EVALUATING MOBILE TEXT ENTRY WITH THE FASTAP™ KEYPAD

Andy Cockburn and Amal Siresena
Department of Computer Science
University of Canterbury
Christchurch, New Zealand
{andy, ans26}@cosc.canterbury.ac.nz

ABSTRACT

This paper describes the evaluation of a new keyboard interface for mobile devices. The Fastap™ keypad is designed to allow one character per key-press input on mobile devices without requiring additional hardware (such as a plug-in keyboard) or a stylus-based input display. Results of the evaluation show that Fastap is both efficient and rapidly learnable when compared with the current industry standard methods for text messaging.

Keywords

Mobile text entry, evaluation, novice and expert use.

1. INTRODUCTION

More than 24 billion short-message-service text messages are sent from mobile devices every month[1]. Despite this heavy use, the text entry methods used by current mobile phones are frustratingly slow and hard to learn. These problem are due to the standard ISO phone layout having only 12 keys ('0'-'9', '#' and '*') to input the entire alphabet and all of the punctuation and numerical characters. Each physical key is therefore overloaded with three or four alphabetical characters: for example, the digit '2' is overloaded with 'A', 'B', 'C'. Consequently, the user or the system must determine the intended character for each keypress.

The most common user-driven technique is called 'multi-tap' in which successive presses of the same key determine the intended letter. For example, to enter 'CAB' the user would press '222' (to get the third letter on the '2' key), then '2', and finally '22'. The break between two letters entered with the same key is normally indicated by a pause.

Multi-tap input schemes have been largely superceded by predictive schemes, such as the widely used T9[2] method. With T9, the device displays the most likely word for the sequence of keys pressed since the last space character. If the predicted word is incorrect, the user scrolls through the alternative dictionary words made from the same key sequences. If the word is not in the dictionary, the user must delete their input and enter the word using a technique such as multi-tap.

This paper describes the evaluation of a new mobile keypad that removes the need for key disambiguation without relying on stylus-based interaction or plug-in expander keyboards. The Fastap[3] keypad, shown in Figure 1, uses small raised keys (primarily used for text) that surround recessed keys (primarily used for numbers). Although the individual keys are small, they allow a contact area that is similar to that of a full-sized keyboard [5]. This is achieved through two techniques. First, the raised letter keys are carefully spaced to allow a large contact area for the small key. Second, "passive chording" is used to select the recessed numerical keys. The simultaneous depression of multiple keys is transparent to the user who simply presses over the target key. In this way, Fastap allows exactly one-press per character for the alphanumeric character-set.

Our evaluation compares the data input rates of Fastap with those of multi-tap and T9 across three stages of use: initial user reaction, novice use, and expert use. Subjective preferences and workload estimates are also compared.

2. RELATED WORK

HCI research into mobile text entry has focussed on two different input technologies. First, there have been several developments based on stylus and gesture-based input schemes. Examples include the Unistrokes gestural alphabet [2] and Dasher [9] which predicts probable characters and words while allowing gesture-based selection from alternatives. The main disadvantages of gesture-based schemes are the high training demands and the need for sensing displays seldom found on mobile phones.

There has also been extensive research into improving the efficiency of text input using the ISO standard mobile-phone keypad. Silfverberg et al [8] described a theoretical model to predict multi-tap and T9 text entry rates. They predicted very high text entry rates—21 words per minute (wpm) for multi-tap and 40.6wpm for

[1] Source GSM Association www.gsmworld.com.

[2] T9, trademark of Tegic Communications www.tegic.com

[3] Fastap, trademark of Digitwireless: www.digitwireless.com

Figure 1:The Fastap prototype phone.

T9. One expert user was able to attain input speeds similar to the predicted values (21.0 and 32.9wpm). Dunlop and Crossan [1] predicted slower rates of 14.9wpm and 17.6wpm respectively for multi-tap and a predictive scheme similar to T9, yet their 14 non-expert empirical participants achieved only 5wpm and 5.45wpm for multi-tap and the predictive scheme[4]. James and Reischel [4] followed up Dunlop and Crossan's work showing expert text entry rates of 7.93wpm and 20.36wpm with multi-tap and T9.

Finally, MacKenzie [6] describes a 'keystrokes per character' (KSPC) metric for analysing text entry methods. KSPC characterisations can be used to aid the design of text-entry mechanisms prior to labour-intensive implementation and empirical evaluation. They can also be used to predict text entry rates for experts using each method. This theoretical measure does not, however, provide insights into the learnability, perceived workload, or subjective satisfaction with each method.

3. EVALUATION

Our evaluation compares data input rates and workload measures for three mobile text entry methods: multi-tap, T9 and Fastap. These measures are compared across different levels of user expertise and different task types (dialing a number, text messaging, and so on). The mobile phones used in the experiment were an Ericsson T10s (used for multi-tap input), a Nokia 8260 (used for T9), and three identical Fastap prototypes. All tasks were videotaped, and task completion times and error rates were measured through video transcription.

Thirty-six paid participants took part in the experiment, with twelve randomly assigned to each of the three interface types. Although all participants carried out the initial reaction and novice user tasks, data was only analysed from twenty-six participants: twelve with Fastap, and seven each with multi-tap and T9. Data from the other participants was discarded after screening for prior mobile phone experience. Twenty-five participants continued to the expert evaluation phase: the top ten Fastap users, the top eight T9 users, and the top seven multi-tap users. Each expert participant was paid to train with their assigned interface for sixty minutes, spread across six days of daily ten-minute training sessions.

For the initial reaction task, participants entered two text messages without any instruction: "i bought a cellphone" and "033667001". The phones were preset to their text-messaging mode. A two-minute time limit was placed on the task.

After the initial reaction tasks the participants were trained in their assigned interface for five minutes, with dummy tasks. They then proceeded to the novice evaluation, which involved entering three text messages from each of four sentence categories, as follows. *Traditional* sentences contained no capitals, punctuation or non-dictionary words: e.g. "i will be home later". *Numerical* sentences contained typical phone numbers: e.g., "call 039833298". *Non-dictionary* sentences included proper nouns and capital letters: e.g., "See you at Waikuku beach". *Abbreviated* sentences made extensive use of non-dictionary abbreviations: e.g., "we can tlk 2moro". Each participant's sentences were randomly selected from twenty-four sentences in each category.

The expert evaluations ran identically to the novice evaluations, but after the six ten-minute training sessions were complete. During the training sessions the participants entered sentences of the four categories as fast as possible. The training sentences were different to the evaluation sentences.

Dependent measures were text entry rates and number of errors. Subjective measures were recorded using the NASA Task Load Index (TLX) Worksheets [3]. Data entry rates are measured using words per minute, based on 5.98 characters per word (including the terminating space), as in previous studies.

4. RESULTS

4.1 Initial Reaction

There were dramatic differences between the three interfaces in the initial reaction tasks. All twelve of the Fastap users successfully entered the text "i bought a cellphone", with a mean text entry rate of 6.3wpm (standard deviation 1.6). Six of the multi-tap users completed the task within the two-minute time limit, with a mean entry rate (for the six who completed) of 3.6wpm (s.d. 1.3). Only four of the seven T9 users completed the task, with a mean text entry rate of 3.9wpm (s.d. 1,7). The three T9 users who exceeded the two-minute time limit expressed confusion and astonishment when the previously entered letters changed on subsequent keypresses.

[4] Dunlop and Crossan did not report these values. They are calculated from Figure 5 in their paper.

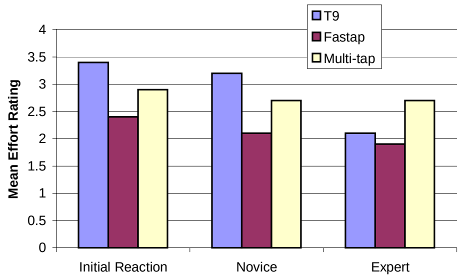

Figure 2: Mean NASA-TLX effort ratings by interface and experience level.

The second initial reaction task involved entering the number "033667001" in text entry mode. Again, all Fastap users successfully completed the task, with a mean entry rate of 7.5wpm (s.d. 2.3). Only three of the multi-tap users completed the task within two minutes, with a mean rate (for those completing) of 2.0wpm (s.d. 1.5). Similarly, only four of the T9 users completed the task: mean 2.2wpm (s.d. 1.3).

Subjective responses to the NASA-TLX questions echoed the performance measures, with users rating the demands of Fastap lowly, slightly higher for Multi-tap, and dramatically higher for T9. Workload assessments for the three interfaces, averaged across the six NASA-TLX measures, are shown in Figure 2.

The high failure rates for multi-tap and T9 indicate that these interfaces rely heavily on user instruction. Fastap, however, is immediately comprehensible.

4.2 Novice Use

Text entry rates for the novice and expert user tasks were analysed using a 3×4 mixed-factors analysis of variance (ANOVA). The first factor was interface-type (between subjects), with three levels: Fastap, Multi-tap and T9. The second factor was sentence-type (within subjects), with four levels: traditional, numerical, non-dictionary, and abbreviated.

ANOVA showed a significant main effect for interface-type ($F_{2,23}$=11.3, p<.01), with mean rates of 7.1wpm (s.d. 2.3), 5.0wpm (s.d. 3.3) and 3.8wpm (s.d. 1.3) for Fastap, T9 and multi-tap respectively. As expected, rates for the four sentence types were reliably different ($F_{3,69}$=12.4, p<.01), with traditional sentences fastest and abbreviated ones slowest.

There was a significant interaction between factors interface-type and sentence-type ($F_{6,69}$=4.8, p<.01), indicating that the relative performance of the interfaces changed across sentence type. The interaction is depicted in Figure 3a, which shows that Fastap and Multi-tap performed relatively constantly across sentence types compared to T9, which performed well with traditional sentences, but very poorly with non-dictionary ones.

Again, the subjective responses to the NASA-TLX questions echoed the performance measures, with users reporting lower workloads with Fastap than Multi-tap and T9 (see Figure 2).

4.3 Expert Use

Results in the expert tasks reflect those of novice use, with significant main effects for both factors, and a significant interaction between them: see Figure 3b. Mean text entry rates with Fastap, T9 and multi-tap were 8.5wpm (s.d. 2.2), 8.2wpm (s.d. 3.9) and 4.8wpm (s.d. 1.4): $F_{2,22}$=8.8, p<0.01.

Comparing the novice and expert rates indicates that training was particularly valuable to T9 users who's mean performance increased by 64% from 5.0wpm to 8.2wpm. Fastap users, in contrast, improved by only 20%: from 7.1wpm to 8.5wpm. Multi-tap users improved by 26% from 3.8wpm to 4.8wpm. Comparing Figures 3a and 3b indicates that training with T9 was particularly important for non-traditional sentences.

NASA-TLX workload measures had decreased from the initial and novice user levels, particularly for the T9 users. Ratings averaged across the NASA-TLX categories were 1.9 for Fastap, 2.1 for T9, and 2.7 for multi-tap (see Figure 2).

5. DISCUSSION

The results are encouraging for Fastap. They indicate that Fastap offers three main advantages over Multi-tap and T9. First, it is immediately usable without any training. This rapid learnability contrasts dramatically with T9, which confused almost all participants. Second, Fastap allows any alpha-numeric data to be input within the same interface mode and at roughly the same text-entry rate. Again, this is in dramatic contrast to T9, which requires a Multi-tap input mode for non-dictionary text. Furthermore, the Multi-tap mode is prone to 'stickiness', with many participants forgetting to return to the T9 mode after entering a non-dictionary word. Third, Fastap's subjective workload measures were lower than Multi-tap and T9, particularly for new and novice users.

Fastap's advantages are most apparent during the early stages of use, but they are not gained at the cost of expert performance. For the fast expert-traditional condition, the difference between T9 (10.8wpm) and Fastap (9.3wpm) is fairly small, and both are dramatically ahead of Multi-tap (5.6wpm). Furthermore, Fastap was dramatically faster with abbreviated sentences (T9's 5.1wpm compared to Fastap's 8.6wpm). This result is potentially important. Given the frustration of using any mobile text entry device, users will continue to use extreme abbreviation. Word-based predictive schemes such as T9 can respond by adding common abbreviations to their dictionaries, but creative abbreviations will remain unpredictable. Prefix-based systems (e.g. LetterWise[5]) disambiguate between overloaded letters based on the frequency of letter sequences in a language. When the

[5] Eatoni Ergonomics www.eatoni.com.

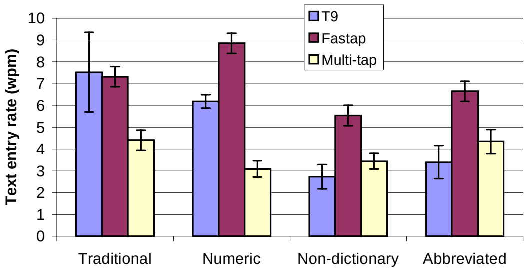

Figure 3a: Novice users.

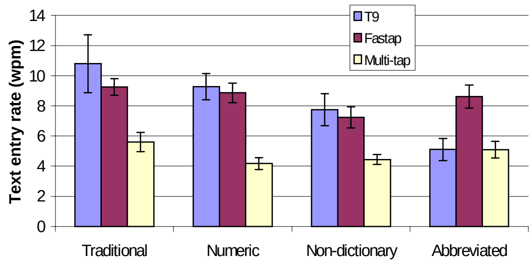

Figure 3b:Expert users.

Figure 3: Mean text entry rates for the three interfaces.

system predicts an incorrect letter, the user presses a ‘Next’ key to advance to the next letter on the key. This technique degrades gracefully towards multi-tap performance as the user enters obscure abbreviations, but it avoids the abrupt modal changes of T9. Digitwireless are developing an analogous word-based prediction scheme for Fastap. Combining Fastap’s one-keypress per character with word prediction should dramatically increase its performance. Furthermore, the absence of keypress disambiguation in Fastap means that previously entered characters will not change, avoiding the confusion experienced by those new to T9.

Our results show that Fastap compares favourably with the dominant schemes currently used for text entry on mobile phones. However, market forces are driving the rapid development of new and improved interfaces for mobile text entry. Top-of the line mobile phones blur the distinction between mobile telephony and palm-top computing, introducing relatively sophisticated display capabilities such as stylus-driven input. Several previous evaluations have shown that miniaturised QWERTY keyboards, such as those provided by the Nokia Communicator 9210i[6], can allow much faster text input than the standard ISO keypad layout [7]. Similarly, gesture-based input mechanisms such as Unistrokes [2] and Dasher [9] have been shown to allow relatively high input speeds with practice. Despite these opportunities, most mobile phones currently support the standard ISO-format keypad, with device cost presumably playing a major role in customers’ selection. Fastap offers increased performance over an ISO keypad while maintaining the cost and size parameters of the current mainstream devices, but it remains to be seen how consumers will assess the trade-off between the increased performance of advanced display/input technologies and their additional cost.

6. CONCLUSIONS

The Fastap keypad places 52 independent keys onto an area the same size as a standard ISO keypad. Through careful design, the contact area for each key is similar to that of a full-size computer keyboard. Our evaluation showed that Fastap users very quickly acquire a comparatively high level of performance, while T9 users struggle until well trained. Although T9 allows fast text entry rates of dictionary words after training, Fastap was faster overall for a realistic mixture of dictionary, non-dictionary, abbreviated and numeric use. The addition of word-based prediction schemes is likely to further enhance Fastap’s efficiency.

REFERENCES

[1] M. Dunlop and A. Crossan, "Predictive Text Entry Methods for Mobile Phones," *Personal Technologies*, vol. 4, pp. 134-143, 2000.

[2] D. Goldberg and C. Richardson, "Touch-Typing With a Stylus," Proc. INTERCHI'93, Amsterdam, 1993.

[3] S. Hart and L. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Human Mental Workload*, M. Hancock, N, Ed., 1988, pp. 139--183.

[4] C. James and K. Reischel, "Text Input for Mobile Devices: Comparing Model Prediction to Actual Performance," Proc. CHI'2001, Seattle, 2001.

[5] D. Levy, "The Fastap Keypad and Pervasive Computing," Proc. Int. Conf. on Pervasive Computing, Zurich, Switzerland, 2002.

[6] I. MacKenzie, "KSPC (keystrokes per character) as a characteristic of text entry techniques," Proc. Fourth International Symposium on Human-Computer Interaction with Mobile Devices, 2002.

[7] I. MacKenzie, S. Zhang, and R. Soukoreff, "Text Entry Using Soft Keyboards," *Behaviour and Info. Technology*, vol. 18, pp. 235--244, 1999.

[8] M. Silfverberg, I. MacKenzie, and P. Korhonen, "Predicting Text Entry Speed on Mobile Phones," Proceedings of CHI'2000, April 1--6, 2000.

[9] D. Ward, A. Blackwell, and D. MacKay, "Dasher---A Data Entry Interface Using Continuous Gestures and Language Models," Proc. ACM UIST, San Diego, California., 2000.

[6] www.nokia.com/nokia/