INPUT RATES AND USER PREFERENCE FOR THREE SMALL-SCREEN INPUT METHODS: STANDARD KEYBOARD, PREDICTIVE KEYBOARD, AND HANDWRITING

James R. Lewis
International Business Machines Corp.
West Palm Beach, Florida

The purpose of this study was to compare three input methods (standard keyboard, predictive keyboard, and perfect handwriting recognition) considering user input rates and preferences. Six participants used all three input methods to enter both normal text and addresses. Participants indicated that they preferred to use the standard keyboard. The average input rate for handwriting was fastest, but also much more variable than the other methods. Despite its speed, participants generally found it difficult to write comfortably and legibly on the small (35x115 mm) display. The input rate for the standard keyboard was more than twice the input rate for the predictive keyboard. These results suggest that, for small devices, neither handwriting recognition nor predictive keyboards would effectively replace the standard keyboard layout. Even with perfect handwriting recognition, users seem to prefer tapping on a small standard keyboard unless the device's hardware design allows comfortable handwriting input.

INTRODUCTION

As touchscreen interfaces become smaller, emulating keyboard input becomes more difficult. Providing a method for effective data input for small, handheld devices is one of the major usability problems facing personal digital assistants (PDAs) and personal communicators (PCs). For such devices there are a number of methods currently available, including: 1) an on-screen keyboard with a standard layout, 2) an on-screen predictive keyboard, and 3) handwriting recognition. Each of these approaches has its benefits and drawbacks.

The most obvious and easiest solution is an on-screen version of a standard (QWERTY) keyboard layout. Sears (1991) reported that users could type an average of 25 words-per-minute (wpm) on touchscreen keyboards with keys 22.7 mm square, and could type an average of 20 wpm using a keyboard only 70 mm wide (implying keys about 7 mm square). As the size of the keys decreases, typing speed decreases.

A number of studies have shown that the standard layout is superior to alphabetic layouts, regardless of user experience with standard layouts or size of keyboard (Francas, Brown, and Goodman, 1983; Mackenzie et al., 1994; Norman and Fisher, 1982), so any full-keyboard layout should use the standard layout rather than an alphabetic layout.

Another solution to on-screen keyboard emulation is a predictive keyboard -- one that only displays a most-likely subset of the full set of keyboard keys, with predictions based on tables of letter frequencies. The software uses these tables to arrange the letters of the alphabet by most-likely frequency of occurrence, and displays the letters that the user is most likely to type next. If the desired letter is not in the initial set, users can touch an "Other" button to display the next most-likely letters. For small devices, users also need to touch buttons to display numbers and punctuation. The main advantage of a predictive keyboard is that its keys are much larger than the keys for a standard layout, given the same area in which to work. The key disadvantages are 1) the requirement to touch buttons to display characters, numbers, or punctuation not currently visible and 2) that the letters do not always appear in the same position, but instead appear in an order determined by their likelihood of use.

A third approach is handwriting recognition. A fast handwriter can produce about 25 wpm (Bailey, 1989), so handwriting is a feasible alternative to selecting letters from a standard layout, given perfect handwriting recognition. However, current handwriting recognition algorithms do not produce perfect recognition. Consequently, people who use a handwriting recognition system must constantly check the accuracy of recognition, which slows their writing considerably and may lead them to choose to stop using the handwriting recognizer and, instead, use the alternative method of an on-screen standard keyboard layout.

The purpose of this study was to compare these approaches (standard keyboard, predictive keyboard, and simulated perfect handwriting recognition) considering user input rates and preferences. The results of this comparison can help PDA and PC designers decide which type of input method to emphasize in their interfaces.

METHOD

Participants

Six IBM employees participated in the experiment. Two were permanent employees, and the other four were graduate students interning at IBM. Three were male, and two were left-handed. Their reported typing speeds ranged from 25 to 60, and averaged 52.5. Three had used reduced-size keyboards previously, and those same three participants had previous experience using pen-based systems. Four participants had some PDA experience.

Materials

The participants used a Simon™ to input eight sentences with the Simon's built-in standard keyboard (typing), predictive keyboard (typing), or sketch pad (handwriting). The touch-sensitive screen on a Simon is 35 mm x 115 mm. The standard keyboard layout occupies a space 25 mm x 115 mm, with keys 9 mm by 5 mm. The predictive keyboard's area is 35 mm x 50 mm, with typing keys 13 mm x 13 mm. The user interface for Simon's predictive keyboard shows six letters at a time in this space, with a spacebar and basic punctuation always available. The drawing (writing) area of Simon's sketch pad is 33 mm x 83 mm.

Input stimuli were 24 sentences randomly selected from the Brown corpus (a collection of magazine and newspaper articles representative of contemporary English-language usage), and randomly distributed across three sets of sentences. All sentences ranged in length from 90 to 110 characters. Three sets of addresses (each composed of eight addresses) contained addresses randomly selected from the 1994 Human Factors and Ergonomics Society directory. The final three stimulus sets included one sentence set and one address set, with the sentence-address set pairing determined by chance.

The other materials were an instruction sheet, a background questionnaire, a layout rating form for each input method, a keyboard layout attribute importance form, and a layout ranking form.

Procedure

I used a digram-balanced Greco-Latin design (Lewis, 1993) to simultaneously counterbalance the presentation of input method, stimulus set, and the pairing of input methods and stimulus sets. Participants used all three input methods to enter both normal text (sentences) and addresses. They always entered the sentences first. The experimental design was within-subjects, with three independent variables (three types of input method, two types of text to enter, and eight trials per input method and text combination).

Participants completed a background questionnaire, then started the experiment. Participants sat in a chair during the experiment, holding the Simon in one hand and a small stylus in the other, using the stylus to tap on the screen in the keyboard conditions and to write on the screen in the handwriting condition. Participants were required to produce completely accurate text, so input times include the time for both initial text production and error correction. After finishing the sentences and addresses with each input method, they completed a form to rate the input method. After completing all input methods, they completed one form to rate the importance of various layout attributes, and a final form to rank (and discuss their reasons for ranking) the three input methods.

RESULTS

Performance

The results of an analysis of variance on the entry rate (in characters per second, or cps) showed significant main effects for Input Method ($F(2,10)=44.8$, $p=.0001$), Text Type ($F(1,5)=4.9$, $p=.018$), and Trial ($F(7,35)=5.0$, $p=.001$). The Input Method by Text Type interaction was also significant ($F(2,10)=5.2$, $p=.029$), but none of the interactions with Trial were significant.

The results indicated that handwriting was significantly faster than the standard layout, which was significantly faster (in fact, over twice as fast) as the predictive layout (see Table 1 and Figure 1). Even though the interaction of Input Method and Text Type was significant, the interaction was due to differences in degree of difference between input speed for sentences and addresses – there were no crossed effects. Entering addresses was always slower than entering sentences (see Table 2 and Figure 2).

Because learning usually follows an exponential function, one would expect any rapid learning effect to show up as a steep, positively sloped function of Trial. Although the main effect of Trial was significant, the Input Method by Text Type by Trial interaction was not ($F(14,70)=0.9$, $p=.59$), and showed no evidence for a rapid learning effect for any of the input methods or text types (see Figure 3).

Table 1. Mean Throughput (with 95% Confidence Intervals) for Each Input Method in CPS (with WPM conversion)

Input Method	Lower Limit	Mean	Upper Limit
Standard	1.10 (13.2)	1.18 (14.1)	1.25 (15.0)
Predictive	0.49 (5.8)	0.52 (6.2)	0.55 (6.6)
Handwriting	1.34 (16.1)	1.88 (22.6)	2.43 (29.2)

Note: The formula used to convert cps to wpm was
wpm = cps * 60 *sec/min* * 1/5 *word/characters*

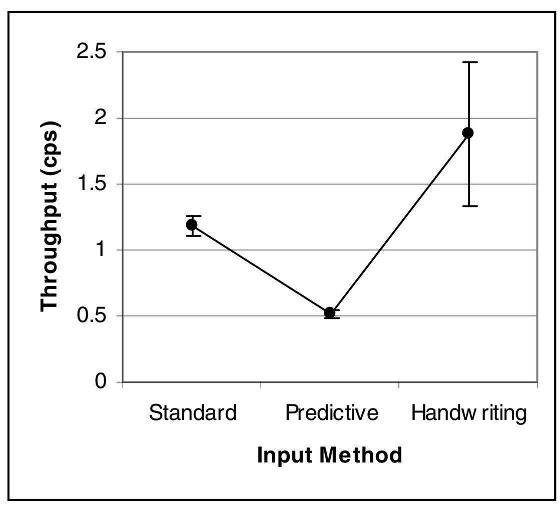

Figure 1. Throughput as a Function of Input Method (CPS) (Error bars represent 95% confidence intervals)

Table 2. The Input Method by Text Type Interaction in CPS (with WPM conversion)

	Type of Text	
Input Method	Sentences	Addresses
Standard	1.42 (17.0)	0.93 (11.2)
Predictive	0.60 (7.2)	0.44 (5.2)
Handwriting	1.97 (23.6)	1.80 (21.7)

Note: The formula used to convert cps to wpm was
wpm = cps * 60 *sec/min* * 1/5 *word/characters*

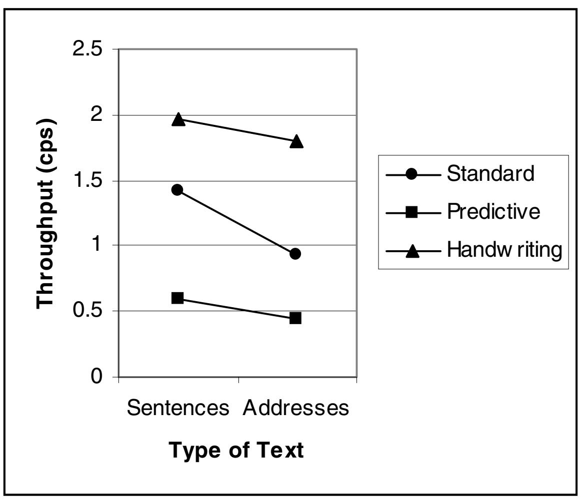

Figure 2. The Input Method by Text Type Interaction

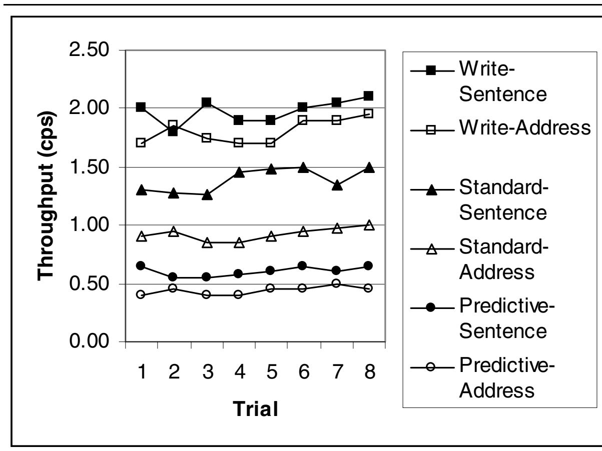

Figure 3. The Input Method by Text Type by Trial Interaction

Ratings

The rating questionnaires used 7-point bipolar scales, with lower ratings better than higher ratings. The items covered the topics of ease of finding letters, ease of rapid input, ease of accurate input, ease of learning letter locations, ease of typing, and acceptability of key layout (with suitable modifications for the questionnaire given after the handwriting session). The average ratings for the input methods (collapsed across items) were significantly different ($F(2,10)=8.2$, $p=.008$). The average rating for the standard layout was 1.5, that for the predictive layout was 3.6, and the handwriting rating was 4.6. Bonferroni *t*-tests indicated that participants provided more favorable ratings for the standard layout than the predictive layout ($t(5)=4.6$, $p=.006$) and handwriting ($t(5)=4.3$, $p=.007$). The difference between the predictive layout and handwriting was not significant. Participants rated all the questionnaire attributes as important, with average importance ratings exceeding 6.0 for all attributes (except acceptability of key layout, which averaged 5.5).

Rankings

The mean ranks for the input methods were 1.33 for the standard layout, 2.17 for the predictive layout, and 2.50 for handwriting (a lower mean rank is better). A Friedman test indicated a marginally significant difference among the mean ranks ($\chi^2(2)=4.4$, $p=.10$). The pattern of ranking results was similar to that of the rating results.

Participant Comments

The main disadvantage of handwriting was the difficulty of handling a small device and writing comfortably. Also, the resolution and parallax of the display made it difficult to dot "i's" and cross "t's". The primary disadvantage cited for the predictive layout was the constantly changing character location. The main advantage cited for the standard layout was its familiarity.

DISCUSSION

Participants indicated (with both rankings and ratings) that they preferred to use the standard keyboard for text entry with Simon. The average input rate for handwriting was fastest, but also much more variable than the other methods. Despite its speed, participants generally found it difficult to write comfortably and legibly on the small display. The input rate for the standard keyboard was more than twice the input rate for the predictive keyboard.

These results suggest that, for small devices such as the Simon, neither handwriting recognition nor predictive keyboards can effectively replace the standard keyboard layout. Even with perfect handwriting recognition, users will prefer to tap on a small standard keyboard unless the device's hardware design allows comfortable handwriting input with near perfect accuracy. Without perfect handwriting recognition accuracy, handwriting is likely to provide slower throughput than tapping a soft standard keyboard (Mackenzie et al., 1994).

Experience in providing product demonstrations of Simon has shown that the predictive layout has a strong initial appeal, so it is reasonable to continue to include it as an option in small devices. However, the predictive keyboard is markedly inferior to the standard layout for both sentence and address entry. It also appears that the problems that participants experience using the predictive keyboard prevent it from being an input method that could benefit from practice. Uncertainty concerning letter location and the need to touch buttons to display numbers and punctuation cause user performance to be very slow with this input method. This is similar to the usability problems reported by Koester and Levine (1994) for word prediction interfaces. Although word prediction reduced the number of required keystrokes, "the cognitive cost of using word prediction had a major impact on the performance of these subjects" (Koester and Levine, 1994, p. 177). The current study showed a similar cognitive cost associated with using letter prediction.

A problem with experiments comparing the throughput for different input devices is that it is impossible to prove that the pattern of results observed in a specific experiment generalizes to all possible designs of the input methods under investigation. This is particularly true if previous work has not demonstrated that the devices' designs are optimal, as is the case in this experiment. We compared the input methods given on a particular commercially available device, but it is likely that it is possible to make these specific input methods better. For example, it should be possible to improve the throughput for the predictive keyboard by providing eight rather than six letters at a time (Lewis, Allard, and Hudson, 1999). It might be possible to improve user acceptance of handwriting by providing a better physical interface for the support of comfortable handwriting (although less than perfect handwriting recognition will probably continue to be a problem for the foreseeable future).

Despite potential improvements in predictive keyboards, user performance with Simon's predictive keyboard was not encouraging, with measured throughput less than half that of the on-screen standard keyboard. However, in addition to other limits to generalization of the current study, the results might be specific to the amount of screen real estate available on a Simon. Once the screen real estate of a device becomes too small to support a standard keyboard layout, the predictive keyboard could become the input device of choice.

Also, the devices examined in this experiment do not include all possible small-screen input methods. For example, it is possible to develop alternative layouts based on English-language digram analysis (Lewis, Kennedy, and LaLomia, 1999; Lewis, LaLomia, and Kennedy, 1999). A digram-based layout takes advantage of the statistical characteristics of English letter distributions, but should be more learnable than a predictive layout because the letters are always in the same position. Also, after the development of Simon, a number of newer handwriting methods (for example, Unistrokes™) appeared as alternative handwriting recognition methods (Mackenzie and Zhang, 1997; Goldberg and Richardson, 1993). This study did not investigate these input methods.

Even after optimization, none of these input methods are likely to effectively replace the use of a full-sized, standard keyboard (such as a computer keyboard), especially if the user has even moderate typing skill. For this reason, designers of PDAs should provide a computer version of the PDA's databases (with data upload from and download to the PDA) or an attachable full-sized keyboard to allow users to rapidly enter their information into the PDA.

TRADEMARKS

Simon is a trademark of BellSouth Corp.
Unistrokes is a trademark of Xerox Corp.

REFERENCES

Bailey, R. W. (1989). *Human performance engineering: Using human factors/ergonomics to achieve computer system usability*. Englewood Cliffs, NJ: Prentice Hall.

Francas, M., Brown, S., and Goodman, D. (1983). Alphabetic entry procedure with small keypads: Key layout does matter. In *Proceedings of the Human Factors Society 27th Annual Meeting* (pp. 187-190). Santa Monica, CA: Human Factors Society.

Goldberg, D., and Richardson, C. (1993). Touch-typing with a stylus. In *Conference Proceedings on Human Factors in Computing Systems – CHI '93* (pp. 80-87). New York, NY: ACM.

Koester, H. H., and Levine, S. P. (1994). Modeling the speed of text entry with a word prediction interface. *Transactions on Rehabilitation Engineering*, *2*, 177-187.

Lewis, J. R. (1993). Pairs of Latin squares that produce digram-balanced Greco-Latin designs: A BASIC program. *Behavior Research Methods, Instruments, & Computers*, *25*, 414-415.

Lewis, J. R., Allard, D. J., and Hudson, H. D. (1999). Predictive keyboard design study: Effects of different word populations, number of displayed letters, and number of transitional probability tables. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (in press). Santa Monica, CA: Human Factors Society.

Lewis, J. R., Kennedy, P. J., and LaLomia, M. J. (1999). Development of a digram-based typing key layout for stylus input. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (in press). Santa Monica, CA: Human Factors Society.

Lewis, J. R., LaLomia, M. J., and Kennedy, P. J. (1999). Evaluation of typing key layouts for stylus input. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (in press). Santa Monica, CA: Human Factors Society.

MacKenzie, I. S., Nonnecke, B., Riddersma, S., McQueen, C., & Meltz, M. (1994). Alphanumeric entry on pen-based computers. *International Journal of Human-Computer Studies*, *41*, 775-792.

Mackenzie, S. I., and Zhang, S. X. (1997). The immediate usability of Graffiti. In *Proceedings of Graphic/Vision Interface '97* (pp. 129-137). Toronto: Canadian Information Processing Society.

Norman, D. A., and Fisher, D. (1982). Why alphabetic keyboards are not easy to use: Keyboard doesn't much matter. *Human Factors*, *24*, 509-519.

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