

# The Design and Evaluation of a High-Performance Soft Keyboard

**I. Scott MacKenzie**

Dept. Computing and Information Science  
University of Guelph  
Guelph, ON N1G 2W1 Canada  
+1 519 824 4120 x8268  
smackenzie@acm.org

**Shawn X. Zhang**

Dept. Computing and Information Science  
University of Guelph  
Guelph, ON N1G 2W1 Canada  
+1 519 824 4120  
shawnz@acm.org

## ABSTRACT

The design and evaluation of a high performance soft keyboard for mobile systems are described. Using a model to predict the upper-bound text entry rate for soft keyboards, we designed a keyboard layout with a predicted upper-bound entry rate of 58.2 wpm. This is about 35% faster than the predicted rate for a QWERTY layout. We compared our design ("OPTI") with a QWERTY layout in a longitudinal evaluation using five participants and 20 45-minute sessions of text entry. Average entry rates for OPTI increased from 17.0 wpm initially to 44.3 wpm at session 20. The average rates exceeded those for the QWERTY layout after the 10<sup>th</sup> session (about 4 hours of practice). A regression equation ( $R^2 = .997$ ) in the form of the power-law of learning predicts that our upper-bound prediction would be reached at about session 50.

## Keywords

Soft keyboards, mobile systems, stylus input, pen input, linguistic models, Fitts' law, digraph probabilities

## INTRODUCTION

Besides handwriting recognition, one popular method of text input for pen-based mobile systems is through a soft keyboard. Users enter text by tapping on the image of a keyboard on the system's display. Although the QWERTY layout is entrenched for physical keyboards, soft keyboards are easy to implement and modify. So, a reasonable goal is to search for an alternate, perhaps better, layout than the venerable QWERTY. This paper describes the design and evaluation of one such layout, which we call OPTI in reference to our goal of designing an optimal soft keyboard.

## Learning Time and the Elusive Crossover Point

A usability issue is learning time. Users who bring desktop computing experience to mobile computing may

fare poorly on a non-QWERTY layout — at least initially. Thus, longitudinal empirical testing is important. We want to establish not only a layout's potential for experts, but also the learning time for typical users to meet and exceed entry rates with a QWERTY layout.

In a general sense, we are comparing the viability of a "new technique" against "current practice". We expect lower performance measures for the new technique initially, but these should eventually "crossover", wherein performance with the new technique exceeds that with current practice. This is illustrated in Figure 1.

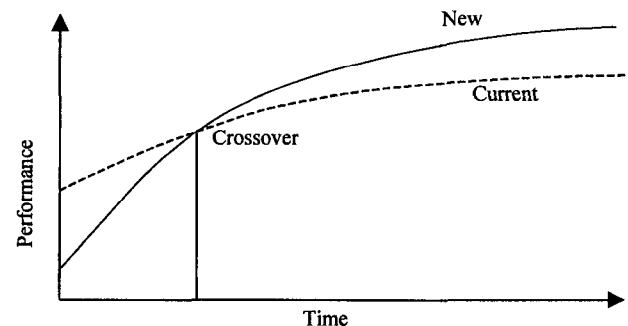


Figure 1. The elusive crossover point

The crossover point may not be achieved, however; and there are a variety of possible explanations. Perhaps the new technique was simply not as good, or perhaps further refinement was needed. It is also possible that the study was terminated before the crossover point could be reached. Two examples are cited below.

McQueen et al. [6] tested six participants in a study on numeric entry with a stylus. Two methods were tested, one using a standard numeric keypad ("current practice") and one using numeric pie menus ("new technique"). In a study involving six participants in 20 20-minute sessions, the crossover point occurred at the 7<sup>th</sup> session. In another study, Bellman and MacKenzie [1] compared two text-entry techniques for small hand-held devices such as pagers. The technique used five finger-operated buttons to move a cursor and select characters on a small display.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI '99 Pittsburgh PA USA

Copyright ACM 1999 0-201-48559-1/99/05...\$5.00

The character:: were displayed either in a fixed alphabetic pattern ("current practice") or in a pattern that fluctuated after each entry to minimize the required cursor movement ("new technique"). In a study involving 11 participants in 10 30-minute sessions, a crossover point was not attained. Although a crossover point may have been reached with further practice, a detailed analysis suggested that further refinements to the technique were warranted.

### Modeling Text-Entry with Soft Keyboards

Since longitudinal user studies are labor intensive, we developed a prediction model. With pen-based mobile systems, the input channel for text entry is reduced from ten-fingers (as in touch typing) to one finger or pen. This is fortunate because modeling the psychomotor act of stylus tapping is simple compared with touch typing.

Our model has several components, including linguistic data, Fitts' law, a shortest-path model, and a key-repeat-time measure. The model generates a theoretical text entry rate (in words per minute) for any layout of soft keyboard. This allowed us to evaluate alternate designs "on paper" before proceeding to an "empirical" evaluation.

Our linguistic model is for "common English" — the assumed language for text entry. Tables are available giving the 26 letter frequencies and the 26 × 26 letter-pair (digraph) frequencies in common English [5]. Because linguists focus on language, the frequency data do not include spaces — the most common character in text entry tasks. Soukoreff and MacKenzie [7] extended the tables to include space characters. Their data provide 27 letter frequencies and 27 × 27 digraph frequencies. For entering common English, the most common character is space ( $p = .1863$ ) and the most common digraph is e-space ( $p = .0457$ ).

We use Fitts' law to predict the time to tap a key given any previous key. We compute the amplitudes ( $A$ ) for all the 27 × 27 digraph movements in a given keyboard layout, and, for each, compute the movement time ( $MT$ ) using a Fitts' law model for stylus tapping [3]:

$$MT = \frac{1}{4.9} \log_2 \left( \frac{A}{W} + 1 \right)$$

where  $W$  is the width of each key. The mean  $MT$  is then computed by summing the 27 × 27  $MT$ s, each weighted by the digraph probability. The mean  $MT$  is then converted to text entry speed in words per minute (wpm), assuming five characters per word. The result is an "upper bound" prediction, since it assumes the *visual scan* time to find a key is zero.

We included two refinements to the model. The first is a "shortest path" model. When a long key (such as key2 in Figure 2) is involved in a key1-key2-key3 sequence, we

compute the shortest path among several discrete paths. This is an assumed behavior for experts.

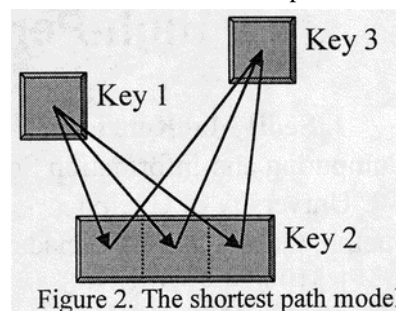


Figure 2. The shortest path model

The second is a key-repeat-time measure. Double letters (e.g., t-t) require no lateral pen movement. They require a repeat tap on the same key. To model this, we conducted a simple experiment. We asked five individuals to make 25 quick taps on the same key on a soft keyboard 10 times. The software captured the entry time between taps. The average key repeat time was 127 ms.

The components of our model were implemented in a spreadsheet; thus, rapid predictions were available for any soft keyboard design. As an example, the QWERTY layout in Figure 3 has a predicted upper-bound entry rate of 43.2 wpm.

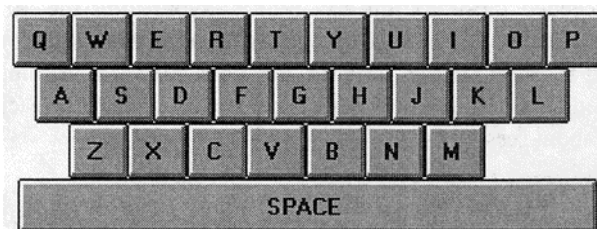


Figure 3. A QWERTY soft keyboard layout

### DESIGNING A HIGH-PERFORMANCE SOFT KEYBOARD

In addition to the model described above we used the following design rationale:

- The soft keyboard should have no dead space between keys, where no action is assigned.
- The entire keyboard shape should be rectangular to fill in a typical application window.
- There is no limit on how many sizes or shapes can be used. We decided to use only two sizes: square (1 x 1 unit) for alphabet keys and square or rectangular (1 x 2 units) for space keys.

Following substantial trial and error -- where each iteration yielded a higher prediction than the previous -- we settled on the design in Figure 4, which we call OPTI. The predicted upper-bound text entry rate for OPTI is 58.2 wpm. This is 35% faster than for the QWERTY layout. It is also about 5% faster than our predicted entry rate for a commercial soft keyboard known as *Fitaly*,

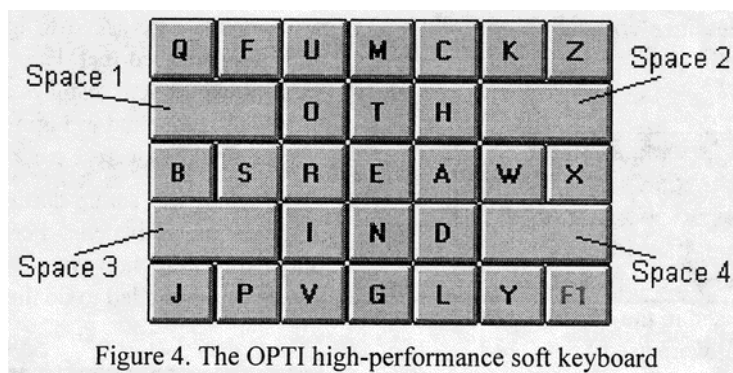


Figure 4. The OPTI high-performance soft keyboard

from Textware Solutions (Boston, MA).<sup>1</sup> Many commonly used sequences such as “THE”, “WH”, “EA”, “CK”, “LY” or “ING” are tightly located, so the pen travel distance is shorter. The design is nearly symmetrical, making it suitable for either hand. Note that the four space keys are very accessible (about 36% of the digraphs in common English involve the space character).

### METHOD

A usability study was undertaken to evaluate the OPTI soft keyboard. Since QWERTY is the most common layout, we included both layouts in the study. Our goal was not just to evaluate the OPTI layout, but to determine its performance relative a QWERTY soft keyboard. We fully expected the QWERTY layout to be better initially — because of users' familiarity with QWERTY — but we wanted to establish the learning trends for both layouts and to determine if and when a crossover point would occur, wherein OPTI would fare better than QWERTY.

### Participants

In longitudinal studies, fewer participants are usually engaged but they are tested over a prolonged period of time. In this study, we used five participants. All were university computer science students, four male, one female. All were right handed and used desktop computers on a regular basis. None had regular experience using a pen-based computer. They were recruited from a pool of subjects who participated in other unrelated experiments. Since we are testing a keyboard specifically designed for English, we picked only participants whose first language was English. All were well informed on the time commitment required for the experiment.

### Apparatus

The experiment software was developed with Borland C++ 4.0 and ObjectWindows Library (OWL 2.0). The

host system was a Packard Bell 486DX-50 PC running Microsoft *Windows for Pen Computing* 1.1. A Wacom *PL-IOOV* combining an LCD display and digitizing tablet was attached to the system and was the only device used by the participants.

The experiment was conducted in the HCI Lab at the University of Guelph's Dept. of Computing and Information Science. To minimize interference from any other source the lab was completely booked for the experiment. The entire experiment took about four weeks. A special web site was created for information updates and participant scheduling.

### Design

The experiment was a 2 x 20 within-subjects factorial design. The two factors were:

- Keyboard layout (QWERTY, OPTI)
- Session {20 sessions}

Each session lasted about 45 minutes and was divided into two 20-22 minute periods. One of the two layouts was assigned in each half-session period in alternating order from session to session. The order of the conditions was balanced between participants to reduce interactions.

Each half-session contained several blocks of trials. The number of blocks for each half-session period was controlled such that as many blocks as possible were collected within the allotted time. Therefore, in the early sessions, fewer blocks (5 to 6) were administered than in later sessions (9 to 11). A five-minute break was allowed between the two half-sessions.

Each block contained 10 text phrases of about 25 characters each. These 10 phrases were randomly selected from a source file of 70 phrases. Phrases were not repeated within blocks but repeats were allowed from block to block. The phrases were chosen to be representative of English and easy to remember (see Figure 5). The sample phrase set was tested for its correlation with common English using the frequency counts in Mayzner and Tresselt's corpus [5]. The result

<sup>1</sup> The *Fitaly* keyboard layout fares quite well according to our prediction model. To our knowledge, however, no empirical evaluation of this product has been published. For more information see <http://www.twolutions.com/>.

was  $r = .9845$  for the single-letter correlation and  $r = .9418$  for the digraph correlation.

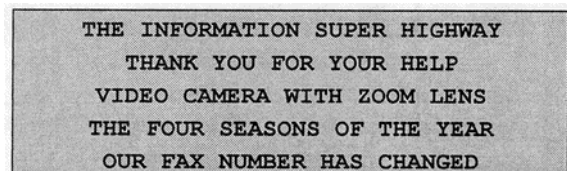


Figure 5. Sample phrases used in the experiment (70 phrases in total were used.)

Each participant completed 20 sessions. Sessions were scheduled Mondays through Saturdays, separated by at least two hours but no more than two days. This was to simulate “regular use” of the system while trying to avoid fatigue and accommodating participants’ daily schedules.

This was a longitudinal study attempting to practice participants toward expert performance. Data collection included numerous measurements on user input. For each key tapped, the following was collected.

- Given character
- Given character’s position (ID on the keyboard)
- User entered character
- User entered character’s position. (The four space keys were identified separately.)
- Elapsed time (ms) from previous tap
- Error (1 if the user character was different from the given character; 0 if user character was correct)

### Procedure

Each participant was given written instructions explaining the task and the goal of the experiment. They were asked specifically to aim for both entry speed and accuracy. The instructions also stated that if they made more than 10% errors within a phrase (about 3 mistakes) they should slow down on the next phrase to increase accuracy.

As designed, the length of each half-session period was controlled with a timer. Once started the software was self-administered. The entire session was monitored on a separate CRT connected to the system.

Participants were then given the tablet and the stylus. The tablet was tilted off the desk to provide a good viewing angle (about 25°). It was also adjusted to have appropriate contrast and brightness. The overhead lights were turned off to reduce the glare on the tablet’s display panel. The height of the desk was 26 inches, a standard height for typing. The desktop could be adjusted by about two inches to allow for different body sizes.

The participants were asked to copy each short phrase by tapping on the soft keyboard. A soft audio feedback “tick” was heard for each character entered. When an error occurred a more prominent “click” was heard. The

participants were asked to ignore errors and to carry on with the next correct letter pointed at by the cursor. Typical experiment displays are illustrated in Figure 6 for the OPTI layout and in Figure 7 for the QWERTY layout. The square keys were 1 cm x 1cm.

A plot chart was set up during the experiment to keep the participants motivated. Performance expectations were not explained, however. Instead, participants were constantly reminded to do their best on both layouts.

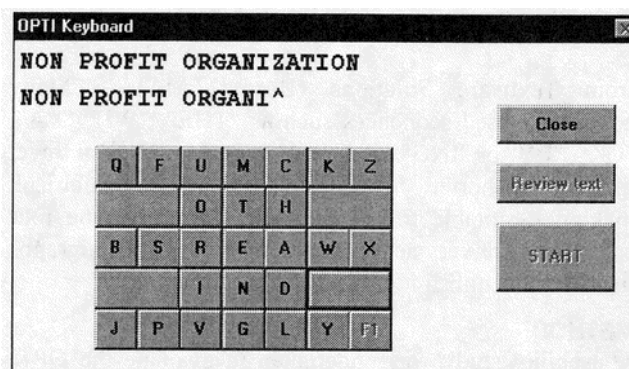


Figure 6. Experiment screen with the OPTI layout

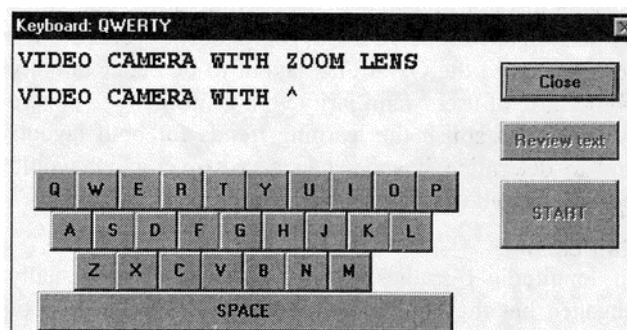


Figure 7. Experiment screen with the QWERTY layout

## RESULTS AND DISCUSSION

### Text Entry Speed – The Learning Curve

The analysis of variance of text entry speed showed no main effect for keyboard ( $F_{1,4} = 0.60, p > .05$ ). There was a significant effect of session ( $F_{19,76} = 89.2, p < .0001$ ) and a significant keyboard-by-session interaction ( $F_{19,76} = 34.3, p < .0001$ ).

The results above were as expected. That is, the OPTI layout fared poorly initially (17 wpm) in comparison with the QWERTY layout (28 wpm). With practice, however, the OPTI layout eventually out-performed the QWERTY layout (see Figure 8). The crossover occurred

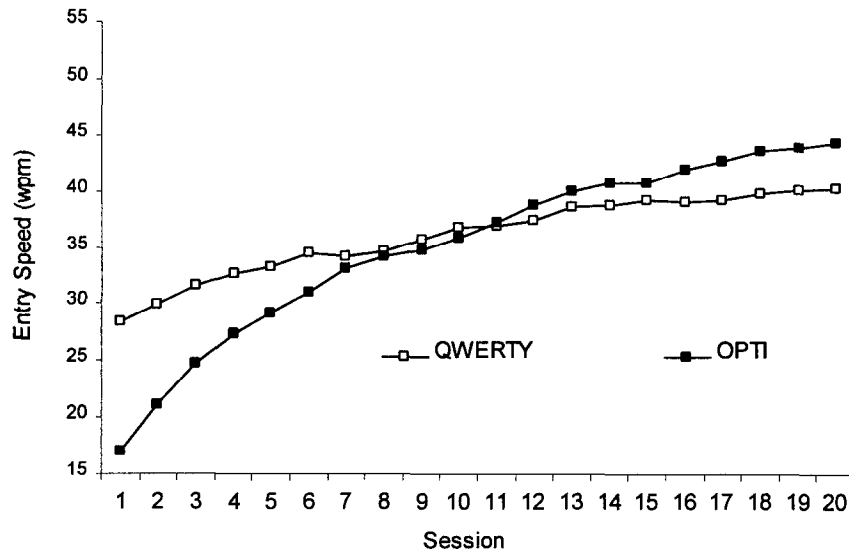
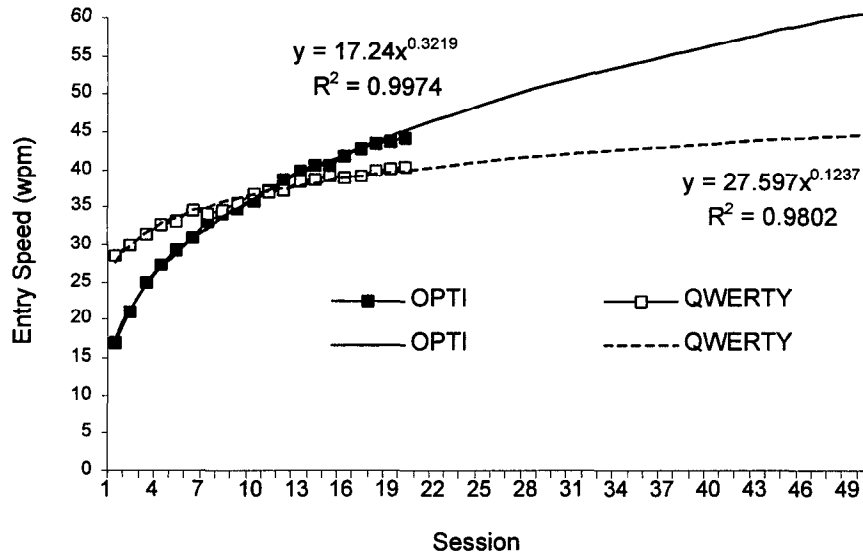


Figure 8. Entry speed by keyboard layout and session

Figure 9. Learning curves and extrapolations to 50<sup>th</sup> session

at the 10<sup>th</sup> session. This is just under four hours of practice.

As the experiment progressed performance continued to improve with the OPTI layout, whereas performance showed signs of leveling off with the QWERTY layout. The average text entry rate for the OPTI layout reached nearly 45 wpm by the 20<sup>th</sup> session and the performance of the QWERTY layout reached about 40 wpm.

For each layout, we derived standard regression models in the form of the power law of learning (e.g., see [2]). The prediction equations and the squared correlation coefficients are illustrated in Figure 9. The high  $R^2$  values

imply that the fitted learning models provide a very good prediction of user behaviour. In both cases over 98% of the variance is accounted for in the models. The somewhat lower  $R^2$  value for the QWERTY layout may be explained as follows. Since our participants were experienced computer users, they were familiar with the QWERTY layout at the start of the study. By no means is the prediction model for the QWERTY layout capturing users' learning behavior from their "initial exposure" to the layout; subjects were "well along" the learning curve. For the OPTI layout, however, users had no prior experience with the layout, and, so, the learning model is

more representative of the initial exposure and the learning thereafter.

Although our longitudinal study lasted 20 sessions, the participants had not become "experts" on the OPTI layout by a mere seven hours of use. So, we mathematically extended the learning curves for another 30 sessions to project performance with further practice (see Figure 9). The extrapolation for the QWERTY layout was 44.8 wpm and for the OPTI layout, 60.7 wpm. These two values, representing about 17 hours of practice with each layout, are close to our theoretical upper-bound predictions noted earlier.

### Error Rates

An error was recorded when the user-entered character differed from the given character. The error rates ranged from 2.07% for OPTI and 3.21% for QWERTY on the first session to 4.18% for OPTI and 4.84% for QWERTY on the 20<sup>th</sup> session.

An analysis of variance revealed a significant difference in error rates between the two keyboard designs ( $F_{1,4} = 12.30, p < .05$ ). QWERTY had consistently higher error rates throughout the experiment. There was also a significant increase in error rates over sessions ( $F_{19,76} = 4.42, p < .001$ ). This may have occurred because entry speed increased over sessions, thus participants' input tended to continue into the reaction time following an error. This behaviour has been noted in other text entry studies (e.g., [4]).

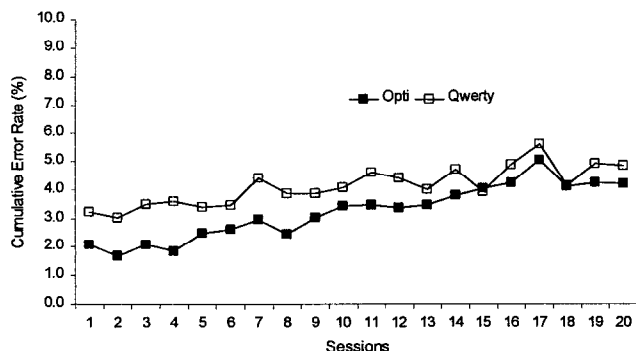


Figure 10. Error rates by layout and session

### Use of the Space Keys

Since the space character is so prominent in the text-entry task, it is worth examining participants' behaviour in their use of the four spaces keys in the OPTI layout. For any character-space-character sequence at least one space key would create the shortest path. We call this the "optimal space key". Participants were allowed to use space keys at their discretion; however, the data file distinguished among the four space keys. As learning progressed, a few patterns could emerge, such as a tendency to use (a) the optimal space key, (b) a randomly-chosen space key, (c)

the closest space key following a character, or (d) a "favorite" space key.

A favorite space is a personal choice, and is not necessarily optimal. It might be the one that stays visible more often than others, for example. Note that for each space entered there was one optimal space key and three non-optimal space keys.<sup>2</sup>

Figure 11 shows a slight increase in participants' use of the optimal space keys, from 38% in session 1 to 47% in session 20.

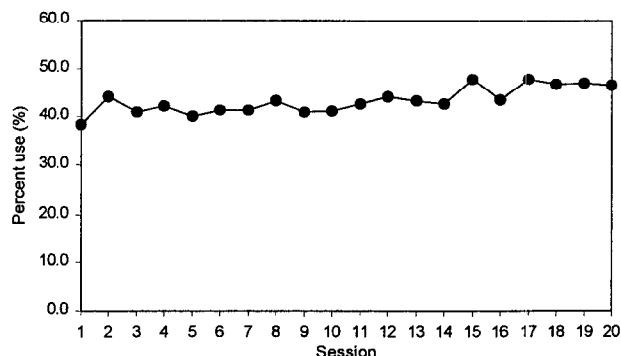


Figure 11. Use of optimal spaces with OPTI over sessions

Although the percentages in Figure 11 are well above random choice (25%), they do not suggest a strong tendency to use the optimal space key. Having four space keys is convenient; but, using the optimal space key requires extra judgement on-the-fly and this is not likely to occur — at least within the confines of the limited practice in the present study.

In re-examining our digraph table, we noted that the ratio of character-only digraphs to digraphs involving spaces is 62:38. This means that 62% of the time, the pen travels from character to character when entering common English. Thus, behavioral improvements or further efforts to optimally accommodate space key usage will have a limited payoff.

### CONCLUSIONS

We have described the design and evaluation of a high-performance soft keyboard for mobile systems. Our results indicate that after about four hours of practice users' entry rates will be higher with the OPTI layout than with a QWERTY layout. After about seven hours of practice users achieved a mean entry rate of 44.3 wpm. Our model predicts that entry rates will edge upward, reaching about 58 wpm for expert users.

<sup>2</sup> When the character-space-character pattern is symmetrical there would be two or more space keys that create the same shortest path. In an extreme case with E-Space-E, all the four space keys are equally optimal.

These results are important for designers of pen-based systems supporting text entry. Tapping on a soft keyboard is a viable alternative to handwriting recognition, and the OPTI keyboard layout represents one possible approach to this interesting design problem.

#### REFERENCES

- [1] Bellman, T., and MacKenzie, I. S. A probabilistic character layout strategy for mobile text entry, In *Proceedings of Graphics Interface '98*. Toronto: Canadian Information Processing Society, 1998, pp. 168-176.
- [2] Card, S. K., English, W. K., and Burr, B. J. Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT, *Ergonomics* 21 (1978), 601-613.
- [3] MacKenzie, I. S., Sellen, A., and Buxton, W. A comparison of input devices in elemental pointing and dragging tasks, In *Proceedings of the CHI '91 Conference on Human Factors in Computing Systems*. New York: ACM, 1991, pp. 161-166.
- [4] Matias, E., MacKenzie, I. S., and Buxton, W. One-handed touch typing on a QWERTY keyboard, *Human-Computer Interaction* 11 (1996), 1-27.
- [5] Maynzer, M. S., and Tresselt, M. E. Table of single-letter and digram frequency counts for various word-length and letter-position combinations, *Psychonomic Monograph Supplements* 1,2 (1965), 13-32.
- [6] McQueen, C., MacKenzie, I. S., and Zhang, S. X. An extended study of numeric entry on pen-based computers, In *Proceedings of Graphics Interface '95*. Toronto: Canadian Information Processing Society, 1995, pp. 215-222.
- [7] Soukoreff, W., and MacKenzie, I. S. Theoretical upper and lower bounds on typing speeds using a stylus and keyboard, *Behaviour & Information Technology* 14 (1995), 370-379.