

A COMPARISON OF THREE METHODS OF CHARACTER ENTRY ON PEN-BASED COMPUTERS

I. Scott MacKenzie[†], R. Blair Nonnecke[†], J. Craig McQueen[‡], Stan Riddersma[†], and Malcolm Meltz[§]

[†]Dept. of Computing & Information Science
University of Guelph
Guelph, Ontario, Canada N1G 2W1

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

[§]Architel Systems Corp.
Toronto, Ontario
Canada, M8V 1A4

Methods for entering text on pen-based computers were compared with respect to speed, accuracy, and user preference. Fifteen subjects entered text on a digitizing display tablet using three methods: hand printing, QWERTY-tapping, and ABC-tapping. The tapping methods used display-based keyboards, one with a QWERTY layout, the other with two alphabetic rows of 13 characters. ABC-tapping had the lowest error rate (0.6%) but was the slowest entry method (12.9 wpm). It was also the least preferred input method. The QWERTY-tapping condition was the most preferred, the fastest (22.9 wpm), and had a low error rate (1.1%). Although subjects also liked hand printing, it was 41% slower than QWERTY-tapping and had a very high error rate (8.1%). The results suggest that character recognition on pen-based computers must improve to attract walk-up users, and that alternatives such as tapping on a QWERTY soft keyboard are effective input methods.

INTRODUCTION

Pen-based computers have received much attention in the media recently, primarily due to new technologies entering the marketplace. They are appearing as personal computers, Personal Digital Assistants (PDAs), and large "whiteboard" displays. The market for pen-based computers includes people who work intensively with information and who work away from a desk (e.g., field service personnel, couriers, doctors).

It has been suggested that the ability of pen-based technology to recognize handwriting makes it revolutionary and will change the way people enter information into computers. Already, many products have been released, such as the IBM *ThinkPad* or the Fujitsu *325Point*, where printed characters are recognized. Unfortunately consideration and evaluation of alternate entry methods is often ignored. Impartial empirical evaluations are necessary to determine which method is optimal for text entry. This paper compares three methods of character entry for pen-based computers.

Printed Character Recognition

Handwriting has received the most attention as an obvious and preferred input method for pen devices. Few pen-based computers have recognizers capable of interpreting cursive text (although this may change with products such as the Apple *Newton*). Most commercial recognizers convert the strokes of a printed character to an ASCII value. There are a variety of recognizer "engines" on the market; for example, Gibbs (1993) surveys 13 handwriting recognizers from seven different vendors. Recognizers are most effective with block-printed characters. Their performance improves when they exploit context, dictionaries, constrained symbol sets, user

profiles, and training. For example, constraining the symbol set to lower-case letters is effective if lower-case characters are expected in the entry field since this reduces the search space to 26 symbols.

Since the purpose of using a pen as an input device is to capture the skill of using a regular pen, the performance of an "ideal" recognizer should be transparent to the user. That is, a perfect recognizer will accept and interpret natural handwriting at a rate controlled by the user and the accuracy of the recognizer would be equivalent to the accuracy of a human attempting to read the writing. People have a high tolerance for handwriting anomalies, as they draw on semantics, syntax, and context when interpreting.

The accuracy of recognizers is the key to their success. Of the 13 recognizers surveyed by Gibbs (1993), seven quoted untrained, 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. Gibbs also notes: "there is no accepted standard for evaluating accuracy. Each vendor assesses their own accuracy as they please" (p. 31). If these accuracy figures are correct, then the performance of the recognizers on lower-case-only entry should be much better because the character set is constrained to just 26 symbols.

Keyboard Tapping

An alternative entry method is to select characters on a soft keyboard with the tap of a pen or stylus. Soft keyboards exist in current graphical user interfaces, and are also common in interfaces for the disabled (Shein, Treviranus, Brownlow, Milner, & Parnas, 1992). Sears (1991) reported an entry rate of 17 wpm selecting characters with a mouse on a soft

QWERTY keyboard.

Impartial empirical tests of typing speeds and error rates for pen-tapping have not been published, although proprietary data are available (e.g., Carr & Shafer, 1991). The closest published input scheme is a touch screen keyboard with text entry using fingers. For touch entry, Gould, Greene, Boies, Meluson, and Rasamny (1990) reported typing rates of 12 wpm; and Wiklund, Dumas, and Hoffman (1987) found speeds of 14-18 wpm with error rates under 1%. Sears (1991) tested a group of expert users and found speeds of 25 wpm. In the latter study, subjects used both hands, so comparisons with pen-tapping are weak.

There are measurable but relatively small speed differences between various keyboard layouts for practiced users (e.g., DVORAK vs. QWERTY). Display-based keyboards must be unobtrusive, for example, by reducing the size of the keyboard while retaining the QWERTY layout, or elongating the keyboard so it occupies less vertical or horizontal space. In the latter case, the characters are usually arranged in alphabetic order.

A disadvantage of pen-tapping is the lack of kinesthetic feedback and the inability to have a reference point (Wiklund et al., 1987). Hence, visual contact with the on-screen keypad must be maintained during entry. For example, a border-crossing guard using a pen computer to enter license plate numbers would be severely constrained if the system required on-screen eye fixation.

An experiment is described in the next section that investigates hand printing and two variations tapping on a soft keyboard with a pen.

METHOD

Subjects

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

Apparatus

Software to run the experiment was developed in C using Microsoft's *Pen For Windows* and the Microsoft character recognizer. The recognizer was constrained to 26 lowercase letters. 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 input and a 640 x 480 LCD screen. Using the combination of the tablet and high speed personal computer allowed the experiment to run without introducing lag and also allowed character entry to be observed on a regular VGA monitor. Characters were randomly ordered from a fixed set of phrases. The single letter frequency count table of Mayzner and Tresselt (1965) was used to create a character-balanced phrase set.

Procedure

The task consisted of entering characters provided by the software using one of the three methods. The conditions were (a) hand printing, (b) tapping on a QWERTY soft keyboard (QWERTY-tapping), and (c) tapping on an ABC soft keyboard (ABC-tapping), as illustrated in Figure 1. 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. Nine blocks were used for each condition for a total of 594 characters (including blanks).

Subjects performed all three conditions in a one hour session. Conditions were counterbalanced using a Latin square to minimize transfer effects.

Subjects were instructed to aim for both speed and accuracy when entering the characters. As well, they were 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 preferred by several subjects.

Execution of a condition consisted of a brief practice session of 3 phrases and then 9 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.

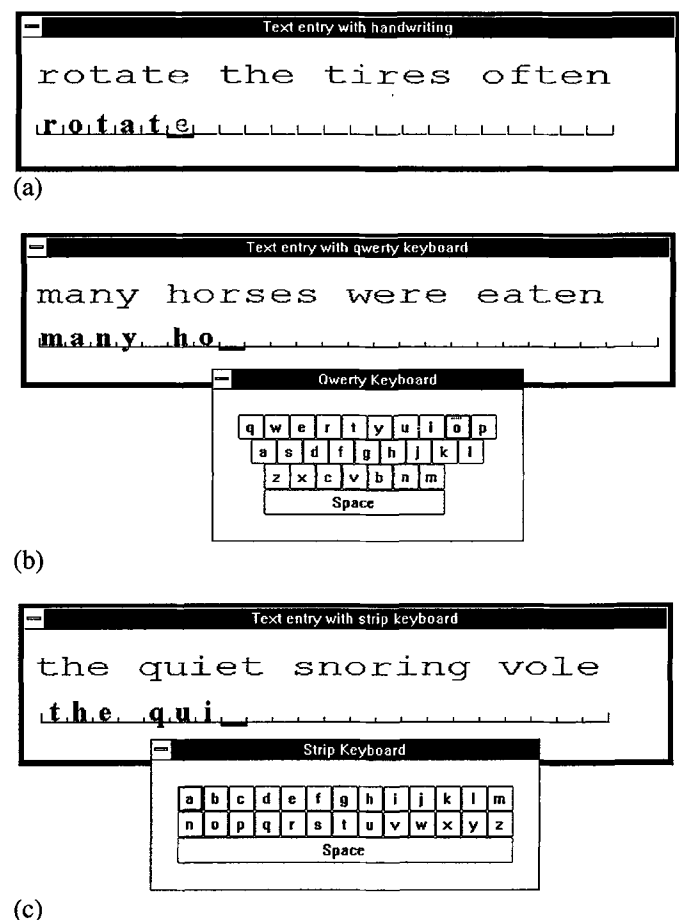


Figure 1. The three experimental conditions were (a) hand printing, (b) QWERTY tapping and (c) ABC-tapping.

RESULTS

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 for each of the 15 subjects.

There was a significant main effect for condition on users' entry time ($F_{2,28} = 95.6, p < .0001$) and error rate ($F_{2,28} = 33.6, p < .0001$). The mean values for each condition are shown in Figure 2. Entry times were converted to "words per minute" (wpm) for comparison with other studies. We used the typists' definition of a word – five characters including spaces.

Accuracy for hand printing was more highly varied than for the other two conditions, as seen in Figure 3. 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%) better than the performance of another subject using QWERTY-tapping (3.1%).

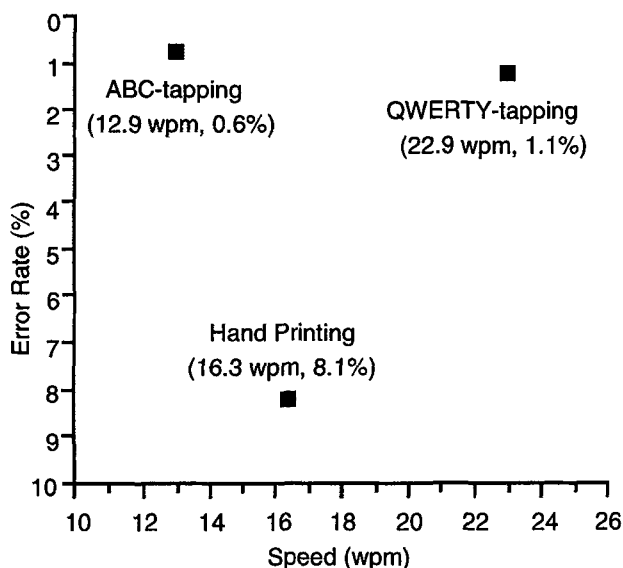


Figure 2. Comparison of the three conditions for error rates and entry speed.

Learning

Although there was no effect across blocks for accuracy ($F_{8,112} = .74$), there was a significant main effect across blocks for entry time ($F_{8,112} = 12.9, p < .0001$). Apparently, subjects did not improve their accuracy with practice, however, they did get faster as seen in Figure 4. This is consistent with Bailey's (1989) observation that "in activities where 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).

QWERTY-tapping and hand printing showed the greatest improvement in absolute speed over the 9 blocks (increase of 2.8 wpm and 2.9 wpm respectively); however, hand printing had the highest relative rate of improvement (19.9%) due to the lower initial value. The ABC-tapping condition improved the least over the 9 blocks (2.0 wpm for a 16.3% increase in speed).

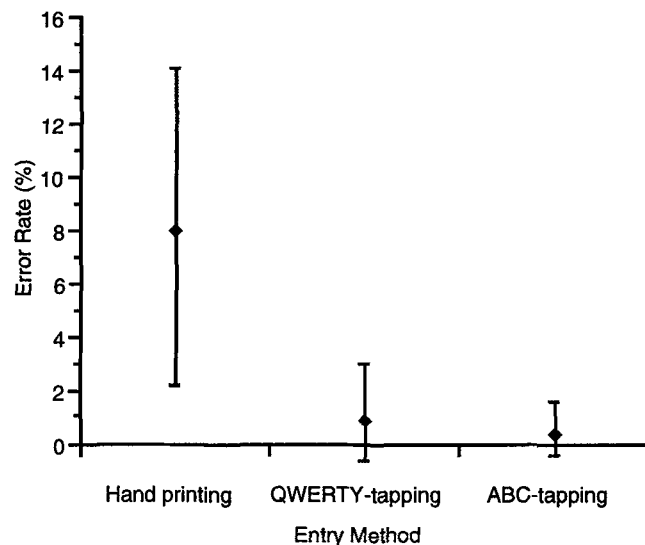


Figure 3. Error rate for each condition with standard deviation error bars.

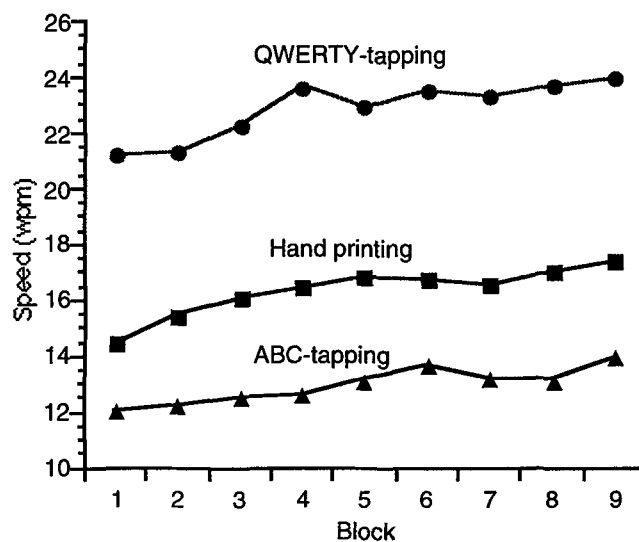


Figure 4. Learning as increasing entry speed.

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 decomposed by character. The most frequently misinterpreted character was the letter "n" (13.4% of all errors, as shown in Figure 5). For each character, there are 25 possible misinterpretations. As shown in Figure 6, the Microsoft recognizer posted the letter "c" most frequently when a recognition error was made (35.9% of all errors).

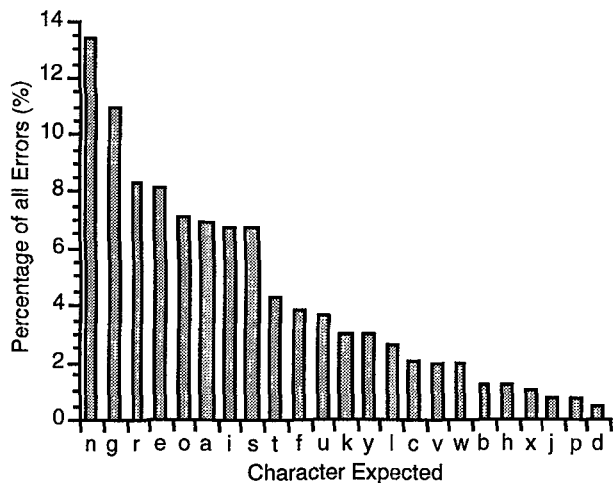


Figure 5. Characters expected by the subjects that were posted as some other character.

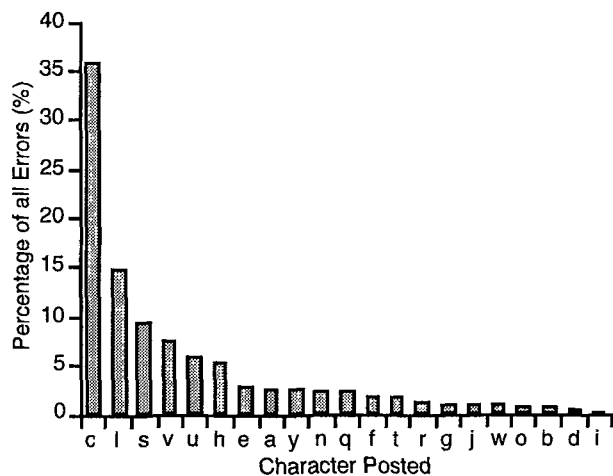


Figure 6. Characters posted by the recognizer in error.

In all, there were 81 unique error pairs (character expected vs. character posted). Figure 7 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 characters 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 Figure 8. Hand printing and QWERTY-tapping received equally high first-choice

ratings, each being preferred by 7 subjects, while the ABC-tapping was 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).

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

Figure 8. Subject preferences (frequency, $n = 15$)

DISCUSSION

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 could achieve high entry rates using ABC-tapping given sufficient practice. As well, suggestions were made to improve the performance of ABC-tapping: placing the characters in a 5 x 6 matrix rather than a 2 x 13 matrix, and putting them in one long row or column. Subjects believed that the 5 x 6 matrix would provide a smaller visual scanning area and reduce the confusion caused by the arbitrary break in the ABC keyboard.

None of the users suggested that the keys were either too small or too large, even though most of the 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.

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 (lowercase 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.

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

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

Figure 7. The 10 most-frequent translation error pairs for the hand printing condition.

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.

CONCLUSION

Of the three conditions, QWERTY-tapping was fastest and was most preferred by the subjects. It also had a low error rate. Subjects disliked the unfamiliar and slow ABC-tapping condition. Although it had the highest error rate, hand printing was preferred nearly as much as QWERTY-tapping.

For hand printing to be more readily accepted by walk-up users, the performance of character recognizers must be improved beyond the current state. At the same time, alternate input methods should be sought for users who prefer not to print.

ACKNOWLEDGMENTS

This research is supported by Architel Systems Corporation and the University Research Incentive Fund (URIF) of the Province of Ontario. We gratefully acknowledge this support without which this research would not have been possible.

REFERENCES

- Bailey, R. W. (1989). *Human performance engineering* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Carr, R., & Shafer, D. (1991). *The power of Penpoint*. Reading, MA: Addison-Wesley.
- Gibbs, M. (1993, March/April). Handwriting recognition: A comprehensive comparison. *Pen*, pp. 31-35.
- Gould, J. D., Greene, S. L., Boies, S. J., Meluson, A., & Rasamny, M. (1990). Using a touchscreen for simple tasks. *Interacting with Computers*, 1, 59-74.
- Mayzner, M. S., & Tresselt, M. E. (1965). Tables of single-letter and digram 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.
- Shein, G. F., Treviranus, J., Brownlow, N. D., Milner, M., & Parnas, P. (1992). An overview of human-computer interaction techniques for people with physical disabilities. *International Journal of Industrial Ergonomics*, 9, 171-181.
- Wiklund, M. E., Dumas, J. S., & Hoffman, L. R. (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*, 585-589. Santa Monica, CA: Human Factors Society.