-gestures for text-editing. *International Journal of Man–Machine Studies,* **27,** 91–102.

Paper accepted for publication by Associate Editor Professor P. Barker.