User Experience with Chinese Handwriting Input on Touch-Screen Mobile Phones

Qin Gao[1], Bin Zhu[1], Pei-Luen Patrick Rau[1], Shilpa Vyas[2], Cuiling Chen[1], and Hui Li[1]

[1] Department of Industrial Engineering, Tsinghua University, Beijing, 100084, China
{trishagao,binzhuw,plprau,cathychencl,hli.sunshine}@gmail.com
[2] Research in Motion UK Limited, Centrum House, 36 Station Road, Egham, Surrey, England
Shvyas@rim.com

Abstract. The fast development of mobile services in China and the recent trend of touch-screen handheld devices precipitate researchers to understand user experience with mobile Chinese handwriting input (HWI). This research attempts to provide an integrated picture of user experience with Chinese HWI on touch-screen mobile phones. Five usability experts were invited to inspect usability problems of seven handwriting-enabled mobile phones, which varied in type of screens, writing tools and interaction approaches, operation systems, and handwriting recognizers. In a following usability evaluation with five novice users, we collected both quantitative data and qualitative data to give an overall performance assessment to the seven mobile phones and furthermore to analyze the usability issues with HWI. As a result, we identified 16 usability issues related with handwriting recognizers, input interfaces, and devices/users respectively.

Keywords: Usability evaluation, heuristic evaluation, mobile phone, user interface, Chinese, handwriting input.

1 Introduction

The popularity of mobile data services in China (e.g. web surfing, IM, emailing, etc.) highlights the importance of mobile Chinese input methods. Recently the trend of touch-screen handheld devices in China, such as iPhone, triggered a new round of enthusiasm on handwriting input (HWI), which is considered as a superior solution for languages using a large and complicated character set such as Chinese (Mackenzie & Soukorff, 2002). Theoretically anyone who can write Chinese is able to input with HWI, including people with little computer experience, such as older people. Different from the popular Pinyin-based method on keypad-equipped phones, HWI is independent of the dialect spoken by the user. Furthermore, the “naturalness” brought by the pen-and-paper metaphor upon which HWI is based (Frankish, Morgan, &Noes, 1994) is very appealing, especially for novice users.

As Chinese HWI becomes more and more popular, people are raising their expectation about its ease of use. Currently there has been a few studies on the usability of mobile HWI of alphanumeric text (Bouteruche, Deconde, Anquetil, &Jamet, 2005;

P.L.P. Rau (Ed.): CCD/HCII 2013, Part I, LNCS 8023, pp. 384–392, 2013.

Frankish, Morgan, &Noes, 1994; Frankish, Hull, &Morgan, 1995; MacKenzie & Zhang, 1997; Ren & Mizobuchi, 2005). None of these studies, however, have tried to provide a comprehensive understanding of issues influencing user experience throughout the entire inputting session, which involves interaction with the handheld device, the back-end handwriting recognizer, and the front-end input interface as a whole. Tasks in these studies were often constrained to inputting alphabet-numerical characters. In real use, users often need to carry out text input together with interface management operations (e.g., pointing, selecting), and often input characters from different vocabulary sets (e.g., insert an email address and a phone number in a message written in Chinese). Furthermore, the popularity of capacitive screen put forwards problems related to writing with fingers.

The current study aims to specify usability issues that influence user experience throughout the whole inputting session. In order to accomplish this goal, we started with a thorough review of existing usability studies on mobile HWI with particular focus on Chinese input. Then we invited five experts to identify usability problems on seven touch-screen phones that differ in brands, type of screens, writing tools, operation systems and handwriting recognizers. Following that, we compared user performance among the seven phones by usability testing with five novice users.

2 Related Work

There are mainly two alternative solutions for inputting text with touch-screen mobile phone: HWI and virtual keyboard (VKB). An early study by Lewis (1999) compared input rates for VKB and a simulated HWI. With the assumption of 100% recognition accuracy, HWI input speed averaged 23.6 words per minute (wpm) for sentences, which outperformed VKB (17.0 wpm). Liu, Ding, and Liu (2009) studied mobile Chinese input rate against novice users, and found that there was no significant difference in input rate between HWI and standard VKB, but users made more errors with HWI than with VKB. The result that even novice Chinese users can achieve a fairly good input rate may be explained by the fact that Chinese uses pictographic characters, and the complex forms and structures of Chinese character may provide more cues to handwriting recognizers (Dai, Liu, &Xiao, 2007), and consequently less selection operations. However, except for Lewis (1999) study that assumed a perfect HWI recognition accuracy, the error rate of HWI is found higher than that of VKB for both alphanumeric and Chinese input.

Error rate with HWI is closely related to the recognition accuracy of the recognizer. LaLomia (1994) found a HWI recognition accuracy of 97% or higher is generally acceptable. We did not find any study reporting such statistics with Chinese HWI, but results from a Japanese HWI study (Ren & Zhou, 2009) can be taken as a reference: the recognition accuracy for Chinese characters mixed with Kana characters was around 90% to 93%. Recognition accuracy is not only related with algorithm design of the recognizer, but also with users' writing style (Lumsden & Gammell, 2004) and the type and amount of training they received (MacKenzie & Zhang, 1997). Frankish (Frankish, Morgan, &Noyes, 1994) suggested recognition accuracy can be improved by putting limitations on the writing place (e.g., discrete cells for letters), the writing style (e.g., prohibiting cursive writing), the size of vocabulary (e.g., alphabetic letter only). But each of these measures may detract from the appeal of "naturally writing".

The importance of input interface quality has been recognized by a number of researches. Bouteruche et al. (2005) proposed a set of design guidelines based on the idea of maximizing spatial contiguity and minimizing attention switching from the writing task. These guidelines were implemented in their interface design of a HWI editor called DIGIME. For HWI using writing boxes, proper geometry and dimensions of the box are required to maximize the inputting speed and minimize the error rate. Ren and Zhou's (2009) study showed that the optimal writing area for inputting Chinese mixed with Kana characters is square box with a dimension of 14*14mm.

The physical aspect of handheld devices and writing tools may also influence usability of mobile HWI, as reflected by user comments in Lewis study (1999). However, we did not find many scholarly researches in this vein, except for Ren and Mizobuchi's study (Ren & Mizobuchi, 2005) on the optimal dimensions of stylus for Japanese people. Through two experiments they found the most suitable dimensions was: pen-length 11 cm, pen-tip width 0.5 cm, and pen width 7mm. To be noted is that participants of the experiment were Japanese, and the results should be generalized with cautious due to anthropometric difference among countries.

Previous studies imply that when a user input with mobile HWI, his/her experience is influenced not only by the handwriting recognizer, but also the input interface and the physical device. Though the impact of some specific features on usability has been studied, a purposeful examination is needed to obtain a big picture of usability issues that influence user experience with mobile Chinese HWI.

3 Methodology

To establish a comprehensive understanding of possible usability issues of mobile Chinese HWI, a wide examination of the state-of-art mobile HWI solutions on touch-screen phones was needed. We selected seven mobile phones covering a wide range of brands (Apple, Nokia, Samsung, LG, HTC, Dopod, and a local brand) and operation systems (iOS, Android, Windows Mobile, Symbian, and model built-in systems), as shown in Table 1. We ran two studies to specify usability issues that influence user experience with mobile Chinese HWI. In the first study, five usability experts inspected usability problems of the seven phones through self-reporting questionnaire, think-aloud tests, and a focus group discussion. The results of study 1 also informed the design of the following usability testing with five novice users (study 2).

Table 1. The seven mobile phones in the expert review and user testing

Mobile Phone	Operating System	Touch Screen	Writing Tool	Handwriting Software	HWI support
iPhone 3GS	iPhone OS 3.0	Capacitive	Finger	OS Built-in	Chinese only
HTC Magic	Android 1.5	Capacitive	Finger	OS Built-in	All (support mixed input)
Dopod T5399	Win Mobile 6.5	Resistive	Stylus	Malanhua	All (no mixed input)
Nokia N97 mini	Symbian S60	Resistive	Finger	OS Built-in	All (support mixed input)
LG KP500	Unspecified	Resistive	Stylus	OS Built-in	All (no mixed input)
Samsung S800	Touch Wiz 2.0	Resistive	Stylus	OS Built-in	All (no mixed input)
K-Touch N77	Unspecified	Resistive	Stylus	OS Built-in	All (support mixed input)

3.1 Study 1: Expert Review

Participants. Five usability experts with abundant touch-screen product experience were invited. Two were faculty majoring in usability engineering and human-computer interaction (HCI), one was a usability researcher who had led and worked in mutiple usability projects, and two were PhD students who had been trained for more than 4 years in the HCI program and also had adequate experience with usability evaluations.

Procedure. Each expert first tried out all the seven models separately in their working place and at their own pace. With each of the seven phones they were asked to complete the following tasks: creating and editing a contact, creating and sending short messages, sending and receiving emails, searching and browsing web contents, using mobile social networking services. For each task, the expert was asked to identify usability issues related to handwriting recognizer, input user interface issues, and general issues. Data were collected with a review form which experts filled in and returned to us at the end of their testing period of a phone. The seven phones were circulated between experts with coordination and a time slot of two hours for each phone with each expert was ensured.

Then each expert was invited to our usability laboratory for a think-aloud test. Two types of tasks were used in the think-aloud test: contact editing and messaging. Based on feedback from experts, the tasks were slightly revised and used in study 2, which were introduced in detail later. When performing tasks, experts were encouraged to speak aloud what they were looking at, thinking, and doing, and these were captured by video and audio recording. The testing sequence of the seven phones was randomized for each expert. Each session took about 3 to 4 hours.

Finally, we held a focus-group discussion with all five experts for discussing issues they found and generating further insights on usability of mobile Chinese HWI.

3.2 Study 2: User Testing

Participants. Five novice user testing sessions were held in our laboratory. All participants, including 3 females and 2 males, were graduate students from different department of Tsinghua University, aging from 23 to 27. They were all experienced mobile phone users, but none of them had ever own a touch-screen phone. It is interesting to find when being asked whether they want HWI feature, only one participant give positive answer. Concerns participants raised included: (1) keyboard (especially physical keyboard) input method is faster compared with HWI; (2) handwriting recognition has a relatively low accuracy rate; (3) a matter of habit and more accustomed to keyboard input mode; (4) the operation of handwriting sometimes needs both hands, thus not very convenient.

Tasks. The participants were required to complete five tasks with different length and content on each of the seven phones. Task design was guided by the goal to be representative of those users would encounter in real use. Task one and two involved creating new contacts by entering name, phone numbers, and email address in the contact manager of each phone. Task one involved a mixture of Chinese characters, English letters, while in task two, no Chinese characters were input. Task three to five involved sending messages to others. Task three was a greeting sentence in English; task four mixed a short English sentence, a Chinese sentence encouraging the recipient, and a phone number; task five was a Chinese ragged verse ended with an emoticon.

Measurement. 1) *Input rate* (characters per minute, cpm) was measured by the number of characters entered per minute. 2) *Error rate* (%) was measured by the ratio between the number of wrongly input characters and the number of total input characters (both correct and wrong). 3) *Correctness rate of the first character* (CRFC, %) was the ratio between the number of characters recognized by the first (default) alternative in the alternative list and the number of total input characters. It means the user does not need to extra selection action for the correct recognition. Compared with error rate, CRFC reflects more precisely the recognition accuracy of recognizer. 4) *Subjective satisfaction* was measured by the means of users' ratings on five satisfaction scales, asking about users' satisfaction with recognition accuracy, recognition speed, input interface, design of the contact/messaging function of the phone, and overall satisfaction with the HWI being used. Five-point Likert scales (1 for highly unsatisfied and 5 for highly satisfied) were employed in the questionnaire.

Procedure. Each user was tested individually in the usability laboratory of Tsinghua University. After being introduced the research, participants filled out a background questionnaire that collected demographic information and past experience with touch-screen phones. Before carrying out tasks, participants were allowed to experience each phone for a couple of minutes. Then, the participants completed the five tasks on each of the seven phones with HWI, except for iPhone, whose HWI recognized only Chinese characters and some simple punctuation but not others (this was no longer true for built-in HWI with latest iOS 4.3). The testing sequence of the seven mobile phones was randomized. Upon completion of tasks with a mobile phone, the participants then completed a satisfaction questionnaire for that phone. They were also shortly interviewed about usability problems they have found. Each session took about 2~3 hours.

4 Results Analysis

Quantitative analysis was first conducted to compare the performance of seven mobile phones. User performance and satisfaction from study 2 was reported in 4.1, and usability issues identified in both study 1 and 2 were synthesized and reported in 4.2.

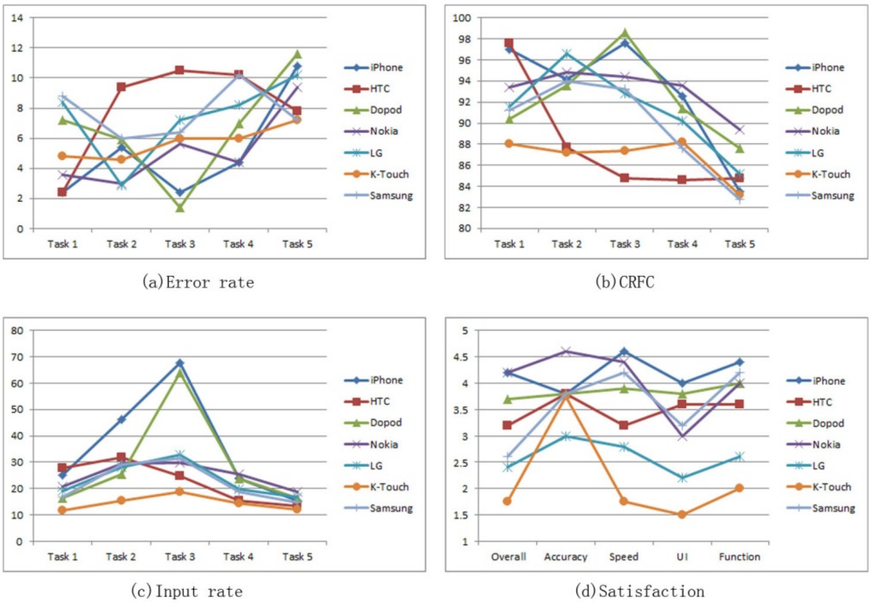
4.1 User Performance and Satisfaction

Error Rate and CRFC. The means of error rate & CRFC for each mobile phone and task are demonstrated in Fig.1 (a) & (b). For Chinese input tasks, the performance of Nokia was the best (low error rate and high CRFC), while the performance of Dopod, LG, HTC, & Samsung was relatively worse. There were several factors influencing input accuracy of Chinese HWI. A major one is how the input method deals with mixed input of Chinese, English, and numbers. For example, Nokia supports mixed-input and also gives priority to the recognition of characters corresponding with the current mode. This setting can greatly decrease the errors, because users don't need to frequently switch recognition mode (e.g., Chinese mode, English mode), which is easily omitted by users. Dopod, Samsung and LG do not support mixed input.

Another key factor is the interference between screen operations and handwriting. The interference problem with Dopod was complained 8 times by participants during the studies. In addition, too few recognition alternatives also influence the accuracy a lot. For example, iPhone only provided 4 recognition alternatives, and it is more likely that no candidate is correct and users need to delete and re-write the character. Besides, the relatively low accuracy of LG was caused by its problematic writing area design. Its writing box looks like a 田, and participants easily regarded it as four independent sub-area, but in fact there is only two independent sub-area.

Input Rate. The average input rate for each mobile phone and task is demonstrated in Fig.1 (c). In terms of Chinese input tasks, Nokia got the highest input rate among the seven mobile phones, which was due to the mixed-input recognition feature of Nokia. The lowest input rate of K-Touch was mainly caused by the following reasons: 1) interference between screen operations and handwriting; 2) the slow reaction speed of back-end recognizer; 3) the difficulty with switching recognition mode. For English tasks, Dopod and iPhone were much faster than the other phones. In particular, Dopod supports continuous handwriting input recognition for English characters and numbers, while on iPhone participants used virtual keyboard for English letters and numbers.

Satisfaction. Satisfaction ratings with the seven mobile phones are shown in Fig.1 (d). Participants were most satisfied with iPhone and Nokia among the seven mobile phones. In particular they were most satisfied with the recognition accuracy of Nokia and the recognition speed/input interface of iPhone. From Fig.1 (d), we also found that in most cases the ratings of user interface were lower than the other items, which indicate that more attention should be paid to the interface design in the future.



(a) Error rate (b) CRFC

(c) Input rate (d) Satisfaction

Fig. 1. The results of quantitative analysis

4.2 Usability Issues of Mobile Chinese HWI

Based upon the study 1 & 2, we compiled a list of usability issues, which were classified into three categories: input interface related issues, handwriting recognizer related issues, and device and users related issues.

- **Input Interface Related Issues**

1. *Visual style of stokes should resemble natural chirography in pen-and-paper writing.* The visual style of K-Touch was most appreciated in the experiment as it resembles the look of strokes in pen-and-paper writing. Participants complained that too thick (LG) or too thin (Dopod & Nokia) strokes were lack of a feeling of "natural writing".
2. *For writing area design, full-screen writing over the current window is preferred to full-screen writing in a new window; writing in a box tiled with the current window is preferred to writing in a box overlapping the current window.* For full-screen writing in a new window, users need to write in a totally new window after they activate the edit function, and return to the previous window after complete the writing. This way of jumping out and backward break the normal workflow of users. For overlapping writing box (e.g. the box mode of Nokia), when the box field overlaps the input field, users need to close the writing box and move the field and re-open the writing box again.
3. *Provide 5~7 recognition candidates.* Too few recognition candidates may reduce the possibility of "hitting" the correct target and also arouse anxiety from users. A number of 5~7 is recommended, though more scientific study is needed to prove its validity.
4. *Prediction characters should not be presented simultaneously with recognition candidates.* Predicting the next character based on available input is helpful for improving input rate and reducing user fatigue. But presenting predictive characters simultaneously with recognition candidates is very error-prone: users may click a predictive character but not a recognition candidate. It is better to present predictive characters after the user select a recognition candidate.
5. *Provide direct shortcuts to frequently used punctuation and other symbols on the input screen.* Displaying a list of frequently used punctuations and symbols on the interface is appreciated by users.
6. *Make correction operation easy to recognize and access.* The backspace key and the delete key are important correction tools, and should be designed visually prominent (e.g., bigger size, placing at the right corner) and physically easy to access.
7. *The icons on the handwriting input interface should be legible and easily recognized by user.* The icons should be simplified and could represent meanings clearly. The elements of icons (e.g. usage, appearance, borders) should be kept consistent.

- **Handwriting Recognizer Related Issues**

1. *User effort for switching between different modes should be minimized.* Repetitive and difficult switching among different modes was complained most. Possible approaches include: a) allowing mixed-input of frequently used symbols, numbers,

and English characters; b) making the switch operation easy and effective; c) the recognizer can adjust its mode to the input context automatically. For example, for form-filling tasks, such as editing contact information, the recognizer can switch its input mode according to current input field, such as digits or English letters.

2. *Users are more accustomed to timeout method instead of discrete writing box method for character segmentation.* The benefit of discrete writing box method is it allows users write in their own rhythm and prevent errors due to incomplete writing. However, we found neither experts nor novice users appreciate this benefit. Some novice users did not understand how this method works and still wrote as with a timeout method: they would not start writing a new character in the next box until the previous input character was correctly recognized.
3. *The timeout threshold should be defined according to the trade-off between speed and accuracy.* We found short timeout threshold leads to recognition errors due to incomplete writing, especially for complex characters with many strokes. According to Cui and Lantz (2005), optimal timeout options for fast, ordinary, and slow writers could be 350ms, 500ms, and 700ms respectively.
4. *It is best to provide continuous input recognition for English and number HWI.* The continuous input recognition can increase the input rate and accuracy greatly.

- **Device/User Related Issues**

1. *Ensure users have proper writing tool.* Writing with fingers or stylus is often decided by the type of screens. For **stylus input**, the size of stylus would influence the posture of pen-holding and writing. For **finger input**, it should not be assumed that all users have long-enough nails or similar size of fingers to produce efficient handwriting.
2. *The required force for writing on screen surface should allow users write for a long time without serious fatigue.* Participants found writing on K-Touch and LG KP500 very laborious since both requires high pressure on the screen. This usability issue relates to the design or selection of touch-screens. Further study should be conducted to find out how big the force should be.
3. *Tracking speed should be fast enough to allow continuous writing experience.* We found slow tracking speed not only lead to longer writing time, but also incurs errors. The recognizer sometimes misinterpreted this pause as a signal of accomplishment and started recognition, which will cause errors. Miller (1968) suggested 0.1 second as the limit of response time for a system response to be perceived as immediate and part of the mechanic action induced by the user.

4.3 Conclusion and Limitations

In this paper, we evaluated usability of mobile Chinese HWI applications of seven mobile phones. Based upon the results from expert inspection and usability testing, we finally identified 16 usability issues related to handwriting recognizers, input interfaces, and devices/users. We expect this list of usability issues give mobile HWI designers a more integrated view of usability requirements from users and provide a handy reference to inspecting usability problems of a certain solution.

The current study is in its nature exploratory, and is inadequate to give definite answers to the identified issues. More strictly controlled experiment research in the future are expected to provide elaborated and validated guidelines based on the framework of the current study. In particular, we found mobile HWI design guidelines for older people in great need due to the sheer volume of the Chinese grey population and potential digital gap resulted from inaccessible informational product design.

Acknowledgements. This study was supported by Research In Motion UK Limited.

References

1. Bouteruche, F., Deconde, G., Anquetil, E., Jamet, E.: Design and Evaluation of Handwriting Input Interfaces for Small-size Mobile Devices. In: 1st Workshop on Improving and Assessing Pen-Based Input Techniques, pp. 49–56 (2005)
2. Cui, Y., Lantz, V.: Stroke Break Analysis: a Practical Method to Study Timeout Value for Handwriting Recognition Input. In: 7th International Conference on Human Computer Interaction with Mobile Devices & Services, pp. 263–266. ACM (2005)
3. Dai, R., Liu, C., Xiao, B.: Chinese Character Recognition: History, Status and Prospects. Frontiers of Computer Science in China 1(2), 126–136 (2007)
4. Frankish, C., Morgan, P., Noyes, J.: Pen Computing: Some Human Factors Issues. In: IEE Colloquium on Handwriting and Pen-Based Input, IET, pp. 5/1–5/3 (1994)
5. Frankish, C., Hull, R., Morgan, P.: Recognition Accuracy and User Acceptance of Pen Interfaces. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 503–510. ACM Press/Addison-Wesley Publishing Co. (1995)
6. LaLomia, M.: User Acceptance of Handwritten Recognition Accuracy. In: Conference Companion on Human Factors in Computing Systems, pp. 107–108. ACM (1994)
7. Lewis, J.R.: Input Rates and User Preference for Three Small-Screen Input Methods: Standard Keyboard, Predictive Keyboard, and Handwriting. In: 43th Annual Meeting of Human Factors and Ergonomics Society, pp. 425–428 (1999)
8. Liu, Y., Ding, K., Liu, N.: Immediate User Performances with Touch Chinese Text entry Solutions on Handheld Devices. In: 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 56:1–56:2. ACM (2009)
9. Lumsden, J., Gammell, A.: Mobile Note Taking: Investigating the Efficacy of Mobile Text Entry. In: Brewster, S., Dunlop, M. (eds.) Mobile HCI 2004. LNCS, vol. 3160, pp. 156–167. Springer, Heidelberg (2004)
10. MacKenzie, I.S., Zhang, S.X.: The Immediate Usability of Graffiti. In: The Conference on Graphics Interface 1997, pp. 129–137. Canadian Information Processing Society (1997)
11. Mackenzie, S., Soukoreff, W.: Text Entry for Mobile Computing: Models and Methods,Theory and Practice. Human-Computer Interaction 17(2&3), 147–198 (2002)
12. Miller, R.B.: Response Time in Man-computer Conversational Transactions. In: Fall Joint Computer Conference, Part I, pp. 267–277. ACM (1968)
13. Ren, X., Mizobuchi, S.: Investigating the Usability of the Stylus Pen on Handheld Devices. In: 4th Annual Workshop on HCI Research in MIS, pp. 30–34 (2005)
14. Ren, X., Zhou, X.: The Optimal Size of Handwriting Character Input Boxes on PDAs. International Journal of Human-Computer Interaction 25(8), 762–784 (2009)
15. Soukoreff, R.W., MacKenzie, I.S.: Recent Developments in Text-entry Error Rate Measurement. In: Extended Abstract of the ACM Conference on Human Factors in Computing Systems (CHI), pp. 1425–1428 (2004)