# A Psychophysical Comparison of Two Stylus-Driven Soft Keyboards

Michael Bohan Chad A. Phipps Alex Chaparro Charles G. Halcomb

Department of Psychology
Wichita State University

*Abstract*

This study compared text entry performance of two stylus-driven soft keyboards for use in hand-held computing devices: the QWERTY and the T9. Participants transcribed text presented on a computer screen into a personal digital assistant (PDA) using a stylus and one of these two keyboards. We introduced a new psychophysical technique for measuring transcription rate that provides a composite measure of speed and accuracy. Using this technique, we calculated the maximum transcription rate for each keyboard. The results show that transcription rates were higher for the QWERTY keyboard than for the T9, despite the T9 keyboard's apparent superior physical characteristics. An ancillary experiment demonstrated that the poorer performance of the T9 layout may have resulted from an increase in visual scanning time due to perceptual grouping of the letters on the keys. Together these findings imply that the QWERTY keyboard layout remains the most effective of the currently available designs for stylus tapping on soft keyboards.

*Key words: Soft keyboards, stylus input, pen-based computing, hand-held devices, transcription rate.*

## 1 Introduction

Recent advances in computing technology have led to a dramatic increase in the availability of hand-held computing devices. From palm-size computers to personal digital assistants (PDAs), personal information managers (PIMs), and pen tablets, the trend has been toward the development of smaller, more mobile devices. This trend has forced manufacturers to consider alternative methods of text entry such as hand-writing and the use of a stylus for tapping on-screen soft keyboards.

While hand-writing recognition capabilities continue to improve, performance under optimal conditions is still fair at best, with reported walk-up accuracy rates ranging anywhere from 85%-93% [1, 2, 3], and entry speeds of around 16-18 wpm [2, 4]. As a result, there remains a good deal of interest in on-screen or soft keyboards as the primary means of input for hand-held computing devices. A wide variety of keyboard designs, differing in the organization of the alphanumeric characters (e.g. QWERTY, ABC, Dvorak, FITALY, etc.), and the manner with which the alphanumeric characters are selected (e.g. tapping, gestures) are being explored [e.g. 4, 5, 6].

In previous studies comparing text entry performance using various soft keyboard layouts, the QWERTY style has consistently outperformed other designs in terms of both speed and accuracy [e.g. 4, 7]. For instance, in a study of five soft keyboard designs, MacKenzie et al. [7] showed that the QWERTY yielded an average text entry rate of 21.1 wpm, with the runner-up (ABC keyboard) attaining an average of only 10.7 wpm. This performance advantage of the QWERTY keyboard is presumably due to the subject's familiarity with computer keyboards [7]. That is, typing on a desk top computer transfers to stylus tapping on soft keyboards. It is important to point out, however, that the nature and extent of this transference is not yet known. Tapping on a soft keyboard with a stylus is different than touch-typing on a standard keyboard. The former uses one-handed input, and due to the lack of tactile feedback, requires continuous visual guidance of the finger or stylus. The latter employs two hands and relies primarily on tactile rather than visual feedback for movement guidance. Nevertheless, familiarity with the location of the letters on the keyboard does appear to facilitate tapping performance [7].

Any alternate keyboard design for use with hand-held computing devices must allow the user to be proficient very quickly since these devices are typically designed for the "walk-up" market. That is, consumers who want to be able to approach the device and begin to use it immediately. The challenge therefore is to come up with alternate key layouts that are organized in a meaningful fashion, allowing faster and more accurate input rates without the need for extensive practice.

The T9 keyboard (see Figure 1), recently developed by Tegic Communication [8], offers an interesting new keyboard alternative. The layout mirrors that of a touchtone telephone (without the numbers), with individual letters grouped onto a single key (e.g. ABC…DEF…GHI). Users are offered larger blocked keys to tap instead of the smaller individual lettered-keys. Furthermore, with the T9, the user need not make multiple taps to identify each letter, as with some other types of telephone keyboards. Instead, this system uses

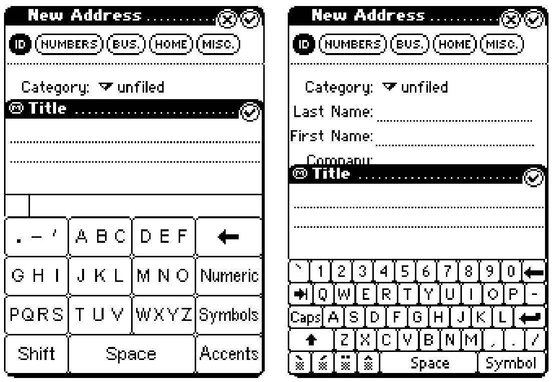


Figure 1. Pictorial representation of the T9 (left) and QWERTY (right).

keyboards.a disambiguation algorithm, in conjunction with a dictionary, to determine the word intended. Obviously, the limitations of such a system are highly dependent upon the ability to accurately differentiate between potential words (e.g. “bat” and “cat”). Nevertheless, this keyboard design is appealing as an alternative to the QWERTY for two reasons. First, the natural familiarity with the layout of the letters on the T9 keyboard, due to exposure during telephone use, may allow users to become proficient at entering text with very little training. Second, because multiple letters are placed on a single key, the keys are necessarily larger than those on a comparably-sized QWERTY keyboard. The larger keys would be expected to yield faster tapping speed and improved accuracy.

Therefore, the primary objective of the present study was to compare text entry rate on the T9 soft keyboard to that of the QWERTY keyboard. A secondary objective was to introduce a new technique for measuring text entry performance, which we call transcription rate.

### 1.1 Transcription Rate

To measure participants’ rate of text entry using these two keyboard layouts, we adapted a psychophysical technique originally developed by Legge and colleagues [9] for measuring rate of reading. Participants are required to read words aloud that are presented on a monitor. The presentation rate (words per minute) is then increased incrementally until errors are made. Reading rate (product of text presentation rate and the percentage of words read correctly) in words per minute is then calculated, and from this, a maximum reading rate can be determined. That is, the maximum rate at which the words can be read without errors. This technique has been shown to provide a convenient and accurate measure of reading performance [10].

In the present experiment, transcription performance was measured in a similar fashion. In this case, rather than reading aloud words presented on the screen, participants transcribed the words using either the QWERTY or T9 soft keyboards. Employing this technique, we were able to arrive at a maximum transcription rate for each keyboard. The primary advantage of measuring transcription rate in this way is that it provides a composite measure of both speed and accuracy. It also allows a rapid and easy evaluation of stimulus parameters (e.g. keyboard layout, font size and shape, etc.) on text input performance.

## 2 Method

### 2.1 Participants

Five right-handed graduate students ranging from 25-28 years of age (M = 27.4, SD = 1.94), including two of the authors, from the psychology department at Wichita State University participated in the study. All of the participants reported having at least 20/20 or corrected to 20/20 vision. In addition, all of the participants had extensive computer experience, but only two had any previous experience with the T9 keyboard.

### 2.2 Materials

Data input was performed on a Texas Instruments Avigo 10 personal digital assistant (PDA). The additional hardware used in this study included a 33 MHz personal computer with a VGA display monitor (14-in diagonal screen running a 640 x 480 pixel resolution) for presenting the on-screen text to the participant. An additional VGA monitor was used to display information to the experimenter. The QWERTY and T9 soft keyboards came standard on the PDA. The software used to display the on-screen text was developed by one of the authors for a previous study [10].

### 2.3 Design and Procedure

Data were analyzed using a completely randomized block design, with keyboard layout being the within-participants variable.

The experimental set up is represented in Figure 2. The participants stood in front of the monitor, located at a height of 147.32 cm from the floor, such that the center of the screen was approximately at eye level. The PDA was attached to a stand located just below the monitor, at a height of 124.46 cm from the floor, and at an angle of ~45 degrees.

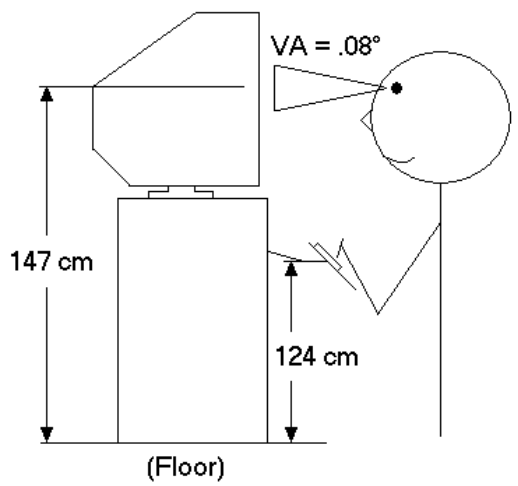


Figure 2. Pictorial representation of the experimental setup.

The overall dimensions of the QWERTY keyboard were 5 x 2.2 cm, and each letter key was 0.4 $cm^2$. The space bar was 1.9 x 0.4 cm. The overall dimensions of the T9 keyboard were 5 x 3 cm, and each key was 1.2 x 0.8 cm. The space bar was 1.9 x 0.8 cm. The height of the letters was equivalent for both keyboards (0.3 cm).

In an experimental trial, each participant was shown text consisting of ten four-letter words randomly selected from a list of 240 words. Words could be repeated within a 10-block trial, but never in succession. Moreover, the words selected for use in the experiment were those that the T-9 algorithm identified as the most probable word for that series of inputs. This was done to keep the display of transcribed words equivalent in both conditions. The height of the characters in each word subtended about 0.8 degrees of visual angle. The words were flashed serially onto the screen. Serial presentation of the text was used, as opposed to scrolling from right to left, removing the participant's need to make saccadic eye movements during reading, and thereby eliminating the additive effects of eye movements on performance. When a word was flashed on the screen, the participant transcribed the word with the memo function of the PDA using one of the two soft keyboards. A word for the purpose of input was defined as the four letters plus a space. Explicit instructions were given to read and transcribe one word at a time. This was to ensure that the participants were not reading ahead and then transcribing multiple words.

The transcription rate data was collected using the method of constant stimuli. The percentage of words correctly transcribed was recorded for each of a set of presentation rates (10, 20, 40, 60, and 80 wpm). These rates were selected such that the resulting transcription performance would span a range from perfect (presentation rate = 10 wpm) to very poor (presentation rate = 80 wpm).

The experiment took approximately 1-hr to complete, during which time five measures of transcription rate were taken for each presentation rate condition. The order of presentation rates was counterbalanced for each participant.

## 3 Results and Discussion

The transcription rate was computed as the product of the presentation rate and the mean number of words correctly transcribed (in percent). In order to estimate the maximum transcription rate, while avoiding experimenter bias, a second order polynomial function of the form $Y = A + BX + CX^2$ was fitted to each participant's data. The maximum transcription rate was the solution to the first derivative of the best fitting function. Figure 3 shows the average transcription rates as a function of presentation rate for each subject. Note that the fit of the second order function to the data was quite good. The lowest $R^2$ value was .69 for participant 4 using the T9 keyboard. The $R^2$ values for all other conditions were in the range of .86 to .98. Furthermore, as can be seen in Figure 3, transcription rate tended to drop off rapidly for presentation rates greater than ~ 40-45 wpm.

The results show that participants were able to transcribe words faster using the QWERTY keyboard than the T9 keyboard. More specifically, the average maximum transcription rate when using the QWERTY was about six wpm higher than when using the T9. A paired-samples T-test revealed that this difference was statistically significant, [t(4) = 4..99, p = .008]. It should be noted that the average maximum transcription rate for the QWERTY keyboard reported here (22.56) is similar to other published figures [e.g. 4,7]. It should also be pointed out that the time required for subjects to shift their gaze back-and-forth between the monitor and the PDA may exacerbate transcription rates. However, this was simply a function of the nature of the task chosen for this experiment (i.e. transcribe written text into the PDA) and would not be expected to differentially affect performance of the two keyboards.

The objective of this study was to investigate the T9 keyboard as a potential alternative soft keyboard design to the standard QWERTY keyboard for use in stylus-driven hand-held computing devices. The T9 was expected to fair well against the QWERTY given its similarity to a telephone layout, and the potential speed and accuracy advantages afforded by its larger keys. Yet despite its apparent physical superiority, transcription rates for the T9 were lower than that for the QWERTY keyboard. It is not entirely clear why we found this discrepancy, however there are at least two potential explanations. First, it is possible that the participants' actual level of experience with the touch phone layout in terms of letter placement was less than originally expected. That is, although people use telephones on a daily basis, these interactions typically involve the numbers not the associated letters.

An alternative interpretation in terms of visual scan requirements also seems possible. That is, the individual letters on the QWERTY keyboard are visually separated by the boundaries of the keys. This

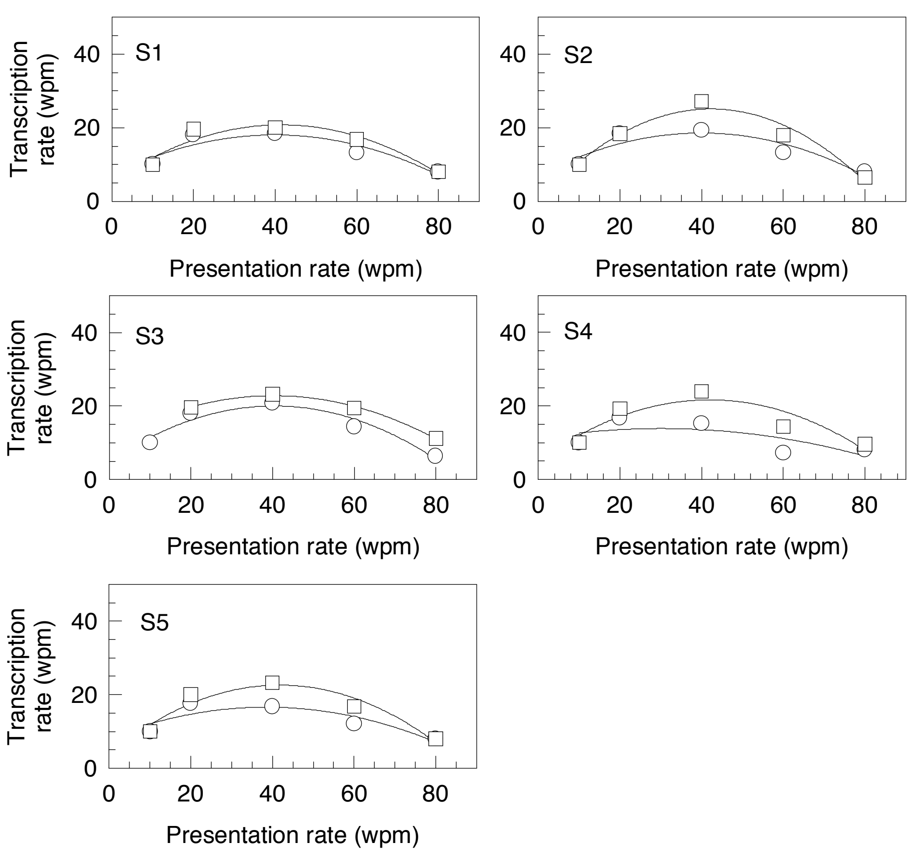


Figure 3. Transcription rate as a function of text presentation rate for the two keyboards. Circle and square symbols represent data obtained for the T-9 and QWERTY keyboards, respectively

may facilitate rapid visual capture of a target letter during transcription. The letters on the T9 keyboard, however, are arranged in groups of 3 or 4, and so by a common border. This arrangement may result in perceptual grouping of the letters within a key, according to the principle of *common region* [11]. This principle states that elements will be perceived as grouped together if they are located within a common region of space or an enclosed boundary. This grouping effect may in turn make it more difficult to process the individual elements within the group…a "can't see the trees for the forest" effect. Therefore, in the T9 layout it may be more difficult for the user to individuate the letters within a key, and thus may require a greater amount of visual scan time.

We tested this hypothesis in an ancillary experiment in which we compared paper prototypes of the standard QWERTY keyboard with a modified version ($QWERTY_m$). The $QWERTY_m$ used the key layout of the T9, but with the letters rearranged to follow the QWERTY system (e.g. "ABC" to "QWE"). In this way, the sequencing of the letters was the same in both conditions, but in the $QWERTY_m$ the letters were grouped (Note: the keyboard dimensions matched those of the actual keyboards). Four of the five participants from the first experiment transcribed the sentence "the quick brown fox jumped over the lazy dogs" as quickly and as accurately as possible using the stylus. Transcription rate was measured using a stop watch. Consistent with the perceptual grouping hypothesis, text entry using the $QWERTY_m$ keyboard ($M = 11.65$ wpm) was significantly slower than with the standard QWERTY keyboard ($M = 25.68$ wpm), [$t(3) = -6.21$, $p = .008$].

The results suggest that grouping the letters negatively affected transcription performance. One way in which this perceptual grouping effect may increase visual processing time is that when searching for a specific letter on the T9 keyboard, each letter within each group must be visually inspected until the target letter is found. Indeed, it has been shown that time to locate a target letter on a display increases when it is flanked by adjacent irrelevant letters, due to competition for perceptual processing resources [12]. In the present experiment, this added visual scan time may in turn have negated any potential time savings afforded by the larger keys on the T9 keyboard.

## 4 Conclusions

In summary, our findings suggest that the QWERTY keyboard layout remains the most effective of the currently available styles for stylus tapping on soft keyboards. Moreover, given the performance of the three participants that had no prior exposure to the T9, experienced desktop computer users may expect to be able to enter text with an immediate rate of around 23 wpm. It is important to point out however, that although transcription rates in this study were higher for the QWERTY keyboard, performance on the T9 was better, in comparison, than any previously tested alternate key layouts [e.g. 7]. The effects of training on performance was not investigated in this study. However, it has been suggested that text entry rates for the T9 can be expected to reach 25-30 wpm following 1-2 hours of training, and expert rates may reach in excess of 50 wpm [8]. This would imply that with additional training, users may reach proficiency levels with the T9 that are comparable or possibly even exceed those of the QWERTY. Further research is required however to determine the maximum transcription rates of expert T9 users, and the amount of training required to reach expert levels of proficiency.

Finally, this paper described a new technique for comparing text entry performance, called transcription rate, that provides a composite index of speed and accuracy. This technique enables rapid and objective comparisons of different keyboard designs, as well as easy experimental testing of stimulus parameters.

## 5 Acknowledgements

The authors are greatful to Ray Broniak for providing screen shots of the keyboards.

## 6 References

[1] Gibbs, M. Handwriting recognition: a comprehensive comparison. *Pen*, pages 31-35, 1993 (March/April).

[2] MacKenzie I.S., & Chang, L. A performance comparison of two handwriting recognizers. *Interacting with Computers*. (in press).

[3] McQueen, C., MacKenzie, I.S., Nonnecke, B., & Riddersma, S. A comparison of four methods of numeric entry on pen-based computers. *Proceedings of Graphics Interface '94,* Toronto: Canadian Information Processing Society, pages 75-82, 1994.

[4] MacKenzie, I.S., Nonnecke, B., Riddersma, S., McQueen, C., & Meltz, M. Alphanumeric entry on

pen-based computers. *International Journal of Human-Computer Studies, 41,* pages 775-792, 1994.

[5] Perlin, K. Quikwriting: Continuous stylus-based text entry. In UIST '98, San Francisco, CA, pages 215-216, 1998.

[6] Venolia, D. & Neiberg, F. T-Cube: A fast, self-disclosing pen-based alphabet. In *Proceedings of the ACM Conference on Human Factors in Computing Systems, pages 265-270,* 1994.

[7] MacKenzie, I.S. Zhang, S.X., & Soukoreff, R. W. Text entry using soft keyboards. Submitted for publication.

[8] TegicCommunication,
http://www.t9.com/faq.html, 1998.

[9] Legge, G.E., Pelli, D.G., Rubin, G.S., & Schleske, M.M. Psychophysics of reading–I: normal vision. *Vision Research*, 25(2), pages 239-252, 1985.

[10] Chaparro, A., & Young, R., S., L., Reading with rods: the superiority of central vision for rapid reading. *Investigative Ophthalmology & Visual Science, 34(7),* pages 2341-2347, 1993.

[11] Palmer, S.E. Common Region: A new principle of perceptual grouping. *Cognitive Psychology, 24,* pages 436- 447, 1992.

[12] Eriksen, B.A., & Eriksen, C.W. Effects of noise letters upon the identification of a target letter in a non-search task. *Perception & Psychophysics, 16,* pages 143-149, 1974.