An Empirical Study of Typing Rates on mini-QWERTY Keyboards

Edward Clarkson, James Clawson, Kent Lyons, and Thad Starner
College of Computing and GVU Center
Georgia Institute of Technology
801 Atlantic Drive, Atlanta, GA 30332-0280
{edcclark, jamer, kent, thad}@cc.gatech.edu

ABSTRACT

We present a longitudinal study of mini-QWERTY keyboard use, examining the learning rates of novice mini-QWERTY users. The study consists of 20 twenty-minute typing sessions using two different-sized keyboard models. Subjects average over 31 words per minute (WPM) for the first session and increase to an average of 60 WPM by the twentieth. Individual subjects also exceed the upper bound of 60.74 WPM suggested by MacKenzie and Soukoreff's model of two-thumb text entry [5]. We discuss our results in the context of this model.

Author Keywords

Text entry, keypad input, mobile input, mini-QWERTY, thumb keyboard.

ACM Classification Keywords

H.5.2 User Interfaces: Input devices and strategies.

INTRODUCTION

Mobile computing devices' popularity has increased rapidly—some estimates place global mobile phone use at more than 1.52 billion [1]. Along with this rapid expansion, the development of advanced two-way pagers, personal digital assistants (PDAs) and hybrid phone/PDA/pager devices has also contributed to a similar growth in mobile text messaging: 135 billion messages were sent in just the first quarter of 2004 [1].

Considering the amount of text being input on mobile platforms, research has targeted the development and evaluation of mobile text entry methods. One technique is to use a miniature version of the traditional desktop QWERTY keyboard (referred to as 'mini-QWERTY' or 'thumb' keyboards). Despite the presence of mini-QWERTY keyboards in the mobile computing marketplace, there is very little published data on user typing rates with these devices. We present the results of a longitudinal study of novice mini-QWERTY users, the results of which

indicate users can reach typing rates that meet or exceed many other current mobile input technologies. These first-hand observations can provide a comparative basis for future mobile text-input methods.

RELATED WORK

Several examples of commercial mini-QWERTY devices are shown in Figure 1. The RIM Blackberry mobile information device has included a mini-QWERTY keyboard since 1997, and it continues to attract a loyal customer following. The Danger HipTop (also known as the Sidekick) is a similar, newer device that includes a sizeable mini-QWERTY keyboard under a flip-up screen. Nokia has taken a somewhat different approach with its 6800 series of mobile phones. Its front face can flip open to reveal a split mini-QWERTY layout, with the screen set in the middle of the keyboard.

MacKenzie and Soukoreff have created a theoretical model of two-thumb text entry on miniature keyboards [5]. Using English language letter frequency distributions and Fitts' Law calculations, they predicted a peak expert rate of 60.74 WPM on mini-QWERTY keyboards. A sensitivity analysis of the model to various parameters (e.g., Fitts' Law coefficients, word corpus effect, etc.) yielded no more than a +/-10% variation from their original figure. Validation of the model was left for future work.

Researchers at Canesta, Inc. reported a study that included mini-QWERTY typing speeds at CHI 2003 [8]. In evaluating their virtual projection keyboard, they tested it against a desktop QWERTY keyboard, Grafitti pen input and a thumb keyboard. They recruited 11 subjects who used each method in random order, typing a phrase repeatedly for 2 minutes. Subjects achieved an average of 27.6 WPM on the thumb keyboards, 64.8 WPM on the conventional keyboard, 46.6 WPM on the Canesta keyboard, and 14.0 WPM with Grafitti. The authors state their subjects included both novice and expert Canesta keyboard users but do not mention participants' experience with any of the other input devices. For a more comprehensive review of mobile input technologies, we direct the reader to the review of text entry techniques found in MacKenzie and Soukoreff [6].

Figure 1. Nokia 6820 (top), RIM Blackbery (bottom left) and Danger/T-Mobile Sidekick (bottom right).

RESEARCH METHOD

A number of aspects of our design are directly influenced by other work. In particular, the compensation arrangement and session structure follow those in Lyons, *et al.* [3].

Participants

We recruited 21 subjects, none of whom had ever used a mini-QWERTY keyboard more than once before. Each subject was randomly assigned one of two different mini-QWERTY keyboards to use throughout the study (Dell and Targus branded, discussed further below). Similar to the Lyons, *et al.* study [3], participants were compensated at a rate proportional to their WPM x Accuracy over the entire session: $0.125 x WPM x Accuracy, with a $4 minimum for each session. Participants were asked to complete 20 twenty-minute sessions over the course of 11 days. Four subjects did not complete all 20 sessions, leaving 10 subjects in the Targus group and 7 in the Dell. We excluded the data from the last three subjects in the Targus group in order to match the number of subjects in the Dell group. This change allows an analysis of variance (ANOVA) calculation, and all statistics reported are drawn from this smaller sample. Excluding the three subjects' data resulted in a mean change of less than 1 WPM on the last session. Our final sample size was 14 subjects (7 in each group), all ages 18-34. Six participants were female, eight male, and all but one was right-handed.

Equipment and Software

We used two mini-QWERTY keyboards in the experiment, one by Dell (for the Dell Axim) and one by Targus (for the Palm m505). We modified each keyboard to connect to a standard desktop computer serial port. The Dell and Targus keyboards transmit at 4800 and 9600 baud, respectively. The letter keys on both keyboards are oval-shaped as shown in Figure 2. The Dell keyboard has a single space key in the middle of the bottom key row, while the Targus has two

triangular space keys set below the rest of the keys. The Targus keys measure 6.73 mm. along their major axes and 4.83 mm. along their minor axes. The Dell keys are 5.99 x 4.06 mm. (by comparison, the Danger Sidekick keys are approximately 8 x 6 mm.). We chose to use two similar keyboards to test more than one point in the keyboard design space.

The study occurred in our usability lab with each of the two keyboards connected to a separate Pentium III workstation. We employed the Twidor software package (used in the series of studies on the Twiddler chording keyboard [2, 3]) and adapted it to accept data from our modified keyboards. The Twidor software was configured to use the MacKenzie and Soukoreff phrase set [7], a set of 500 phrases representative of the English language. The phrases range from 16 to 43 characters with an average length of 28 characters. The phrase set was modified to use only American English spellings and display only lower case letters and spaces (no punctuation or capitalization). Each test phrase was shown on the screen and remained there while the subject typed.

Figure 2. The modified Targus (top) and Dell (bottom) keyboards.

Design and Procedure

We structured the study as a 2 x 20 factorial mixed design, with the keyboard as the between-subjects factor and the 20 sessions as the within-subjects factor. Sessions were completed in pairs with a 5-minute break after the first 20-minute session. Each session pair was separated by at least two hours and no more than two days. As mentioned, each subject was randomly assigned a keyboard that was used throughout his or her participation.

Each session was preceded by a warm-up phase, which consisted of the phrases "abcd efgh ijkl mnop" and "qrst uvwx yz" repeated twice. The warm-up phase was not counted in the session statistics. The remainder of the session consisted of a number of trial blocks, containing ten randomly selected phrases each. Each participant completed as many blocks as he or she could in a twenty-minute session. The subjects were instructed to type using only their two thumbs and to type as quickly and accurately as possible. The test software provided statistical feedback in the form of WPM and accuracy data for the most recent sentence typed and the current session average.

In addition to mini-QWERTY rates, we also collected desktop QWERTY typing speeds averaged over 20 phrases at each subject's first and last sessions. Participants completed a demographics survey before the first session and a debriefing survey after the end of the last session.

RESULTS

The 14 participants typed 33,945 sentences across all sessions, encompassing over 950,000 individual characters. Session statistics are weighted by the number of characters in each sentence, and error rates were calculated using Soukoreff and MacKenzie's total error metric [10]. Averaged over both keyboards, participants had a mean first session typing rate of 31.72 WPM ($\sigma = 7.00$) and a session twenty mean of 60.03 WPM ($\sigma = 8.40$). The average error rate for session one was 6.12% ($\sigma = 3.46\%$),

rising gradually to 8.32% ($\sigma = 4.13\%$) by session twenty. The slight upward trend in error rates is consistent with our previous mobile text entry research [3].

Learning Rates

We derived an exponential regression curve for each participant and the group averages. The per-group data correlated closely with the derived curves ($R^2 > 0.98$; see Figure 4). Power curves fitted to individual users showed high correlation, with R^2 values ranging from 0.76 to 0.97.

Keyboard Differences

The Targus group typed faster than the Dell group. The mean session one rates were 34.33 WPM ($\sigma = 8.15$) on the Targus and 29.12 WPM ($\sigma = 4.88$) on the Dell; session twenty average speeds were 61.44 ($\sigma = 8.16$) on the Targus and 58.61 WPM ($\sigma = 9.03$) on the Dell (see Figures 3 and 4). We performed a two-way analysis of variance (ANOVA) with the session number and the keyboard type as our factors. The ANOVA showed significant effects for both the session and keyboard variables with extremely high confidences ($p < 0.0001$). Interaction between the two was not present ($p > 0.9999$). These results indicate there was a statistically significant difference between the two keyboards—the Targus was faster than the Dell—and, unsurprisingly, typing speeds improved over time.

The mean desktop QWERTY typing speed for the Targus group was 84.99 WPM ($\sigma = 19.84$) before the study and 98.31 WPM ($\sigma = 19.25$) at the end and 77.38 WPM ($\sigma = 15.03$) and 86.79 WPM ($\sigma = 14.20$) at the start and end for the Dell. We measured the correlations of the starting desktop QWERTY speeds with session one mini-QWERTY rates, and the end desktop QWERTY speeds with session twenty mini-QWERTY rates. The correlations were mild ($R^2 = 0.31$ and 0.57, respectively). The increase in desktop QWERTY speeds may be due to increased familiarity with the test environment and is consistent with previous work [3].
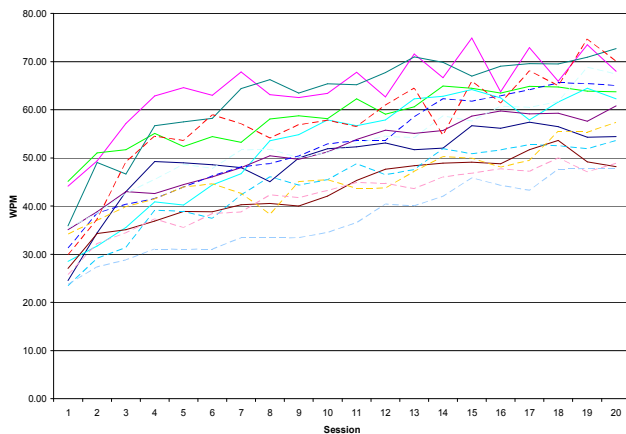

Figure 3. Individual session average WPM curves. Dell curves are dashed; Targus are solid.
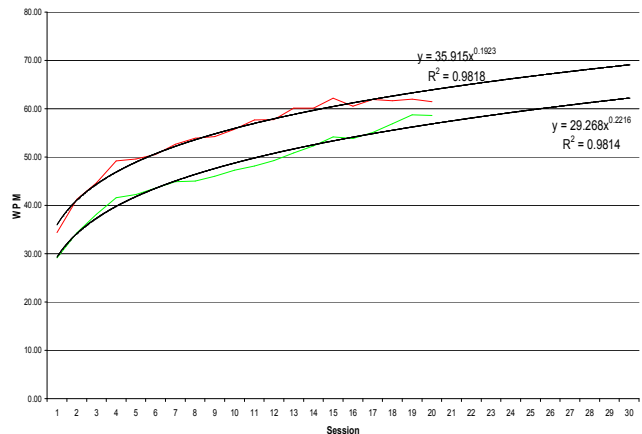

Figure 4. Per-group session average WPM with regression curves, equations and correlations. Dell curves are on the bottom; Targus curves on top.

Survey Results

After the end of the last session, each participant completed a debriefing survey, answering a number of free-form and 7-point Likert scale questions regarding how they used the keyboards and their comfort level. Users found the mini-QWERTY keyboards marginally comfortable overall (mean 4.00, $\sigma = 1.41$) and much less comfortable than a full-size keyboard (mean 2.29, $\sigma = 0.99$). There was no statistical difference in comfort ratings between the two keyboards ($p = 0.36$). Subjects also reported they were less likely to look at the screen after the last session than at the start; however, this was a marginally significant result ($p = 0.056$). This result conforms to some anecdotal responses (e.g., "Not looking at the screen vastly improved my speeds!").

DISCUSSION

MacKenzie and Soukoreff reported a predicted peak expert mini-QWERTY rate of 60.74 WPM. We cannot make direct comparisons of our results to their model, since the model prediction was derived from a keyboard with significantly different properties. To enable an accurate comparison, we replicated MacKenzie and Soukoreff's calculations [5, 9] using measurements from our Dell keyboard. We used the Dell as our reference since it most closely resembled the model assumptions. The resulting calculations on the MacKenzie phrase set yielded a model prediction of 57.89 WPM. Three subjects in the Dell group exceeded the two-thumb model's predicted maximum for our keyboard on at least one session. A total of 23 Dell sessions exceeded the predicted speed, and the single fastest session speed was 74.69 WPM.

One potential reason for the empirical deviation from the model relates to key width measurement. MacKenzie and Soukoroff noted that the relative width of the thumb could confound the width measurement. In addition, our keyboards are built with oval keys with major axes inclined 45° from horizontal, putting the major axis of each key parallel to the natural movement of the thumb. As a result, using the width of the keys along the line of approach (denoted W' [4]) may be more appropriate for similar keyboards.

FUTURE WORK

Several avenues exist for future work in this area and on the data we have presented in particular. We wish to fully reconcile the two-thumb model with our empirical results. Determining what key width measurement method is most appropriate for mini-QWERTY keyboards is necessary, as is clarifying the degree to which thumb size confounds those measurements. Other model assumptions to consider include the minimum time required between key presses and the Fitts' Law coefficients for our keyboard.

Our anecdotal and qualitative data suggest that users increase the time they spend looking at the keyboard as they gain more experience. Yet the ability to type effectively without looking at the input device ("blind typing") can be beneficial when mobile; we wish to examine how well our experts can type without visual feedback.

CONCLUSIONS

We have presented a longitudinal study of mini-QWERTY keyboard use, finding significant differences between our two keyboard groups and expert typing rates exceeding 60 WPM. These rates compare well with other mobile devices in terms of both novice and expert typing rates, and provide a performance benchmark for further research in this area.

ACKNOWLEDGMENTS

REFERENCES

1. CellularOnline, "Latest Mobile, GSM, Global, Handset, Base Station, & Regional Cellular Statistics". http://www.cellular.co.za/, September, 2004.
2. Lyons, K., Plaisted, D. and Starner, T., Expert Chording Text Entry on the Twiddler One-Handed Keyboard. In *Proc. ISWC '04*.
3. Lyons, K.*, et al.*, Twiddler Typing: One-Handed Chording Text Entry for Mobile Phones. In *Proc. CHI '04*.
4. MacKenzie, I. S. and Buxton, W., Extending Fitts' Law to Two-Dimensional Tasks. In *Proc. CHI '92*, pp. 219-226.
5. MacKenzie, I. S. and Soukoreff, R. W., A Model of Two-Thumb Text Entry. In *Proceedings of Graphics Interface 2002*, pp. 117-124.
6. MacKenzie, I. S. and Soukoreff, R. W. Text Entry for Mobile Computing: Models and Methods, Theory and Practice. *Human-Computer Interaction,* Vol. 17, pp. 147-198.
7. MacKenzie, I. S. and Soukoreff, R. W. Phrase Sets for Evaluating Text Entry Techniques. In *Proc. CHI '03*, pp. 754-755.
8. Roeber, H., Bacus, J. and Tomasi, C., Typing in Thin Air: The Canesta Projection Keyboard - a New Method of Interaction with Electronic Devices. In *CHI '03 Extended Abstracts* pp. 712-713.
9. Soukoreff, R. W., *Text Entry for Mobile Systems: Models, Measures, and Analyses for Text Entry Research*. M. Sc. Thesis, Department of Computer Science, York University, 2002.
10. Soukoreff, R. W. and MacKenzie, I. S., Metrics for Text Entry Research: An Evaluation of MSD and KSPC, and a New Unified Error Metric. In *Proc. CHI '03*, pp. 113-120.