Comparing the Immediate Usability of Graffiti 2 and Virtual Keyboard

Thomas Költringer, Thomas Grechenig
Research Industrial Software Engineering, Vienna University of Technology
Operngasse 9, A-1040 Vienna, Austria
Thomas.Koeltringer@rise.tuwien.ac.at

Abstract
This paper presents the results of an empirical study on the input system of the most frequent PDA operating system, PalmOS from Palm Inc. In an experiment with novice users we compared the stroke based alphabet Graffiti 2 with the Virtual Keyboard and the predictive add-on WordComplete from CIC Software for Graffiti 2. We found that although text input with Graffiti 2 was significantly slower and generated a higher error rate (9 wpm; 19%) than text input with the Virtual Keyboard (13 wpm; 4%), there was no significant difference in usability and task load rating. WordComplete for Graffiti 2 had no significant impact on performance but enhanced user comfort.

Categories & Subject Descriptors:
H.5.2 [Information Systems]: User Interfaces - *Input devices and strategies (e.g., mouse, touchscreen).*

General Terms:
Human Factors, Measurement, Performance.

Keywords:
Mobile computing, Pen-based computing, Stylus input, Graffiti 2, Virtual Keyboard, WordComplete.

INTRODUCTION

Mobile devices like mobile phones and personal digital assistants (PDAs) have become very popular and widely used consumer products. In ever shorter time intervals new and allegedly improved devices are launched on the market, often introducing - and therefore forcing users to accustom to - new methods of interaction.

For many years now we have rather contentedly been using PDAs based on the PalmOS, from Palm Inc., the world market leader since 1996. One of the success factors of Palm Inc. was the easy-to-use pen-based text entry method called Graffiti. Graffiti is a single stroke alphabet which resembles the Roman alphabet and is based on Unistroke [3]. It includes letters, numbers, special characters and backspace. To achieve higher accuracy, the area in which Graffiti has to be entered is separated into two sections, the left side for alphabetic and the right side for numeric characters. In various studies Graffiti has proved to be an efficient text entry method [2, 6].

In January 2003 Palm introduced the advanced - and Palm claims improved - Graffiti 2 alphabet [2] which is based on the CIC-Software Jot alphabet. Besides other changes, Graffiti 2 offers more intuitive entry of accents and umlauts and more consistent entry of special characters. The most important change, however, is that with Graffiti 2 some letters (I, K, T and X) have to be written with two strokes.

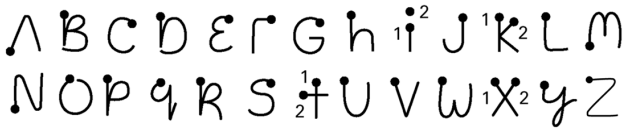

Figure 1. Graffiti 2 alphabet [7]

PalmOS also provides an alternative input method, the on-screen QWERTY keyboard with switchable layouts for entering numbers and other special characters.

To allow easier and faster input on PDAs based on PalmOS CIC Software introduced WordComplete, a predictive add-on for Graffiti (2) and the Virtual Keyboard. When entering a word, after [n] letters (default: n=2) a list of predicted words pops up, from which the user can choose one to complete the input.

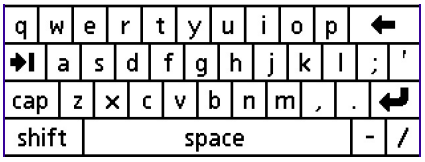

Figure 2. Virtual Keyboard in PalmOS 5.2

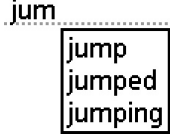

Figure 3. WordComplete 3.0

Which of the described state-of-the-art text input methods works best? We wanted to find out how efficient and usable the new Graffiti 2 is, especially compared to the Virtual Keyboard.

RESEARCH QUESTIONS

The primary research question concerned the immediate usability of the text input systems provided by PalmOS, Graffiti 2 and Virtual Keyboard. We wanted to examine which method is faster and more accurate and is rated higher in usability and task load.

We hypothesized that users enter text slower and make more mistakes using Graffiti 2 than using the Virtual Keyboard. This hypothesis is also supported by the work of Fleetwood et al. (2002) [2], who have evaluated Graffiti 1 in a similar study. Because the cognitive effort (learning, memory) with Graffiti 2 is higher than with Virtual Keyboard we assumed that the immediate usability of Virtual Keyboard is higher. Correspondingly the task load of Graffiti 2 is expected to be higher. We decided to use different text types (normal phrases, numbers and commands), because we expected numbers to be generally easier and commands to be more difficult to enter than normal text, independent from the input method. For the text type "normal phrases" we also examined the effects of using the add-on WordComplete for Graffiti 2.

METHOD

Participants

We recruited 12 voluntary subjects, including students, university employees, research staff and consultants. There were two female and ten male participants and the average age was 27.17 years (SD=4.91). All subjects had basic skills in text input on mobile devices. Ten subjects specified that they had basic experience (they had used a friends PDA one or two times etc.) entering text using Graffiti 1 and the Virtual Keyboard. None of the subjects had experience with Graffiti 2 or WordComplete.

Apparatus and Software

Emulator

The PDA environment was emulated using a Palm Pilot emulator (PalmSource Palm OS 5.2 Simulator) and a graphic tablet on a standard desktop PC platform. Additionally, WordComplete 3.0 (DEMO) from Communications Intelligence Corp. was installed. All software versions were in English language (including the WordComplete dictionary) and all options were set to standard. On the screen of the 19 inch display (1280x1024, 32bit colors) the emulator occupied 640x910 pixels (184x240 millimeter) and the workspace background was set to black. The input was performed with an Aiptek Hyper Pen 5000U, a graphic tablet with a pressure sensitive pen in absolute mode. The workspace on the tablet measured 127x102 millimeter and the maximum resolution was 120 lines per millimeter (3048 lines per inch).

Logging

A dual-head graphics card was used to produce a copy of the display for a video cassette recorder. The entered characters and their timestamps were then manually extracted from the tape and analyzed using the 'analyse34' AWK-script by William Soukoreff[1].

Text Phrases

For every text type three examples were generated.

1. irregular verbs are the hardest to learn.
2. this is the biggest hamburger i have ever seen.
3. the quick brown fox jumped over the lazy dog.
4. type /mydocuments/masterthesis.pdf
5. mailto://thomas.koeltringer@gmx.net
6. http://www.tuwien.ac.at/pr/events
7. 841320734687433
8. 5049832761
9. 544391463337740

The normal text phrases (1, 2, 3) have been taken from MacKenzie and Soukoreff (2003) [5] and represent simple grammatically correct English sentences ($r=0.8199$[2] correlation with the English language). Upper case letters have been removed. The phrases in the command category are a UNIX command (4), an E-Mail address (5) and a URL (6). The phrases in the numbers category (7, 8, 9) have been generated randomly. Entering text with Graffiti 2 + WordComplete was only reasonable with the input phrases of normal text, because WordComplete can only predict words.

Procedure

First, every participant received a short introduction on how the experiment would be conducted. Before every input task (grouped by input method) a learning phase was accomplished. In this learning phase every subject entered the alphabet from A to Z, followed by the numbers 0 to 9. Special characters like '.', ':', '/', '@' and space as well as the commands return and backspace were trained. The participants had to re-enter the most problematic characters until they got them right. The learning phase ended by entering the sentence 'elephants are afraid of mice'.

Every input phrase was printed on a piece of paper and presented to the user. The user was requested to read the presented phrase carefully and keep it in mind to keep the side effects of reading minimal. Moreover, the user was briefed to input the phrases efficiently and preferably without errors. After entering all phrases per method, the subjects answered the SUS and NASA-TLX questionnaires concerning the input method they had just used.

After entering the input phrases with all three input methods the users had to enter the alphabet (A to Z) with Graffiti 2 five times. The users were told not to correct characters which the system didn't recognize correctly in order to identify the problematic characters.

At the end of the test every participant was interviewed to figure out the user's personal preferences and problems with the input methods.

---

[1] William Soukoreff; The Human Computer Interaction Lab at York University; Toronto, Ontario; Canada; will@acm.org

[2] Calculated with http://www.yorku.ca/mack/AnalysePhrases.zip

Design

The empirical study measured the performance of the different pen-based text input systems by variation of the text type. The independent variables were the three input methods (Graffiti 2, Graffiti 2 + WordComplete, Virtual Keyboard) and the three text types (normal phrases, numbers, commands). The design was a within-subject design with repeated measures. Before changing the input method a learning phase was introduced in order to keep memory effects low. The subjects were randomly assigned to the nine test conditions. Usability (measured by System Usability Scale – SUS [1]) and task load (measured by the NASA Task Load Index – TLX [4]) were evaluated after every change of input method. The SUS ranging from 0 (low) to 100 (high) is a 10-item questionnaire that gives an overview of satisfaction with the system. The NASA-TLX is a multi-dimensional rating procedure ranging from 0 (low) to 600 (high) that derives a workload score based on a weighted average of ratings on six subscales. These subscales are mental demand, physical demand, temporal demand, own performance, effort and frustration. After the test the subjects answered qualitative questions and gave personal feedback regarding the test.

RESULTS

The collected data was analyzed using the new error metric introduced by Soukoreff and MacKenzie (2003) [8].

Table 1. Entry Speed and Total Error Rate

	Graffiti 2	Virtual Keyboard	Graffiti 2 + WC
text	9.24 wpm	13.64 wpm	8.22 wpm
	19.35 %	4.11 %	18.82 %
numbers	7.52 wpm	14.55 wpm	-
	21.07 %	1.69 %	
commands	6.64 wpm	10.25 wpm	-
	24.95 %	5.23 %	

Speed

Entering text with Graffiti 2 was significantly faster than numbers ($t(11)=2.719$, $p<0.05$; calculated with the t-test) and commands ($t(11)=4.427$, $p<0.01$). Commands were the slowest to enter (numbers-commands: ($t(11)=3.232$, $p<0.01$). For the Virtual Keyboard normal text and numbers showed no statistical difference. Here, too, commands were the slowest to enter (numbers-commands: ($t(11)=6.551$, $p<0.01$), normal text-commands: ($t(11)=6.391$, $p<0.01$).

In all three text categories input with the Virtual Keyboard was significantly faster than with Graffiti 2 (normal text: $t(11)=5.782$, numbers: $t(11)=7.782$, commands: $t(11)=7.783$, $p<0.01$;). Using the add-on WordComplete did not significantly increase the input speed of the users.

Total Error Rate

Entering numbers and normal text resulted in equal error rates, but the errors made with commands were significantly higher ($t(11)=3.299$, $p<0.01$). For WordComplete the average total error rate was not significantly different to Graffiti 2. With the Virtual Keyboard commands and normal text have no significant deviation in error rate, and numbers are the text type with least errors (numbers-commands: $t(11)=3.302$, $p<0.01$; numbers-normal: text $t(11)=2.57$, $p<0.05$)).

Under all text type conditions the subjects made less errors with the Virtual Keyboard than with Graffiti 2 (normal text: $t(11)=9.634$, numbers: $t(11)=13.071$, commands: $t(11)=9.198$, $p<0.01$). Using the add-on WordComplete (for the input of normal text) did not significantly lower the error rate of the users.

Accuracy of Graffiti 2

Entering the alphabet five times with Graffiti 2 resulted in an accuracy of 86.03% (SD=10.00). If we weight the single accuracy measures for every character by appearance in the English language the accuracy will be 88.17%. The hardest to recognize characters are Q (48.3%), X (71.7%), Y (76.7%), K (78.3%), G (80%), T (80%) and I (85%). Least problems appeared when entering V (91.7%), A (93.3%), P (93.3%), Z (93.3%), C (95%), B (96.7%) and H (96.7%).

Entering the Graffiti 2 alphabet the subjects achieved 9.00 wpm and a total error rate of 11.50%. When entering the alphabet no errors were corrected, so the results are not comparable to the metrics calculated with the phrases.

Task Load (NASA-TLX)

The average task load for Graffiti 2 was 292.42, for Graffiti 2 + WordComplete 187.33 and the Virtual Keyboard had a TLX of 271.75. There were no significant differences between the task load of Graffiti 2 and the Virtual Keyboard. The subjects rated WordComplete less exhausting than Graffiti 2 ($t(11)=2.650$, $p<0.05$) and the Virtual Keyboard ($t(11)=2.350$, $p<0.05$).

Usability (SUS)

Graffiti 2 achieved least points with average 63.75, followed by the Virtual Keyboard with 68.96 and Graffiti 2 + WordComplete with 75.42. The results did not show any significant differences; the difference between Virtual Keyboard and Graffiti 2 is only nearly significant ($t(11)=2.092$, $p=0.06$).

DISCUSSION

Comparing the results with Fleetwood et al. (2002) [2], we measured lower text entry rates with the Virtual Keyboard. This could be due to the desktop apparatus we used: Unlike a real PDA the Virtual Keyboard and stylus were not co-located. We also used non-native English speakers for our experiment which may have influenced our results.

Graffiti 2 vs. Virtual Keyboard

Concerning the quantitative results entering normal text, commands and numbers the error rate is higher and entry speed lower using Graffiti 2 than using the Virtual Keyboard. Nevertheless the users did not rate the Virtual Keyboard more usable than Graffiti 2 (SUS). According to their statements, all subjects (except one) would rather use Graffiti 2, because it was more intuitive, closer to handwriting, and the subjects expected to gain speed and accuracy by training Graffiti 2. However they preferred to enter numbers and special characters with the Virtual Keyboard, even after having gained advanced Graffiti 2 skills.

Entering different text types

According to our hypotheses results for different text types differed depending on text entry method. The subjects reported that using Graffiti 2 numbers were easiest to enter, because of the limited number of characters. Commands were most difficult to enter, because the special characters they contained were hard to remember and achieved the highest error rate. Using the Virtual Keyboard numbers were also easiest to enter. This is because of the small distance between the individual virtual number keys and their bigger size. Commands were ranked last; the main difficulty was to switch between different keyboard layouts in order to enter special characters.

Graffiti 2 usability deficits

The biggest problems occurred when entering the characters Q, X, Y, K, G, I, T and @. These results are found looking at the input stream and questioning the subjects. All subjects had problems entering characters with two strokes (I, K, T and X). We assume that more training is needed to enter these characters with a lower error rate. The subjects often forgot the dot in the character I and the stroke up finishing Q (see Figure 1). This resulted in higher error rates. The subjects complained that these letters were not very intuitive to write. They described entering the letter X with two strokes as a good solution, although it should not matter which stroke to begin with. Problems also occurred when entering U-V and H-N, because the system did not recognize these characters correctly, e.g. understanding U when a V was intended and vice versa. U and V may be easier to enter in the old Graffiti, because the direction of the stroke does not matter. H-N errors occurred due to the fact that the subjects entered a lower case N which was not recognized correctly. The subjects described the letters K and T as not very intuitive, and two of them preferred the Graffiti 1 characters. Comparing these results with MacKenzie and Zhang (1997) [5] the problems with Q, T and I did not occur in Graffiti 1.

WordComplete add-on for Graffiti 2

The add-on WordComplete for Graffiti 2 had no impact on speed and error rate; nevertheless the subjects reported that WordComplete had less task load than Graffiti 2 without extension. They described WordComplete as a useful add-on, but they would appreciate the possibility to adjust the configuration (e.g. a longer list of word recommendations, personalized dictionary).

CONCLUSION

The experiment showed that immediate input with Graffiti 2 is indeed slower and more error-prone than input with Virtual Keyboard, no matter if entering normal text, numbers or commands. We found no difference in task load and usability rating. However, almost all subjects stated that they generally preferred Graffiti 2 as an input method, because it is more intuitive and more similar to handwriting. For entering numbers and special characters they would still prefer the Virtual Keyboard. Using WordComplete as an extension for Graffiti 2 resulted in less task load and was described as a useful add-on, but showed no other significant difference to Graffiti 2.

ACKNOWLEDGMENTS

We thank Georgine Beranek and Poika Isokoski for comments and help with this paper.

REFERENCES

[1] Brooke, J. (1996). SUS: A Quick and Dirty Usability Scale. In Jordan, P.W., Thomas, B., Weerdmeester B.A. & McClelland, I.L. (eds). *Usability Evaluation in Industry*, pp. 189-194. London: Taylor & Francis.

[2] Fleetwood, M.D., Byrne, M.D., Centgraf, P., Dudziak, K., Lin, B. & Mogilev, D. (2002). An analysis of text-entry in Palm OS: Graffiti and the Virtual Keyboard. *Proc. HFES 46th Annual Meeting,* pp. 617-621, Santa Monica: HFES.

[3] Goldberg, D. & Richardson C. (1993). Touch-typing with a stylus. *Proc. CHI93*, pp. 80-87, New York: ACM.

[4] Hart, S. & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human mental workload*, Hancock, P. & Meshkati, N. (Ed.), pp. 139-183. North-Holland: Elsevier.

[5] MacKenzie, I.S. & Soukoreff, R.W. (2003). Phrase sets for evaluating text entry techniques. *Ext. Abstracts CHI03*, pp. 754-755. New York: ACM.

[6] MacKenzie, I.S. & Zhang, S. (1997). The immediate usability of Graffiti. *Proc. Graphics Interface '97*, pp. 129-137. Toronto: CIPS.

[7] Palm, Inc. (2003). Graffiti 2 Alphabet, available at <http://www.palmone.com/us/support/graffiti2/>

[8] Soukoreff, R.W. & MacKenzie, I.S. (2003). Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. *Proc. CHI03*, pp. 113-120. New York: ACM.