Ying Liu

Chinese Text Entry with Mobile Devices

ACADEMIC DISSERTATION
To be presented with the permission of the Faculty of Information Sciences of the University of Tampere, for public discussion in Pinni auditorium B1097 on December 3rd, 2010, at noon.

Department of Computer Sciences
University of Tampere

Dissertations in Interactive Technology, Number 12
Tampere 2010

ACADEMIC DISSERTATION IN INTERACTIVE TECHNOLOGY

Supervisor:	Professor Kari-Jouko Räihä, Ph.D., Department of Computer Sciences, University of Tampere, Finland
Opponent:	Professor I. Scott MacKenzie, Ph.D., Department of Computer Science and Engineering, York University, Canada
Reviewers:	Janet C. Read, Ph.D., Reader, School of Computing, Engineering & Physical Sciences, University of Central Lancashire, United Kingdom Shumin Zhai, Ph.D., IBM Almaden Research Center, USA

Dissertations in Interactive Technology, Number 12

Department of Computer Sciences
FIN-33014 University of Tampere
FINLAND

ISBN 978-951-44-8271-7
ISSN 1795-9489

Tampereen yliopistopaino Oy
Tampere 2010

Abstract

Text entry methods enable entry of written languages to computing systems. Chinese has unique characteristics as a logosyllabic language, which bring new challenges to the design and evaluation of Chinese text entry methods in computing systems. Chinese text entry on mobile devices is becoming an important and urgent research topic with the widespread and frequent usage of mobile devices in this large user group.

The overall objective of this thesis is to improve user experience of Chinese text entry on mobile devices, and the focus is on the process of interaction between users and text entry methods. I designed new interaction solutions with the user-centered design (UCD) approach and explored the patterns of user behavior with various approaches, including comparative user studies and performance modeling. This thesis covers four means of Chinese text entry on mobile devices: Chinese handwriting recognition, Chinese indirect text entry with a rotator, Mandarin dictation, and Chinese pinyin input methods with a 12-key keypad.

In this thesis, I present new design solutions for Chinese handwriting recognition and text entry methods utilizing a rotator. Follow-up evaluations showed the new designs to be well accepted by users. The design processes are presented as well, to show how users can be involved in such UCD approaches. Then I present a Mandarin short message dictation application for mobile phones, with two associated studies. The first study involves two-round evaluation of the isolated Mandarin dictation application, and the second is a Wizard of Oz study aimed at understanding effects of speaking styles in combination with two other factors. Two studies were also carried out on Chinese pinyin input methods that are based on the 12-key keypad. The comparative study of five phrasal pinyin input methods reflected design guidelines for the advanced feature of phrasal input. The other study description presents a predictive model addressing users' error-free speeds, which proved to be effective in three evaluation studies.

This thesis presents rich background knowledge and new findings on the process of interaction with various Chinese text entry methods on mobile devices. It is hoped that this work will assist researchers and practitioners to understand more about Chinese text entry on mobile devices and serve as a starting point for further work on the subject.

Acknowledgements

It is unbelievable that the thesis is ready. So many people have helped me during the years for preparing this thesis and I would like to express my sincere gratitude to them.

First I want to thank my supervisor, Prof. Kari-Jouko Räihä. Without his being open-minded and supportive, I could not have become a Ph.D. student in the University of Tampere. Without his sharing of knowledge, constant guidance in my study and research works and spiritual support, I could never have completed this thesis.

It is worth mentioning that the works presented in this thesis have mainly been completed while I have worked in Nokia Research Center (NRC) in Beijing from 2001 to now. I want to acknowledge my colleagues who have worked with me and supported my works for this thesis in Nokia. Without their support and companionship I could not have completed this thesis. Thanks to Panu Korhonen and Jari Kangas, who gave me the chance to work for NRC. Thanks to Fumiko Ichikawa, Ning Liu, Qiqun Wang, Yuan Feng, Xinxing Yang, Kai Ding and Guohong Ding for their support on works presented in this thesis. Thanks to all my managers in Nokia who are so generous to support my thesis work.

It was also my great honor to have Doctor Janet C. Read from University of Central Lancashire, UK and Doctor Shumin Zhai from IBM Almaden Research Center, US as the reviewers of my thesis. Their constructive and valuable comments helped me finalize this work.

Finally, I want to thank my family. They are the people who are always on my side and give me whatever support I need. I want to thank my husband Baozhu Yang for his patience. I also thank my handsome son Ce Yang for his understanding when he needed to play by himself while I was working on my thesis at home. I thank my parents and parents in law, and their supports are so necessary during this journey.

October 17, 2010

Ying Liu

Contents

List of Abbreviations

CCPM	Chinese characters per minute	p. 16
GOMS	Goals, operators, methods, and selection rules	p. 108
HCI	Human-computer interaction	p. 2
HWR	Handwriting recognition	p. 4
ISO	International Organization for Standardization	p. 3
KLM	Keystroke-level model	p. 105
KSPC	Keystrokes per character	p. 27
PC	Personal computer	p. 20
PDA	Personal digital assistant	p. 2
SMS	Short Message Service	p. 1
UCD	User-centered design	p. 2
UI	User interface	p. 5
WOZ	Wizard of Oz	p. 4
WPM	Words per minute	p. 16

1 Introduction

1.1 Objective

The overall objective of this thesis is to provide knowledge that can help to improve user experience of Chinese text entry with mobile devices. To reach this objective, I designed new interaction solutions and explored patterns of user behavior in the Chinese text entry process with various approaches. The studies covered in this thesis focus on Chinese text entry methods for simplified Chinese, mainly used in mainland China.

Text entry methods in computing systems enable users to enter written languages to those systems. Such a transcription from analog written languages to machine-readable digital text makes the storage, transmission, search, and analysis of text more efficient. Since each language has its unique characteristics and must be coded specifically in the computing systems, text entry methods are generally language-specific.

Chinese text entry on mobile devices is an important and urgent area of research (Wang, 2003). First, there is a remarkable amount of mobile phone users in China now, and the number is still growing rapidly. By August of 2009, the number of users of mobile phones in China had exceeded 710 million, much more than that of Internet users, 338 million (CNNIC, 2010). Second, the Short Message Service (SMS), which involves intensive text entry tasks, has seen phenomenal growth in China since its launch in 2000 (MIIT, 2009; Ma et al., 2007). In May 2009, Chinese people sent about 63 billion short messages, with more than three text messages per user per day (MIIT, 2009). Third, the increasing computing capabilities of mobile devices allow for the development of new applications, such as e-mail and instant messaging, in which text entry is the core task (Griffiths, 2006). Optimized text entry methods are compelling for such new applications of mobile devices.

At the same time, the design of Chinese text entry in mobile computing systems faces great challenges. First, Chinese is unique as a logosyllabic language. The current computing systems were invented in Western cultures where alphabetic languages dominate. How to apply existing knowledge and creatively design mobile text entry methods for Chinese is a great challenge. Second, the widespread use of mobile devices in China implies great variety among the users in many respects, such as technology literacy, needs, and aspirations. How to design mobile text entry for such a dynamic and varied user group is a crucial challenge. In addition, the published explorations of patterns of user behavior in Chinese text entry tasks remain limited, despite the tremendous number of Chinese-speakers worldwide (Sacher, 1998; Sacher, Tng, & Loudon, 2001; Marcus, 2003; Wang, 2003).

Motivated by the urgent need for exploration of the interaction process when users enter Chinese text with mobile devices, this thesis has as its aim to address the following main research questions:

- What are the design opportunities and drivers for new Chinese text entry methods in mobile devices? What are the most fruitful ways to design these new methods? How can we involve users in such design processes?

- How can Chinese text entry methods for mobile devices be evaluated? What are the key metrics for use in such evaluations? How do the text entry methods perform according to these measurements?

- What are the core processes involved in Chinese text entry tasks? How are such processes best modeled, and with what theories?

1.2 CONTEXT AND METHOD

The primary works referred to and described in the thesis are in the field of human–computer interaction (HCI). Human–computer interaction is "*a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them*" (Hewett et al., 1992). In this thesis, end users and mobile computing systems are studied alongside the core task of entering Chinese text, where "mobile computing systems" is used mainly to refer to mobile phones and personal digital assistants (PDAs).

The sub-areas of HCI that are relevant for the purposes of the thesis include studies of mobile text entry methods for alphabetic languages, the user-centered design (UCD) process, and user performance modeling.

When I planned and conducted the studies presented in this thesis, I referred to the approaches, practices, and measures of relevant studies of text entry methods for alphabetic languages. Although Chinese is different from alphabetic languages, the experiences shared from those studies still proved of great reference value. In Chapter 2 of this thesis, I review relevant design and evaluation principles from text entry studies for alphabetic languages (mainly English).

User-centered design is the main approach applied in the design work done for this thesis. This is a design philosophy and an approach wherein user needs, limitations, and characteristics are at the focus of the design of interactive systems (Nielsen, 1993; Jokela et al., 2003; ISO, 1999). Guidance on the process of UCD as set forth in the ISO 13407 standard defines its core activities; see Figure 1.1. The figure shows design solutions produced in the UCD process being evaluated against predefined usability or user experience criteria. Hence, UCD cannot be explained clearly without mention of the definitions of usability and user experience.

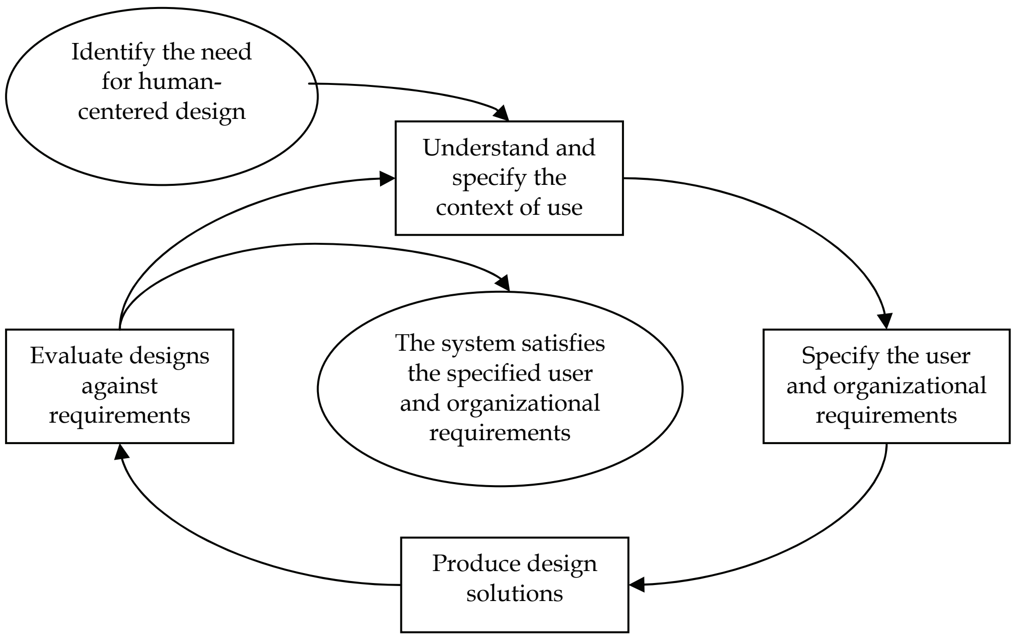


Figure 1.1. The UCD process according to ISO 13407.

A widely cited definition of usability is provided by the International Organization for Standardization (ISO) in ISO 9241-11, where usability is defined as "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*" (ISO, 1998). In recent years, the concept of user experience has become more commonly employed than that of usability, because researchers realized that there are other factors involved in HCI than human cognition and motor performance. A draft definition of user experience from ISO (2008) is "*A person's perceptions and responses that result from the use or anticipated use of a product, system or service.*" The definition was proven to be promising by a survey on the definition of user experience (Law et al., 2009).

User modeling, another sub-field of HCI, is also applied in this thesis, to understand the pattern of user behaviors in the interaction process with mobile Chinese text entry. A model is a simplified description or prediction of reality (MacKenzie, 2003; Carroll, 2003). User models are built to describe or predict user behaviors in human–computer interactions. User models are useful in that they can be utilized to evaluate user performance (e.g., speed and error rates) and identify design opportunities without the need for tedious user studies (Card, Moran, & Newell, 1980, 1983; Soukoreff & MacKenzie, 1995; Dunlop & Crossan, 2000; Silfverberg, MacKenzie, & Korhonen, 2000; Zhai, Smith, & Hunter, 2002; Zhai, Sue, & Accot, 2002; Isokoski, 2004; Myung, 2004; Pavlovych & Stuerzlinger, 2004; Cockburn, Gutwin, & Greenberg, 2007; Dunlop & Masters, 2008; Cockburn & Gutwin, 2009).

Another relevant discipline is linguistics. Linguistic studies of Chinese describe and explain the nature of the Chinese language, which is one of the bases for understanding user behaviors in text entry tasks. Chapter 2 of this thesis summarizes the latest findings from Chinese linguistic studies, as well as other characteristics of the Chinese language.

1.3 CONTRIBUTIONS

This thesis presents the following work:

– design of a Chinese handwriting recognition (HWR) method based on a touchpad for mobile phones, together with a series of user studies for different purposes;

– design of a rotator-based Chinese text entry method for mobile phones, along with user evaluation results and discussions;

– evaluations of an isolated Mandarin message dictation application on mobile phones and a Wizard of Oz (WOZ) study to understand effects of speaking style on user performance and satisfaction; and

– a benchmark user study of five Chinese pinyin[1] phrasal input methods for mobile phones, together with discussion of their usability and design guidelines and trends, and a validated

[1] Pinyin (also named as Han Yu Pin Yin) is the standard coding system of Chinese characters based on the Mandarin pronunciations in the form of Latin letters. Pinyin was approved as the standard Romanization system of Chinese characters by China and ISO in respectively 1958 and 1982. For more information about how pinyin text entry systems work in computing systems, please refer to Section 2.3.1 in this thesis.

> predictive model that estimates average text entry speeds for users' error-free Chinese pinyin input with the 12-key keypad on mobile phones.

The design and study practices presented in this thesis enrich knowledge of how UCD can be applied in the specific field considered. For both design cases, the customized UCD processes are explained in detail. They can serve as examples to show how design opportunities were identified and developed and how users were involved in the design processes. Moreover, results of the user studies can guide further design of Chinese text entry methods. Prototypes were built in most user studies, and this work may help readers to understand how they were built and when they are needed in the UCD process.

In this thesis, I also review works on the following topics: general characteristics of Chinese, characteristics of Chinese short messages, Chinese text entry methods for mobile devices, design and user study principles for mobile text entry, selection-based text entry methods and devices, and theories applied for modeling user performance in text entry tasks. These reviews should provide readers with the background necessary to understand the contributions of this thesis.

Some of the results presented in this thesis were published at conferences. However, here they are either reformulated or presented with additional work and content. I am the first author for most of the relevant publications. A summary of these results and the relevant publications is presented below:

In Chapter 3, I present the user-centered design process for a Chinese HWR method based on a touchpad on mobile phones. Three user studies are also presented, to explain how users were involved in the UCD process. The second user study was aimed at understanding the effects of user type and user interface (UI) designs on user performance. Some of the results from this study were published in the following poster paper:

> **Liu, Y.**, Ding, K., & Liu, N. (2009). Immediate user performances with touch Chinese text entry solutions on handheld devices. In *Proceedings of International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI '09)*, ACM Press, 56–57.

In Chapter 4, I present the user-centered design process for Chinese pinyin input methods with a rotator. Two new design solutions – and user evaluations to compare them with the initial design – are presented. Most text and the results presented in this chapter originate from the following publication:

Liu, Y., & Räihä, K.-J. (2008). RotaTxt: Chinese pinyin input with a rotator. In *Proceedings of International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI '08)*, ACM Press, 225–233.

Chapter 5 presents studies on, and for, a Mandarin short message dictation application. In the first section of the chapter, I present two-round user evaluations of an isolated Mandarin dictation application on Symbian Series 60 mobile phones. The second user study was mentioned in the following publication:

Alhonen, J., Cao, Y., Ding, G., **Liu, Y.**, Olsen, J., Wang, X., & Yang, X. (2007). Mandarin short message dictation on Symbian Series 60 mobile phones. In *Proceedings of Mobility 2007*, ACM Press, 431–438.

In the second part of Chapter 5, I present a Wizard of Oz study to explore the possible effects of speaking style on user performance and satisfaction. The content was reproduced for this thesis from a presentation of the study in the following paper:

Liu, Y., Jiang, L., & Yang, X. (2009). Will input style affect Mandarin short messages in mobile device?: a Wizard of Oz study. In *Proceedings of Speech in Mobile and Pervasive Environments Workshop* (*SiMPE Workshop 2009*) in conjunction with Mobile HCI 2009, 41–44.

In Chapter 6, I present two studies on Chinese pinyin input methods for the 12-key keypad. The first section of the chapter presents a benchmark evaluation of five phrasal pinyin input methods, which is based on material in the following publication:

Liu, Y., & Wang, Q. (2007). Chinese pinyin phrasal input on mobile phone: usability and developing trends. In *Proceedings of Mobility 2007*, ACM Press, 548–554.

In the second section of Chapter 6, a predictive model is presented to estimate users' error-free text entry speeds with pinyin character input methods. Five experiments are presented in this chapter, with the first two for eliciting parameters for the model and the last three for evaluations of the model. The model with the first two experiments for parameters and the third experiment for evaluation was published in the following paper:

Liu, Y., & Räihä, K.-J. (2010). Predicting Chinese text entry speeds on mobile phones. In *Proceedings of 28th International Conference on Human Factors on Computing Systems (CHI '10)*, ACM Press, 2183–2192.

2 Background

2.1 GENERAL CHARACTERISTICS OF CHINESE

Chinese is a logosyllabic language in nature. It is different from alphabetic languages like English. For example, the visual form of an English word often indicates its pronuciation. However, the visual form of a Chinese character does not give a clear hint of its pronunciation due to the extensive homophony in Chinese. Instead, the visual form of a Chinese character is often closely associated with its meanings.

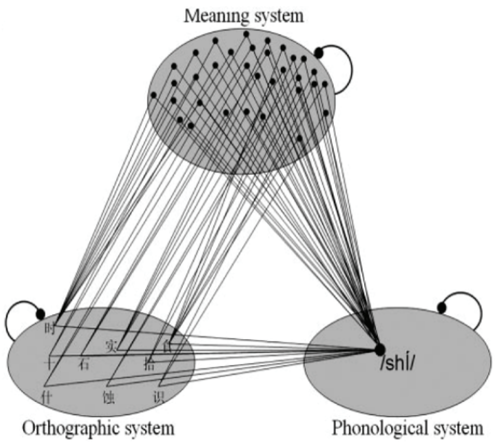


Figure 2.1. An example of the complex relationship among the orthographic, phonological, and meaning system of Chinese. One phonological unit corresponds to many orthographic units and even more meaning nodes due to the extensive homophony in Chinese. For example, the phonological unit of “shi” is shared by nine Chinese characters with the same tone and all meanings associated with those characters (Tan et al., 2005).

Figure 2.1 shows the relationships among the orthographic system, phonological system, and meaning system of Chinese characters. An orthographic system specifies a standardized way of using a specific script to write the characters. It includes both the visual form of the specific script and also how the script should be written. A phonological system refers to the sound systems to encode meaning in human spoken languages. In Chinese, one phonological unit corresponds to many orthographic units and even more meaning nodes. As Figure 2.1 shows, the phonological unit of “shi” is shared by nine Chinese characters with the same tone and all meanings associated with those characters.

There are numerous characters in Chinese: the latest *Kangxi Dictionary* includes 57,557 Chinese characters, though many of them are not frequently used (Zhang et al., 2008). It has been reported that about 2,000 to 4,000 Chinese characters are necessary for basic reading, according to a range of studies (Ann, 1982; Qiao, Qiao, & Qiao, 1990; Gao, Zhong, & Ceng, 1995; Sacher, 1998; Sacher, Tng, & Loudon, 2001). A Chinese character is generally comparable to an English word. For example, the Chinese character of “山” means mountain while Chinese words like “高山” (high mountain) and ”百家姓” (family names) correspond to English phrases. Table 2.1 lists the national standards on character sets that may be supported by computing systems.

Table 2.1: National standards for Chinese character sets for computing systems

<table>
<tr><th>Standard</th><th colspan="2">Number of characters</th><th>Description</th></tr>
<tr><td>GB2312
(1980)</td><td>Level 1: 3,755
Level 2: 3,008</td><td>Total: 6,763</td><td>Simplified Chinese</td></tr>
<tr><td>GB13000
(1993)</td><td colspan="2">20,902</td><td>Simplified and traditional Chinese</td></tr>
<tr><td>GB18030
(2005)</td><td colspan="2">27,533</td><td>Incl. characters used by minority nationalities</td></tr>
</table>

Chinese text can be presented both vertically (top to bottom first, then right to left) and horizontally (left to right first, top downward) (see Figure 2.2), with the former mainly used in ancient times and the latter used nowadays. There are also two written forms for Chinese characters, for historical reasons: the traditional form, used in Hong Kong and Taiwan, and the simplified form, used in mainland China. The simplified Chinese was published in 1956 in mainland China. The latest update for simplified Chinese in 1988 covers 2235 Chinese characters that were simplified based on their traditional written forms. The coding and text entry methods for simplified and traditional Chinese are also different in computing systems. Zhuyin (Bopomofo) and Cangjie are two typical examples of text entry methods for traditional Chinese. This thesis concentrates on text entry for simplified Chinese. It may be hard to apply the results of this thesis to traditional Chinese text entry methods directly. Further explorations are needed to understand how the results can be applied in traditional Chinese.

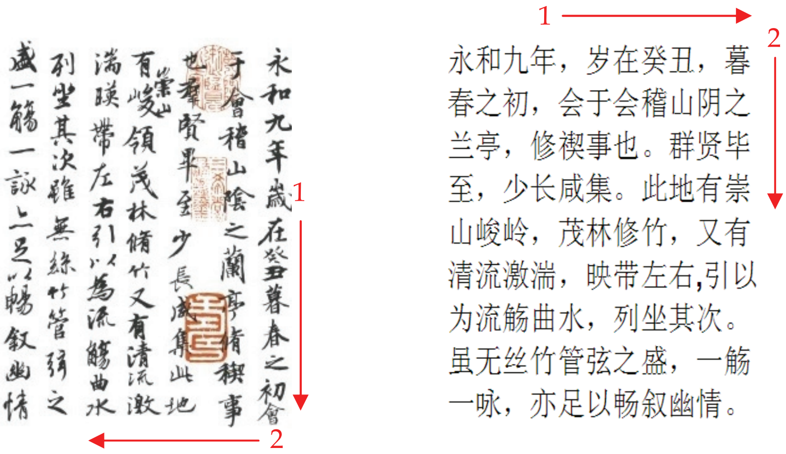


Figure 2.2. Content from Lantingji Xu presented in two ways: vertically in traditional Chinese without punctuation, in handwritten form by ancient calligrapher Wang Xizhi (at left), and horizontally in simplified Chinese, in print form (right). Source: http://www.hrdmv.com/thread-3229-1-1.html.

Chinese characters are complex in structure. There are generally two levels of components forming a Chinese character: radicals and strokes (see Figure 2.3). Strokes are the minimal writing units that can be drawn with a single pen action of lowering and raising the writing instrument (Qiao, Qiao, & Qiao, 1990; Lin & Sears, 2005, 2007). Radicals are the relatively independent components of a Chinese character and usually include multiple strokes (Contemporary Chinese Dictionary, 1996; Liu, Jaeger, & Nakagawa, 2004; Dai, Liu, & Xiao, 2007).

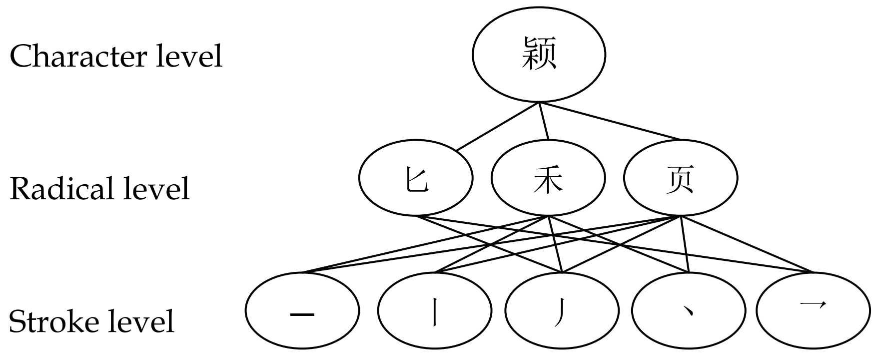


Figure 2.3. An example of a Chinese character with its components at radical and stroke levels (the basic strokes listed here, from left to right, are the horizontal line, vertical line or vertical line with a hook to the left, left curve, dot or right curve, and hook covering all strokes with bends).

There are 10 major and more than 40 minor groups of dialects of spoken Chinese (see Table 2.2), which differ greatly in their expressions, pronunciations, grammars, etc. (Wang, 1999; Li, 2005). In China, people sometimes cannot understand each other with only oral communications if they speak in native dialects that are different. In 1958, Mandarin, based on the local dialect of the Beijing area, became the standard pronunciation system for Chinese (Wang, 1999; Sacher, Tng, & Loudon, 2001), and more people have learned to speak it since then. However, people's native dialect still affects how they speak in Mandarin, which has further influence on their use of the pinyin system that is based on the Mandarin pronunciations of Chinese characters (Liu, Jaeger, & Nakagawa, 2004; Dai, Liu, & Xiao, 2007).

Table 2.2: The 10 main groups of dialects of Chinese and their number of speakers (Wang, 1999)

Dialect	Number of speakers (in millions)	Dialect	Number of speakers (in millions)
北方 Mandarin	662.23	客家 Hakka	35.00
吴 Wu	69.75	湘 Xiang	30.85
闽 Min	55.07	赣 Gan	31.27
晋 Jin	45.70	徽 Hui	3.12
粤 Yue (Cantonese)	40.21	平话 Pinghua	2.00

Chinese is also a tonal language. For example, the most popular flavor of Mandarin Chinese includes five tones: flat (–), rising (/), up-down-up (∨), falling (\), and neutral (·). Cantonese has nine tones. The number of syllables in Mandarin Chinese expands from about 400[2] to 1,300 when the tones are taken into account. Tones in Chinese are semantically meaningful (Li, 2005; Alhonen et al., 2007). First, they decrease the number of homonyms (words that share the same spelling and the same pronunciation but have different meanings). Second, they specify the exact character in spoken language, in combination with language contexts. The characteristics of homophones and the tonal nature of Chinese bring with them greater challenges for speech recognition technology and speech user interface design (Sacher, 1998; Li, 2005; Sacher, Tng, & Loudon, 2001; Chang et al., 2000; Alhonen et al., 2007).

The logosyllabic character of Chinese also affects how language skills, including listening, speaking, reading, and writing, are acquired and associated with each other. Abilities to read Chinese are linked more closely with people's writing than with listening skills (Tan et al., 2005). By contrast, for alphabetic languages, reading skills are more closely associated with listening skills (Ziegler & Goswami, 2005). Another study, on neural bases of language skills, found that reading Chinese involves a

[2] Besides the 400 often used syllables, there are also 18 other rarely used syllables in Mandarin. Thus the counts of syllables could vary between 400 to 418 depending on the corpus used (Zhang et al., 2008).

link between Broca's area and the supplementary motor area while reading of alphabetic scripts involves a link between Broca's area and Wernicke's area (Siok et al., 2004).

2.2 CHARACTERISTICS OF CHINESE TEXT MESSAGES

When a language is used in a different context and with a different device, how people make use of it to express themselves can be different. Since the theme of this thesis is Chinese text entry on mobile devices, it is highly necessary to understand how Chinese text is used in mobile settings. One of the key services with which users create text intensively from mobile phones is the SMS. Therefore, we built a corpus of Chinese text messages that consisted of two parts: one part was collected by us with pen and paper (Ma et al., 2007), and the other part was licensed from a third party. Based on the corpus, we explored how Chinese users use SMS content to communicate, analyzed linguistic characteristics of Chinese short messages, and built and utilized language models in studies conducted for this thesis (Ma et al., 2007; Alhonen et al., 2007; Liu & Räihä, 2008; Liu & Räihä, 2010). The following paragraphs highlight some of the analysis results.

Table 2.3: SMS category, gender and user group by age all affected length of SMS (Ma et al., 2007)

Factors		**Length of SMS with *SD* (average characters per message)**	**F**	***p***
SMS category	Instrumental SMS	14.0 *(0.37)*	33.79	< .001
	Expressive SMS	16.9 *(0.34)*		
Gender	Male	16.1 *(0.46)*	6.22	< .001
	Female	14.8 *(0.20)*		
User groups by age	Students	14.3 *(0.17)*	5.27	< .01
	Working people	15.2 *(0.16)*		
	Retired people	17.0 *(1.08)*		

The Short Message Service was launched in 2000 in China, and since then it has experienced a phenomenal boom among Chinese users. To understand how SMS users make use of the service, we collected 10,843 text messages with pen and paper, from 114 users, and performed some content analysis (Ma et al., 2007). The results indicated that Chinese users send both instrumental and expressive messages (Ling, 2005). The instrumental ones include those used to coordinate things, exchange information, send requests and soft inquiries, and give responses.

Expressive messages, by contrast, are those used to update personal information or indicate status and express personal emotions. Chinese users sent slightly more instrumental (57.7%) than expressive messages (42.3%). Factors of message category (instrumental or expressive), gender, and user group (retired, worker, or student) all affected the length of SMS messages (see Table 2.3). Expressive messages feature more Chinese characters than do instrumental ones. Female users wrote with more characters than male users in China did. Messages sent by retired people have the most Chinese characters, followed by those of working people, with students using the fewest.

Further, we built a larger corpus of Chinese text messages by combining the SMS collected by us with another part of SMS data that was licensed from a third party. In total, the corpus includes 630,000 text messages and a total of 9,200,000 Chinese characters (Ma et al., 2007). Analysis showed that this corpus contains 4,912 distinct Chinese characters, corresponding to 404 syllables or pinyin marks (these statistics exclude punctuation marks). I also analyzed the frequencies of each distinct character in the corpus. The analysis results indicated that the top 385 and 1,461 most frequently used Chinese characters, respectively, accounted for around 85% and 98% of all characters in the corpus. By comparison, GB2312, a standard Chinese character set for simplified Chinese supported by most Chinese text entry systems, includes 6,763 individual Chinese characters with 404 different syllables, covering about 99.75% of all Chinese characters. Since our corpus is a collection of text messages entered with Chinese text entry methods on mobile phones, its characters are a subset of GB2312. However, our corpus is more representative of the current mobile text entry context than is GB2312, which was defined in 1980 for general purposes.

I also examined the phrases in the corpus. In this thesis, a phrase refers to a Chinese word with more than one character that is also the minimal meaning unit in Chinese (Tanaka-Ishii, Zhou, & Kim, 2007). The results of this analysis indicated that the top 388 and 2,088 most frequently used phrases account for 73% and 89%, respectively, of all phrases in the corpus. Figure 2.4 shows the proportions of the phrases with various numbers of Chinese characters in this corpus. Single characters make up the largest proportion of the corpus, followed by two-character and three-character phrases. Based on this analysis result, I also calculated the percentage of the characters that could be entered via predictive input to be 30.3% for the corpus. Predictive input works so that after a character is entered, the most likely characters that could form a phrase in combination with it are presented for users to select. This analysis result is applied in Section 6.2 of this thesis where I build predictive models for Chinese pinyin text entry methods.

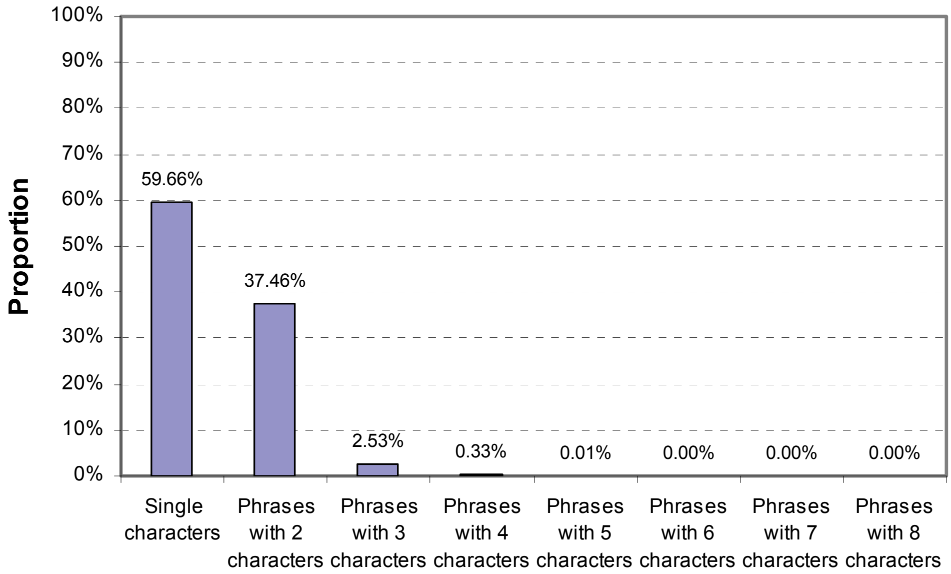


Figure 2.4: The proportion of phrases having different numbers of characters.

For this thesis, we also built and utilized language models from the corpus (Alhonenet al., 2007; Liu & Räihä, 2008; Liu & Räihä, 2010). For Chapter 5, language models were elicited from this Chinese text message corpus and utilized by the speech dictation engine for listing recognition results and predicting the next Chinese characters. For Chapter 6, I transcribed all Chinese characters in the corpus to their pinyin marks and built a language model to estimate the frequencies of any pair of letters in those pinyin marks. I also analyzed characteristics of the pinyin marks from the corpus. This work found that the average number of letters per Chinese character was 3.24 with the pinyin coding system; if the Chinese characters are weighted by their frequency in the corpus, the number drops to 2.88. This is quite different from what Wang, Zhai, and Su (2001) reported, which was probably based on a general corpus of Chinese text: "*On average, each Chinese character's pinyin has 4.2 Roman characters.*" Moreover, a pinyin mark corresponds to about 12 Chinese characters, on average, in this corpus, with a minimum of one to a maximum of 74 characters.

2.3 CHINESE TEXT ENTRY METHODS FOR MOBILE DEVICES

I review the commercial Chinese text entry methods and Chinese speech recognition solutions on mobile devices in this section. The term "commercial solutions" refers to systems that have already been released in commercial products that users can approach and adopt. I cannot cover all the inventions in the domain of Chinese text entry, since there are always new methods emerging – even though they can hardly flourish. The following elements are covered for each type of method: the concept and input process, results of existing research on the interaction process, and advantages and disadvantages.

2.3.1 Keyboard Pinyin Solutions

Pinyin is the standard coding system for Mandarin pronunciation in the form of Latin letters (Zhou, 1953, 2007; Sheng, 1985; Sacher, 1998; Sacher, Tng, & Loudon, 2001; Wang, 2003; Lin & Sears, 2005, 2007; Liu & Wang, 2007; Liu & Räihä, 2008; Liu & Räihä, 2010). A pinyin mark, whose length varies between one and six alphabetic characters, usually consists of a consonant and a vowel, with the exception of a few marks that consist of vowels alone (Zhou, 1953, 2007; Sheng, 1985). Table 2.4 shows the 23 consonants and the 33 vowels.

Table 2.4: The consonants and vowels for pinyin

23 consonants (initials)	b p m f d t n l g k h j q x zh ch sh r z c s y w
33 vowels (finals)	a e i o u v(ü) ai an ao ei en er ia ie in iu ou ua ue ui un uo ang eng ian iao ing ong uai uan iang iong uang

Pinyin is the primary coding system applied with nearly all types of keyboards, including physical and soft keyboards, and for both the QWERTY keyboard for computers and the 12-key keypad for mobile phones (see Figure 2.9). Figure 2.5 (sections a and b) shows the input processes of pinyin text entry methods with a QWERTY keyboard and a 12-key keypad, respectively. Entering Chinese characters with a QWERTY keyboard requires two steps. First, users type in the pinyin mark. Second, the system provides a list of matching Chinese characters sharing the same pinyin mark and users choose the target character. Entering Chinese characters with T9 pinyin input via the 12-key keypad requires three steps. First, the user presses the corresponding keys to enter a pinyin mark. Second, one must select the target pinyin mark from a few options. The key mapping of the 12-key keypad is ambiguous in that one series of key presses may result in multiple options for pinyin marks. If the pinyin mark highlighted happens to be the target one, the user can press the "OK" key to choose it. Otherwise, the user must move the highlighting to the target pinyin mark and choose it. Third, since most Chinese characters are homophonic with several others, the user needs to select the target character from a list of options sharing the same pinyin mark.

Many pinyin text entry methods enable users to complete a pinyin mark by entering its consonant and vowel. An example is the Shuang Pin (Double Spelling) method designed for QWERTY keyboards (Lin & Sears, 2005; Liu & Räihä, 2008). In Shuang Pin, all consonants and vowels including more than one letter (such as "zh," "ch," and "ong") are mapped to a certain key on the keyboard. For example, "zh" (a consonant) and "ong" (a vowel) are mapped specifically to the keys "a" and "y" on the keyboard (Shuang Pin, 2008). Users just need to type "a" and "y" to complete "zhong." The method requires fewer keystrokes; however, users must memorize the key mappings. The Double Spelling soft keyboard in, for example, Windows CE systems includes a consonant keyboard and a vowel keyboard (Shuang Pin CE, 2008). Consonants and vowels are presented to end users separately, and users complete a pinyin mark by typing the required consonant and vowel on the two keyboards. With this method, users do not need to remember any key mapping, but getting familiar with the two-keyboard layouts demands practice. The Double Spelling method is implemented in many pinyin text entry methods as a complement to the normal pinyin method.

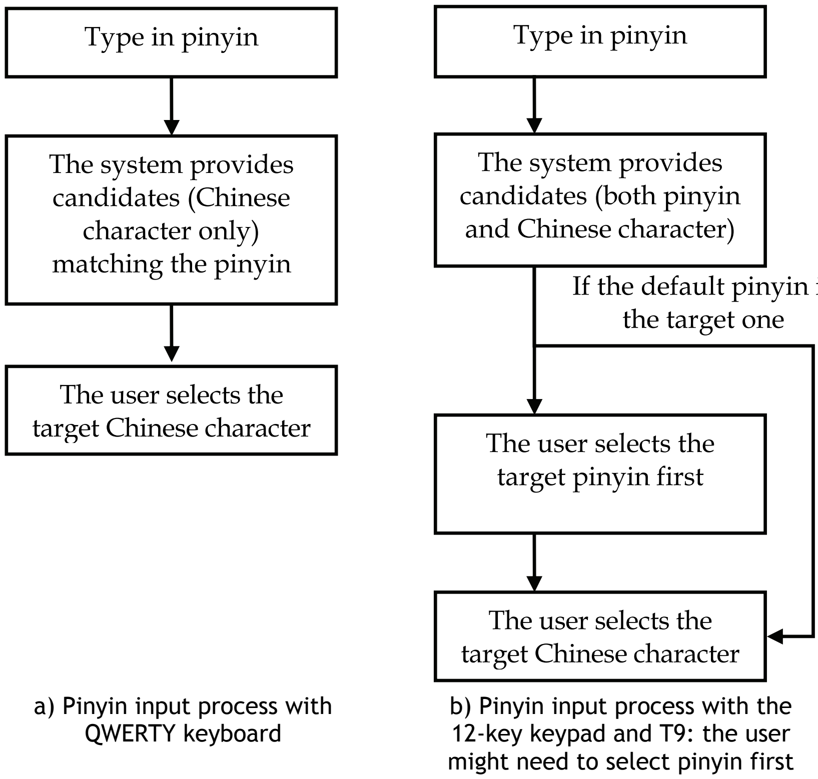


Figure 2.5. The input process with pinyin keyboard solutions.

Phrasal input is a key feature for pinyin text entry. Phrasal input enables users to enter a phrase including more than one character at a time by typing the pinyin marks of associated characters. Phrasal input is already widely supported in pinyin text entry methods designed for the QWERTY keyboard. It significantly decreases the ambiguity level by reducing matching options since characters in a phrase are usually unique. However, ambiguity still exists with the phrasal input methods based on the 12-key keypad. The design and evaluation of those phrasal input methods based on the 12-key keypad are addressed in detail in Section 6.2 of this thesis.

There are two drawbacks to the pinyin coding system that significantly affect user performance. First, there are too many homonyms. With pinyin methods, users are required to select a target character from a list of possible options, which affects user performance. Wang, Zhai, and Su (2001) conducted an anatomical study of a QWERTY-based pinyin method and found that the selection process takes 52% of the total time for the input process. Our analysis of the pinyin character input process in Chapter 6 indicates that the selection processes cost more than 65% of the total time in the input process. Second, completing a pinyin mark requires a number of keystrokes. On average, a pinyin mark includes 2.88 letters according to the analysis of a Chinese SMS corpus in Section 2.2.

There are some published examinations of the Chinese pinyin input process. Lin and Sears (2005, 2007) reported a text entry speed of 4.04 words per minute (WPM) when participants in their study were instructed to enter text with a natural balance of input speed and error rate. Based on how the authors calculated the result, I estimate that their reported text entry rate corresponds to 5.46 CCPM. In Chapter 6 of this thesis, I analyze users' error-free speeds with two types of Chinese pinyin character input. The results indicate that the average error-free speeds were around 19.1 and 21.9 Chinese characters per minute (CCPM) when the predictive feature was, respectively, off and on. There may be a few factors in the great difference between the results from the two studies. First, in the study of Lin and Sears, participants were instructed to enter text at their own speed, with a natural balance between speed and errors, while in my study, participants were asked to enter as quickly and accurately as they could. Moreover, in my analysis of the empirical data, errors and their corrections were removed from the task completion time, so that the empirical results were comparable with the predictive model. Second, people have become more and more familiar with the 12-key keypad and the pinyin input methods in the past few years, which may in another way cause the increase in user speeds.

Shortly after pinyin became the standard coding system for Chinese characters in the Latin alphabet in 1958, the primary schools of mainland China started to teach the pinyin coding system before Chinese characters (Chen & Yuen, 1991). Hence, the pinyin coding system is familiar to many Chinese users. However, it still cannot accommodate all potential users of mobile devices. Many older people have never learned to describe Chinese characters in the form of Latin letters, and this renders pinyin input unfamiliar. Moreover, if people's native dialect is not Mandarin, their dialect will affect how they speak in Mandarin. Mandarin with a heavy accent often results in errors when people produce a pinyin mark (Chen & Yuen, 1991; Li, 2005; Liu, Jaeger, & Nakagawa, 2004; Dai, Liu, & Xiao, 2007). For this group of users, replacement methods may be needed.

2.3.2 Keyboard Stroke Solutions

The stroke methods define the basic strokes composing Chinese characters and map them to several (5–9) keys of the 12-key keypad (Lin & Sears, 2007; Tanaka-Ishii, Zhou, & Kim, 2007). Extra labels on keys for the basic strokes are necessary to assist users in the input process. Users can enter a Chinese character by clicking the corresponding keys for the strokes in a standard order. There are multiple ways to define the basic strokes and their mappings on the 12-key keypad. Figure 2.6 shows five-stroke, eight-stroke, and nine-stroke solutions. Although the three solutions share a few common basic strokes (horizontal line "一," vertical line "|," left curve "/," dot or right curve "、," and hooked "ㄱ"), they are quite different from each other in terms of definitions of the basic strokes and key mappings.

a) The five-stroke method b) The eight-stroke method c) The nine-stroke method

Figure 2.6. The different definitions of the basic strokes and key mappings.

Figure 2.7 shows the input process with stroke methods. Users click corresponding keys to enter strokes of a character, and the system will provide a list of options in real time, from which users select the target character. When entering a complex character, users usually need not input all of its strokes. After entry of a certain number of strokes, the target character will appear in the option list. But for simpler characters, such as "三" ("three" in Chinese), all of the strokes need to be entered, because there are many characters whose first three strokes are identical.

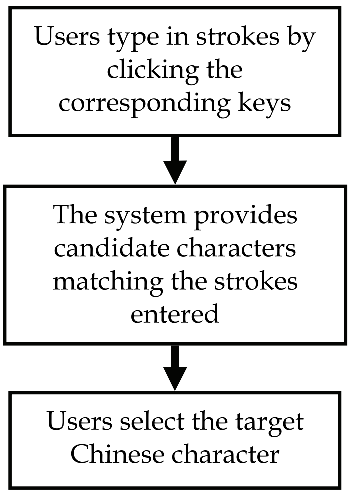


Figure 2.7. The input process with a typical stroke method.

Lin and Sears (2005, 2007) found that labels of strokes on the keys significantly affected user performance with the nine-stroke method. Therefore, they proposed three new label solutions to decrease errors. A follow-up user study showed that user performance with respect to speeds and error rates increased with the addition of more specific stroke labels or examples on the keys. Lin and Sears (2005) also conducted a longitudinal study of the nine-stroke method and found that users can reach a text entry rate of 7.8 WPM after five days of practice. Stroke methods have begun to support phrasal input as well. Users can enter a predefined number of strokes for each character in a phrase, and the system will provide options based on those strokes and language models.

The stroke methods require users to enter Chinese characters in a standard order. Although in most cases that order is coherent with the principle applied normally for writing Chinese characters, which is from left to right first and then from top to bottom, there are still many exceptions. In reality, people actually develop different orders for writing Chinese characters, which do not match this principle. Thus users have to adapt to truly using the standard order, which involves extra learning efforts.

2.3.3 Chinese Handwriting Recognition

Handwriting recognition (HWR) enables people to convert handwritten characters to machine-readable text. HWR technologies can be divided into two categories: on-line and off-line recognition. The term "HWR" is used in this thesis to refer to on-line HWR, which captures and utilizes the trajectory information in the writing processes. As mentioned above, Chinese characters are complex in structure and usually comprise several strokes. This characteristic of Chinese characters is useful for the HWR technology, since complex characters provide more information to the recognition system for differentiation from each other (Dai, Liu, & Xiao, 2007).

Research into Chinese HWR technology began in the late 1970s in China (Ding, 2002; Liu, Jaeger, & Nakagawa, 2004; Dai, Liu, & Xiao, 2007). The technology is mature enough. Recognition rates for freely written and non-cursive Chinese characters, respectively, reach about 98% and 99.8% (Ding, 2002; Wang, 2003; Liu, Jaeger, & Nakagawa, 2004; Dai, Liu, & Xiao, 2007). Moreover, Chinese HWR has been adopted by many commercial products, including computers, mobile devices, and public service systems. There were already about 300 million users of Chinese HWR technique in 2003 (Wang, 2003).

Recent studies of Chinese HWR technology have focused on continuous handwriting recognition, to propose new segmentation methods or possibilities of combining HWR with other input modalities (Wang et al., 2006; Ao et al., 2007; Guo & Jin, 2007; Zou, Yu, & Wang, 2008). Because of the low computing capacities and small displays of mobile devices, discrete Chinese HWR that enables text entry character by character is still the main stream. With discrete Chinese HWR, users first write a character; then the system recognizes the character after it has been completely written. After that, several options for the recognition results are presented, and users can select the target one.

a) Two-box design b) One-box design c) Full-screen design

Figure 2.8. Examples of the full-screen and the box UI designs for Chinese HWR on mobile devices with touchscreens.

There are two types of user interface design for discrete Chinese HWR on mobile devices: full-screen and box designs (see Figure 2.8). With full-screen designs, users may write characters anywhere on the screen. The box designs usually define one or more specific areas (usually two or three), which are squares with frames – hence my term "box designs" – in which one is to write characters.

There are also two ways to signal the completion of a Chinese character to the recognition system and initiate a recognition: via a timeout and by writing in a different box. A timeout is a duration preset in Chinese HWR systems for segmenting Chinese characters. Once the timeout is reached in the writing process, the systems will initiate recognition of the strokes written before it as a Chinese character. The timeout is usually set to between 200 ms and 1 s and can be customized by the end user. Cui and Lantz (2005) conducted an empirical study to consider the most appropriate value for the timeout for Chinese HWR. They collected a large number of data points on the periods between strokes when users were instructed to write Chinese characters on a PDA with a touchscreen. On the basis of the results, they suggested that the optimal timeout values for slow, ordinary, and fast writers should be 750 ms, 500 ms, and 350 ms, respectively. Box designs with more than one box allow users to write Chinese characters in alternate boxes. Once a user starts to write strokes in a different box, the systems start to recognize the strokes written in the previous box as a character. Such a design requires no timeouts and therefore saves some time.

There are also many other design details to consider with Chinese HWR user interfaces. For example, some user interfaces provide soft buttons or a soft keyboard so that users can enter other types of characters – for example, symbols and Roman letters – in the system. Editing functions are also provided in different ways. In Chapter 3, I explore the advantages and disadvantages of the detailed design options in empirical studies.

Wu, Zhang, and Hu (2003) analyzed the discrete Chinese HWR input process on personal computers (PCs) and built models to explain the effects of recognition rate, timeout length, and error corrections on user performance. However, in presentation of the results concerning user performance, task completion time instead of text entry rate was applied.

Chinese HWR is a natural way of entering Chinese text. However, it also has apparent drawbacks. First, both hands are required for use of Chinese HWR on mobile devices. However, mobile devices are usually used in dynamic contexts, where at least one hand is not available. Second, since touchscreens are expensive, mobile products with touchscreens are designed to target those in China who have a relatively high social and economic status.

2.3.4 Chinese Speech Dictation

Speech dictation solutions enable people to enter text by speaking to mobile phones. For mobile speech recognition solutions, the speech recognition engine can be incorporated either into a server or into the client mobile devices directly (Price & Sears, 2005; Alhonen et al., 2007). Speech recognition on mobile devices has been judged to be useful in contexts where there are no hands available and for people who have some physical disability (Silfverberg, 2007). Moreover, the technology of speech recognition has been considered a good opportunity to overcome the obstacles of entering Chinese characters in computing systems. However, commercially successful applications or systems based on speech recognition technology are still limited, as a result of many factors, including that recognition accuracy in real usage contexts is still vulnerable to many constraints (Karat et al., 1999; Karat et al., 2000; Shneiderman, 2000; Silfverberg, 2007).

Speech recognition technology faces two general challenges: issues of speaker-independence and continuous speaking (Entwistle, 2003). But there are greater challenges for Chinese in particular (Li, 2005; Jeng, 2005; Alhonen et al., 2007). First, most Chinese characters are homophonic with some others, which implies that there is no one-to-one relationship between a syllable and a character. Therefore, even if a syllable is correctly recognized, extra user interactions are required for choosing the target character from among many options. Second, there are 10 main dialects of Chinese, which are different in their pronunciation, expression, grammar, etc. Clearly, no single speech recognition engine can meet the requirements of all users. Third, there are multiple tones in almost all dialects. They are usually meaningful and used to specify a character in oral communication in combination with language contexts. Many studies of Mandarin speech recognition have explored how tones might be detected and used in speech recognition systems (Chen et al., 1997; Wang et al., 1997; Chang et al., 2000).

Numerous Mandarin speech recognition systems on personal computers have been built since the early 1990s, and high accuracy levels have been reported for many of them (Gao, Zhong, & Ceng, 1995; Fu et al., 1996; Chen et al., 1997; Wang et al., 1997; Huang et al., 2000). However, in mobile devices, speech recognition technology has seen only limited applications and functions – for example, name-based dialing, menu navigation, or information retrieval (Viikki, Kiss, & Tian, 2001; Chang et al., 2002).

Speech dictation is easy for users to understand because oral communication is one of the most natural means of interpersonal communication (Furui, 2007; Cox et al., 2008). However, explorations of the human–computer interaction processes with Mandarin speech recognition systems are rare. By comparison, many HCI studies have been done to understand effects of recognition rates on user acceptance of English speech recognition and HWR systems (Wolf, Glasser, & Fujisaki, 1991; LaLomia, 1994; Frankish, Hull, & Morgan, 1995; Karat et al., 1999; Karat et al., 2000; Ai, Harris, & Rose, 2006; Munteanu et al., 2006; Zhou et al., 2006; Cox et al., 2008).

Chapter 5 of this thesis presents two user studies on and for an isolated Mandarin short message dictation application running on Symbian Series 60 mobile phones. The first study involved two rounds of user evaluation of the application. The second study used a Wizard of Oz experiment to explore effects of speaking style on user performance and satisfaction alongside other factors, including recognition rate and message length.

2.4 Design and Evaluation of Mobile Text Entry Methods

Mobile devices such as mobile phones cannot gracefully accommodate a QWERTY keyboard and large displays. The 12-key keypad, also called the ITU-T keypad, is the standard keypad incorporated by many phone products (Pavlovych & Stuerzlinger, 2004). With such a keypad, three or four letters are mapped to each key. Hence, when entering different letters on a key, users have to press the key a differing number of times – for example, pressing the "2" key once to get the letter "a," twice in rapid succession for "b," and three times rapidly to get "c." The "multi-tap" method significantly decreases user performance, because, on average, entering a character requires multiple key presses. To improve user performance with the 12-key keypad, many new text entry methods have been proposed and studied with different approaches. A review of the design drivers and approaches will aid researchers and practitioners in their future work.

Figure 2.9. The standard 12-key keypad, also called the ITU-T keypad.

In addition, no design of new mobile text entry methods can advance without involving users (MacKenzie & Soukoreff, 2002; Soukoreff, 2002). Well-planned user studies, via collection of data on user performance and preferences, may assist designers to compare different solutions and make clever decisions. However, practices in conducting user studies are quite different from one case to the next, and a review of the relevant guidelines would be a beneficial guide for user studies on mobile text entry methods.

In the next two sections of this chapter, I review first the design drivers and approaches for new text entry methods and then useful guidelines for conducting user studies of such methods. I do not intend to be exhaustive and cover all existing design and study practices. Instead, I will review those that are relevant for this thesis.

2.4.1 Design Drivers and Approaches

There are a few common goals for designing new mobile text entry methods: high user performance (low error rate and high speed), ease of learning, universality of the solution in both user coverage and contexts, and also being fun to use (Zhai, Kristensson, & Smith, 2005; Zhai & Kristensson, 2007). However, it is hard to meet all goals in one design and reaching one of them often requires the sacrifice of another. Hence, designing a new mobile text entry method often results in a true tradeoff among many factors. For example, a new method with optimal user performance sometimes requires more practice or applies specifically to a certain group of users in particular contexts. For mobile text entry methods specifically, an apparent tradeoff exists between the number of buttons in a keypad (or keypad size) and ambiguity level (MacKenzie & Soukoreff, 2002; Soukoreff, 2002). Fewer keys would probably increase the ambiguity level. Moreover, text entry tasks involve sub-processes at different levels: motor movements, human perception, and other cognitive processes (Zhai & Kristensson, 2007). Optimization of one sub-process often requires devoting more user effort to other sub-processes.

Achieving optimal user performance is crucial motivation for designing new mobile text entry methods. MacKenzie and Soukoreff (2002) summarized two approaches for optimizing user performance with mobile text entry methods: by language prediction and through minimizing motor movement. Stocky, Faaborg, and Lieberman (2004) also reviewed design practices with the same categorization system.

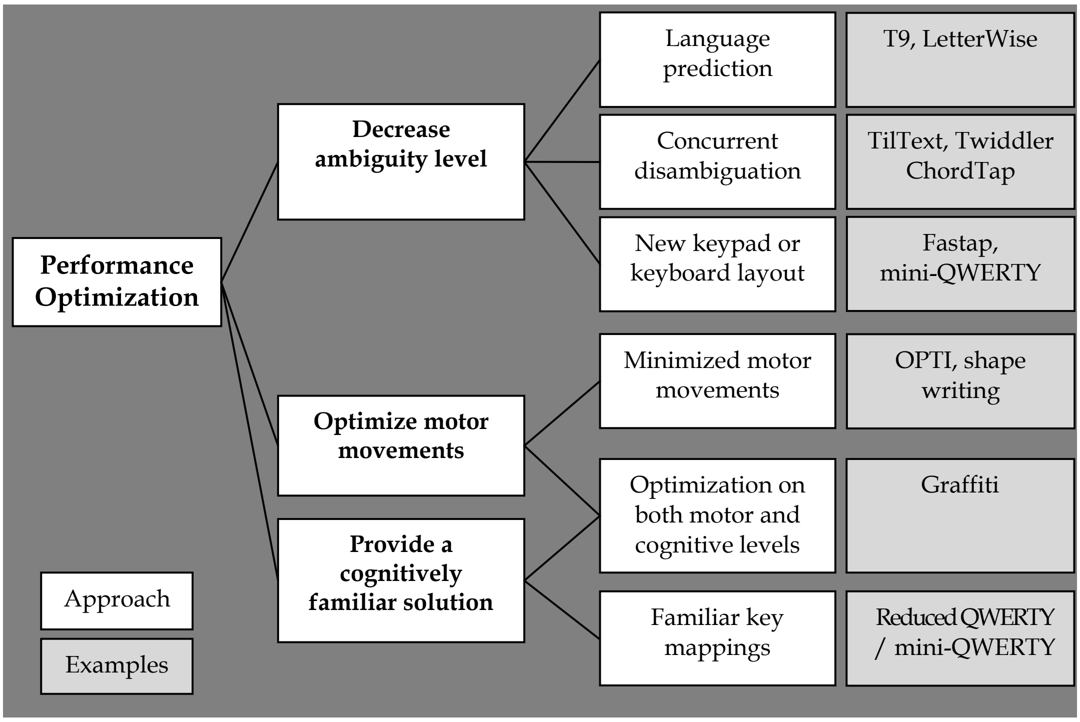


Figure 2.10. Approaches and examples for designing mobile text entry methods with optimal user performance.

In this thesis, I extend the classification to three approaches, to cover more solutions. These approaches are disambiguating, optimizing motor movements, and providing a cognitively familiar solution (see Figure 2.10). The disambiguation approaches are aimed at decreasing the average number of keystrokes per letter, so that the level of ambiguity caused by the key mapping of the 12-key keypad can be decreased. There are three major ways to decrease ambiguity: language prediction, concurrent disambiguation, and new keypad designs with more keys. In contrast with the multi-tap method, the T9 method integrates dictionaries to predict the potential words for a series of key clicks. This allows users to enter a character with a single keystroke, improving user performance. Other predictive solutions taking advantage of language models, such as the "LetterWise" approach, were proposed later to make further improvements (MacKenzie et al., 2001).

Concurrent disambiguation approaches enable users to click more than one key at a time or take some other actions while clicking a key so as to enter a specific letter (Wigdor & Balakrishnan, 2004). ChordTap (Wigdor & Balakrishnan, 2004), Two-Thumb chording (Patel, Clawson, & Starner, 2009), and Twiddler (Lyons et al., 2004; Lyons, Plaisted, & Starner, 2004; Starner, 2004) are typical examples of the former approach and “TilText” (Wigdor & Balakrishnan, 2003) a typical example of the latter. Moreover, some new designs of small keypads that incorporate more keys were also proposed to decrease ambiguity. For example, the Fastap keyboard includes keys for the numbers 0 to 9 and the 26 letters in alphabetical order (Levy, 2002; Cockburn & Siresena, 2003).

Another category of approaches is aimed at optimizing user performance by decreasing the distance of motor movements in text entry tasks. Many approaches for minimizing motor movement have been explored with soft keyboards[3], because layouts of soft keyboards are flexible and cheap to change. Leventhal, McKeeby, and Mynatt (1991) explored effects of shape and character layout on user performance. Isokoski (2004) proposed a pie menu popping up around a pressed key for quicker clicking of the next character. The vowels, “backspace,” and the “space” key were included in the pie menu. MacKenzie, Zhang, and Soukoreff (1999) proposed two new layouts, called OPTI I and OPTI II, minimizing the overall motor movement distance according to Fitts’ law and the relative frequency of bigrams (pairs of letters) for English (Soukoreff & MacKenzie, 1995); I will discuss Fitts’ law in detail in Chapter 6. The Metropolis algorithm was applied to minimize the motor movement time for entering English (Hunter, Zhai, & Smith, 2000; Zhai, Hunter, & Smith, 2000; Zhai, Smith, & Hunter, 2002). Zhai, Smith, and Hunter (2002) also presented other approaches to optimize the layout of soft keyboards.

The last approach for optimizing performance is to provide a familiar layout or key mappings for users so that they can reach optimal performance in a shorter time. In recent years, many new keypads have been proposed that are based on the familiar QWERTY layout. Green et al. (2004) proposed a reduced QWERTY keyboard including only two lines of keys. Hwang and Lee (2005) and Ryu and Cruz (2005) also proposed solutions for mapping the 26 letters on the 12-key keypad by mimicking the QWERTY layout. Although new reduced QWERTY solutions sometimes don’t decrease ambiguity, users find it easier to start with them because the layouts are familiar. Moreover, many mini-QWERTY keyboards are incorporated into mobile products and explored by researchers (Clarkson et al., 2005).

Many new mobile text entry methods have been produced that utilize more than one approach. For example, Dasher (Ward, Blackwell, & MacKay, 2000, 2002) made use of both language prediction and minimizing of motor movements. Another example is shape writing: a

[3] A soft keyboard is a keyboard implemented on a touch-sensitive display. To enter a word, users can type its characters one by one with a stylus or their fingers.

concept based on a soft keyboard with which users can enter an English word by traversing through its soft keys without lifting a pen. After becoming familiar with the gesture patterns of a word, users can write shorthand gestures to enter the word without really reaching the corresponding soft keys (Kristensson & Zhai, 2004; Zhai & Kristensson, 2003, 2007). And the key design approaches utilized by shape writing include minimization of the motor movement distances via shorthand gestures and later the disambiguation process by optimizing the layout of a soft keyboard for English. Another example, MessagEase, made use of both language prediction and minimization of the motor movements and proposed a new keypad for mobile devices (Nesbat, 2003).

Universal mobile text entry methods covering more user groups and usage contexts were also subject to great expectations. When compared with other computing systems, mobile devices are more widely used by end users worldwide. Users of mobile devices are diverse in many respects – for example, literacy level, gender, age, and disabilities. Shneiderman (2003) pointed out that "*designing for experienced frequent users is difficult enough, but designing for a broad audience of unskilled users is a far greater challenge.*" Many techniques have been developed to enable people with disabilities to enter text on mobile devices. Wobbrock and Myers (2006) designed a trackball text entry solution for people with motor impairments. Majaranta and Räihä (2007) worked on methods of text entry by eye gaze so that people with severe motor disabilities can still enter text to computing systems. Lagoá et al. (2007) designed a mobile text entry solution for visually impaired people.

Mobile devices are also frequently used in diverse contexts, which sometimes impose restrictions on the availability of human modalities (Ronkainen et al., 2010). How to make use of the available modalities in such difficult contexts has become an important research topic. Studies have been conducted to understand the effects of different aspects of context, such as moving, on user performance. Multimodal solutions have been proposed to enable the usage of mobile devices in such moving contexts (Brewster et al., 2003; Lumsden & Brewster, 2003; Kernchen, Mossner, & Tafazolli, 2005; Serrano et al., 2006; Ronkainen et al., 2010).

New technologies also play an important role in the design of new mobile text entry methods. For example, the capacitive touch-sensitive display, which supports multiple touch points, has begun to be incorporated into many mobile devices (Benko, Wilson, & Baudisch, 2006; Shin et al., 2009). A recent study reported a new text entry method for Korean that is based on the multitouch technique (Shin et al., 2009). Some new technologies enable more options for feedback in text entry tasks. For example, existing results indicate that the tactile feedback can increase user performance with soft buttons (Poupyrev & Maruyama, 2003; Hoggan, Brewster, &

Johnston, 2008; Dunlop & Taylor, 2009; Hoffmann, Spelmezan, & Borchers, 2009).

2.4.2 User Studies of Mobile Text Entry Methods

User studies are necessary in the interaction design process of new mobile text entry methods. On one hand, user studies are effective ways of involving users in design processes. For example, comparative user studies of design options can aid in smart decisions. On the other hand, diagnostic user studies can help to improve a design by collecting data on user behaviors and feedback. To ensure both internal and external validity of user studies, many guidelines have been proposed and discussed. Internal validity refers to the degree of match between what is tested and what was intended to be tested in a study (Eisenhardt, 1989; Grimes & Schulz, 2002; MacKenzie, 2007). External validity refers to how much the results from a piece of research can be utilized and extended to other contexts (Eisenhardt, 1989; Grimes & Schulz, 2002; MacKenzie, 2007). In this section, I summarize relevant guidelines and measures for user studies of text entry methods.

First, comparative studies are recommended if conditions permit (MacKenzie & Soukoreff, 2002; Curran, Woods, & Riordan, 2006; MacKenzie, 2007); in these studies, more than one design solution is studied with users. Comparative studies have two key advantages: first, comparative studies are more efficient in that more than one solution can be studied at once; second, comparative studies can assist designers in their decision-making by reflecting each other's pros and cons with users.

Second, **a text copying task with memorable short sentences is preferred** over a text creation task in user studies of text entry methods. A text copying task requires participants to enter the text as presented, with the text entry methods given, while a text creation task is what people usually perform normally: thinking of the contents, organizing the text to be entered, and entering it. Text creation tasks are more like what people normally do with text entry systems; however, there are too many factors and processes entangled in such tasks that will negatively affect how the data can be explained and compared with those from other studies. Comparatively, text copying tasks are much easier to control: people just need to follow the text presented and enter it. Moreover, to decrease the conversion involved between entering and following up on the text presented, researchers suggest that the text given be brief and easy to remember at a glance (MacKenzie & Soukoreff, 2002; MacKenzie, 2007). This way, less of various distracting factors is involved in the copying tasks.

Speed and error rate are the two primary objective metrics applied in user studies of text entry methods. In measurement of text entry speeds for English, researchers usually adopt the measure of words per minute (WPM) to present speed results: WPM is calculated by multiplying characters (letters) per second (CPS) by 60 and dividing by 5, which is the number of characters per word set for English. In this thesis, I utilize the measure of Chinese Characters per Minute (CCPM) instead of WPM to present results on text entry rate. A word in Chinese often includes more than one Chinese character and thus corresponds to a phrase in English. Instead, a Chinese character is comparable to a word in English.

There are also many proposals addressing how to define and present results for error rates. A typical categorization defines three types of error rates: corrected error rate, uncorrected error rate, and total error rate (Soukoreff & MacKenzie, 2003; Wobbrock, 2007). The corrected error rate is the number of characters that were wrongly entered but corrected later, divided by the number of all characters, including the transcribed characters and the corrected characters. The uncorrected error rate is the number of wrong characters in the transcribed text divided by the number of all characters. The total error rate is all errors, including the number of corrected and not corrected characters, divided by the number of all characters. For more explanation of the speed and error rate measures, see the review by Wobbrock (2007).

Keystrokes per character (KSPC) is a measure proposed by MacKenzie (2002a) specifically for mobile text entry methods. It refers to the average number of keystrokes or clicks required for entering a character of a specified language with a specified text entry technique (MacKenzie, 2002b, 2007). The KSPC value varies with the text entry technique, the keyboards utilized, and language. For example, for entering English with a QWERTY keyboard, the KSPC figure equals 1. But for smaller keyboards with fewer keys, the KSPC value is usually greater than 1 for English. With a specified text entry technique and language model, KSPC can be calculated with the following equations at character and word level, respectively:

$$KSPC = \frac{\sum (K_c \times F_c)}{\sum (C_c \times F_c)} \tag{2-1}$$

In the above equation, K_c refers to the number of keystrokes required for entering a character c and F_c represents the frequency of the character in the language. C_c represents the size of the character (typically, 1).

$$KSPC = \frac{\sum (K_w \times F_w)}{\sum (C_w \times F_w)} \tag{2-2}$$

In Equation 2-2, K_w represents the number of keystrokes required for entering a word w and F_w is the frequency of the word in the language. C_w is the number of characters in the word. Empirically, the KSPC value is also affected by many other factors, including errors and individual differences. When empirical KSPC results are reported, how the number is calculated should be presented precisely, to avoid confusion.

KSPC is an important and simple characteristic index for mobile text entry solutions. It is quite easy to understand and allows people to compare text entry methods. However, the disadvantages of KSPC should also be noted. With KSPC, two key factors, the key clicks and language model, are taken into account; however, the differences in motor movement times among the pairs of key clicks are overlooked. Therefore, to avoid bias, KSPC is best applied in combination with other measures.

I have conducted both qualitative and quantitative user studies, to seek answers for different research questions in this thesis. Since both independent and dependent variables need to be clearly defined in quantitative studies, such studies are more focused on solving specific design or research questions (MacKenzie & Soukoreff, 2002; MacKenzie, 2007). Results of quantitative studies were analyzed statistically to show the effects of the independent variables on the dependent variables. Qualitative user studies are applied more for seeking user feedback on broader topics (MacKenzie & Soukoreff, 2002; MacKenzie, 2007) – user acceptance or preference of a specific interaction design, comments on design details, etc. Quantitative studies can assist designers to choose better interaction design solutions, while qualitative user study results can be applied for diagnostic purposes and refinement of design details. Readers are directed to comprehensive reviews of user evaluation of mobile text entry solutions in relevant papers and books (MacKenzie & Soukoreff, 2002; MacKenzie & Tanaka-Ishii, 2007).

3 A Touchpad Chinese HWR Solution

3.1 INTRODUCTION

The Chinese text entry solutions based on the 12-key keypad, including the pinyin and stroke solutions, can hardly accommodate all potential users in China, because of the inherent disadvantages of the pinyin and stroke coding systems. First, a non-native speaker of Mandarin often makes errors with the pinyin coding system due to the influence of his or her native dialect (Dai, Liu, & Xiao, 2007; Tang & van Heuven, 2009). Moreover, the pinyin coding system is still young and was made a standard only in 1958, so most senior users are unfamiliar with it. Second, the stroke coding system typically requires users to enter characters exclusively in a standard order. But people actually develop different writing orders for Chinese characters in real life, which do not match the standard order. In those cases, extra training is always needed.

"*Immediate usability*" is seen as important for new text entry solutions designed for mobile devices (MacKenzie & Soukoreff, 2002). A mobile text entry method with good immediate usability enables users to achieve satisfactory performance within a short time. Immediate usability can be measured by studying user performance and preferences while the time of users' exposure to an interactive system is controlled (MacKenzie & Soukoreff, 2002; Zhai & Kristensson, 2007). Chinese handwriting recognition has potential for good immediate usability (Liu, Ding, & Liu, 2009). First, most potential users are already able to write Chinese characters anyway, so a writing-based system is highly likely to accommodate more users. Moreover, since writing Chinese characters is familiar to the prospective users, less training probably is required. Also, since writing practice can improve Chinese children's reading ability, according to recent studies (Tan et al., 2005; Siok et al., 2004), Chinese HWR in mobile computing systems would be more beneficial, since it is not only a text entry tool but also a training tool for children (Tian et al., 2010).

Chinese HWR has already received much attention from both practitioners and researchers. Sacher, Tng, and Loudon (2001) viewed the design of keyboard-based Chinese text entry as a deficient approach and proposed a "keyboard-less" solution for personal computers integrating Chinese HWR and speech recognition technologies. Dai, Liu, and Xiao (2007) pointed out that the complex structure of Chinese characters makes HWR technology feasible. Ding (2002) summarized the technological developments for Chinese HWR and pointed out that the technology was already mature. When summarizing the research and practice of HCI in China, Wang (2003) also commented that Chinese HWR is a successful area in terms of both the technology itself and its applications.

However, acceptance of Chinese HWR by users also faces challenges. First, touchscreens are expensive. Mobile products supporting Chinese HWR usually incorporate touchscreens, which are more expensive than normal screens. A few years ago, such touchscreen-based mobile products were usually designed for those people in China who had higher economic and social status. Second, Chinese HWR requires more computing capacity than does keyboard-based text entry, which in another way increases costs. Third, mismatches exist between product offers and user needs in the market. For example, users who are expecting Chinese HWR may not desire a product supporting touch interactions only, because touch interactions often are associated with little tactile feedback in comparison with physical keys and complex operations such as double clicks.

In this chapter, an affordable Chinese HWR method based on a touchpad is proposed and designed for mobile phones. First, I explore the design space, identify the design opportunities, and propose the new method. Second, I present the user-centered interaction design process and summarize the user studies. Third, I present in detail three representative studies in the UCD process. Finally, I discuss the results and draw conclusions.

3.2 DESIGN SPACE ANALYSIS

Before one designs a new Chinese text entry solution, it would be helpful to explore the design space: identifying the key categorization dimensions, organizing the existing solutions in the space, and indicating design possibilities (Wigdor & Balakrishnan, 2004). As mentioned earlier, Chinese users of mobile phones vary greatly in their profiles: from the very highly educated to semi-literate users and from the wealthy to the poor. For covering the various users of mobile phones in China, immediate usability and affordability of the technology are two critical factors. Thus I define these as the two design dimensions and categorize the existing text entry methods on that basis. Figure 3.1 shows the map of the existing Chinese text entry methods in the space, from our analysis and estimations.

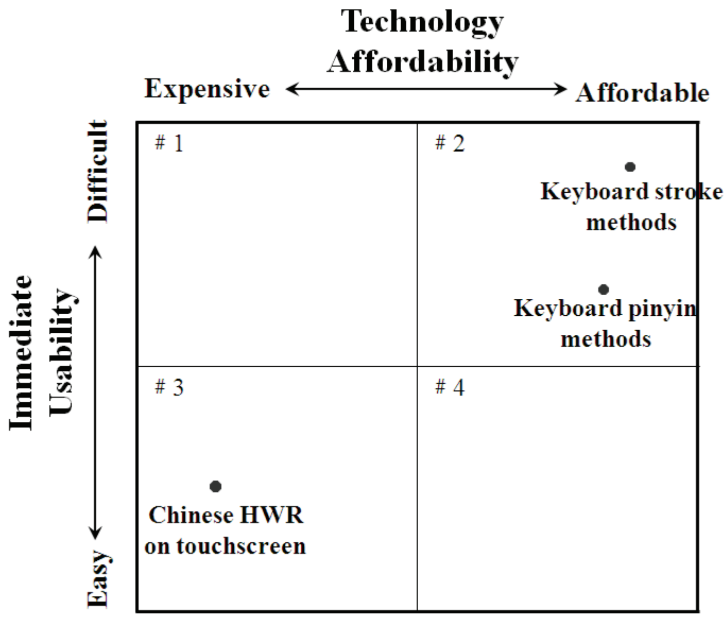


Figure 3.1. Approaches and examples for designing mobile text entry solutions with optimal user performance.

According to the figure, there is a lack of affordable Chinese text entry methods that are good in immediate usability. On the basis of analysis of the design space, we have proposed a Chinese HWR solution based on a touchpad for mobile phones (Liu & Liu, 2008; Liu, Ding & Liu, 2009). A touchpad is a touch-sensitive pad that is often integrated in laptop computers as an input device. In the concept, a touchpad would be mounted in the mobile phone underneath the 12-key keypad (while the keypad can be open) so keypad-based text entry solutions and touchpad-based Chinese HWR are both supported in the product concept. I applied the UCD approach to ensure usability of the interaction designs.

In the next few sections, I first explain the process for development of the idea into specific interaction designs. Then I present three representative user studies completed during the interaction design process for the touchpad Chinese HWR. Not only do the studies show how the UCD paradigm was applied, but also the study results are of value for detailed design of the Chinese HWR technique for mobile devices.

3.3 THE APPLIED USER-CENTERED DESIGN PARADIGM

To ensure immediate usability of the design results, the UCD approach was applied. Figure 3.2 shows the customized UCD paradigm in the interaction design process. It includes three modules for user study and one module for design and prototyping. In the first module for user study (E1 in Figure 3.2), benchmark studies were conducted with users to understand the existing design possibilities and user preferences. In the E2 user study module, the design questions were broken down into smaller ones and explored separately in depth. In the E3 user study module, the initial interaction designs were evaluated with a working prototype. Before the prototype evaluation, there was also a design and prototyping module in which a few interaction designs were designed and a working prototype was built.

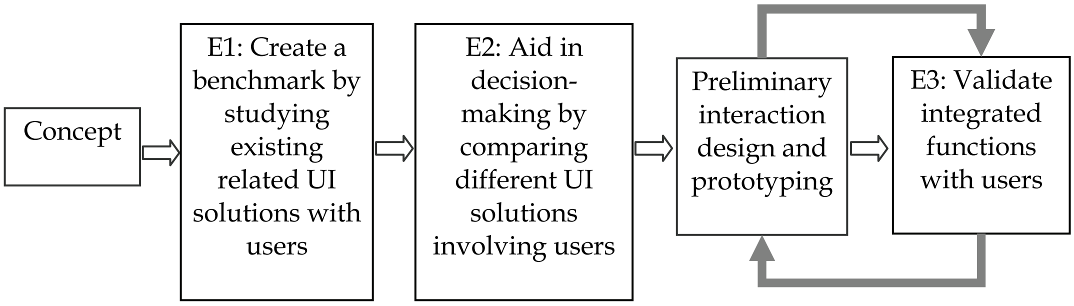


Figure 3.2. The user-centered design paradigm used in the study.

Table 3.1 lists the eight studies that were conducted in the three user study modules. In these studies, both qualitative and quantitative methods were applied to collect user data to answer different design or research questions. In the E1 user study module, I conducted two benchmark studies, on, respectively, the UI layouts and stylus ergonomics for Chinese HWR (S1 and S2 in Table 3.1). In the E2 user study module, I conducted studies of user performance and preferences with different detailed designs (S3 to S7 in Table 3.1). Then in the E3 user study module, I evaluated the concept and designs with end users and a working prototype (S8 in Table 3.1).

Three user studies, two qualitative and one quantitative, are presented in detail in the next three sections of this chapter. They are chosen from the three user study modules, in order, because they were representative in terms of the methodology applied and the problems solved and their results also shaped our design maximally. The first study is the benchmark evaluation of the user interface layouts for Chinese HWR with small devices with integrated touchscreens. The second study was a quantitative one exploring and analyzing user performance with different UI styles: a full-screen design and a design with three input boxes. The third study was a prototype-based evaluation of the initial interaction designs.

Table 3.1: List of user studies conducted in the interaction design process

Module	Objectives	Studies in the module
E1: Build a benchmark by studying related designs	To create a benchmark understanding of the design possibilities and user preferences	S1: Benchmark study of the UI layout for Chinese HWR S2: Benchmark study of stylus ergonomics for Chinese HWR
E2: Narrow down and conduct studies to check user preference and performance with different interaction options	To collect user preference and performance with different interaction options S3: To check whether pen traces for handwriting are needed and whether pen traces displayed on the phone display instead of on the touchpad where people write characters (pen traces displayed at a distance) are acceptable S4: To determine the preferred audio feedback for handwriting S5: To understand the effect of the orientation of the listing of Chinese characters on user performance and preference S6: To understand effects of user type and UI designs on user performance S7: To check the proper sizing of an input box for writing Chinese characters	S3: Pen trace study for touchpad Chinese HWR S4: Proper audio feedback for writing Chinese characters S5: Effect of the orientation of the list of Chinese characters on selection performance S6: User performance study of different UI designs with both novice and expert users of Chinese HWR S7: Sizes of the input area for writing Chinese characters
E3: Evaluate the concept and designs with working prototypes and users	To validate the concept and the interaction designs with a working prototype	S8: Prototype study

3.4 A Benchmark Study of Chinese HWR UI in Mobile Devices

3.4.1 Objectives

The benchmark study of existing user interface layouts for Chinese HWR set out to achieve two goals: first, to understand the design possibilities for our concept and, second, to understand user preferences with the existing designs. This study was part of the E1 module in the UCD paradigm in Figure 3.2.

Understanding the landscape is critical before creation of a specific interaction design. The closest design to our concept is the UI for Chinese HWR in mobile products with integrated touchscreens, because the ways and contexts of using them would be very similar to how our concept will be used. However, a quick analysis of such user interfaces shows that different designs coexist. Moreover, no user studies were found to discuss their advantages and disadvantages. The following is a summary of the options for designing Chinese HWR for mobile devices:

- *Full-screen design vs. box design*: Box designs require users to write Chinese characters in one or a few predefined boxes, while with full-screen designs, users can write characters almost anywhere on a display.

- *Discrete character input vs. continuous input*: Users can enter text only character by character with the discrete character input. Between any two characters, users need to wait a set amount of time to initiate the recognition. I refer to this period of time as a timeout. Once a character is entered, the user can write the next one. With the continuous input style, users can write characters alternately in different input boxes without waiting for timeouts.

- *Presentation of the recognition results*: Some list the recognized characters horizontally and some others in grids.

- *Different system settings*: For example, timeout settings are very different in different designs.

- *Different editing functions:* For example, entry of other characters, including Latin letters, numbers, and punctuation, is supported in different ways; common editing functions, including "Enter" and "Backspace," are also supported in different ways.

Therefore, it is necessary to understand user preferences with such different UI designs, and the results can guide our designs.

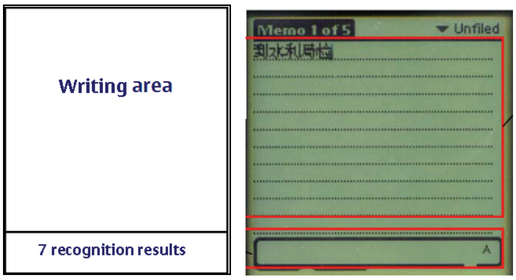
3.4.2 Method

The UI and Apparatus Evaluated

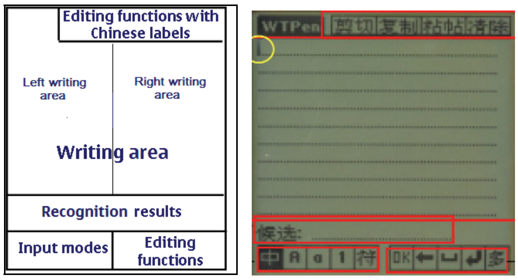
Four Chinese HWR UI solutions were evaluated by participants. Table 3.2 and Figure 3.3, respectively, show the characteristics of each solution and UI layout.

Table 3.2: Characteristics of the UI layouts evaluated for Chinese HWR

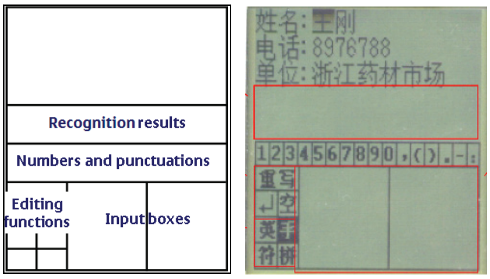
	Devices	UI styles	Input style supported	Number of writing areas
Solution 1	IBM WorkPad	Full screen	Discrete character input	1
Solution 2	IBM WorkPad	Full screen	Continuous input	1 or 2
Solution 3	MBA 998	Input box (Width: 2.2 cm, Height: 2.9 cm)	Continuous input	2
Solution 4	Motorola 6288	Input box (Width: 1.65 cm, Height: 1.65 cm)	Discrete character input	1



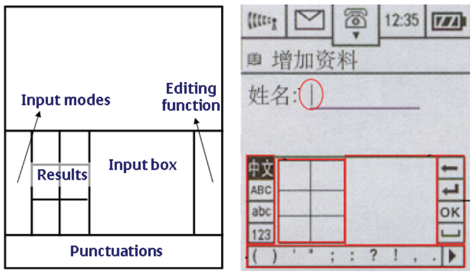
a) UI 1: A full-screen design - users can write a character anywhere in the display area and seven recognition results are shown in the area below.



b) UI 2: A full-screen design - besides a writing area and a result list, this option provides extra soft buttons for editing and different input modes. Users can write more than one character in the writing area.



c) UI 3: A two-box design - use of two input boxes enables continuous writing, and virtual buttons are provided for editing, input modes, numbers, and frequently used punctuation.



d) UI 4: A one-box design - this option offers character input only, with soft buttons provided for editing, input modes, and punctuation.

Figure 3.3. Layouts of the Chinese HWR interfaces under evaluation.

Participants

Sixteen participants, half male and half female, were recruited to take part in the study. Of these, 12 were students and the other four already working. The ages ranged from 16 to 30 years, with an average age of 22.3 (*SD* = 4.61). Half of the participants had experience of using Chinese HWR on mobile devices, and the others were first-time users of it. Fourteen participants were right-handed, and the other two were left-handed.

Design, Tasks, and Procedure

The study was a within-subject design. All participants were instructed to complete three text entry or editing tasks with all designs. Testing orders were counterbalanced with the Latin-square technique. Table 3.3 shows the three tasks.

The study was conducted in a lab environment, with a coordinator and a participant present in one room and observers in another room. Observers in the other room monitored user behavior via a TV set connected to a video camera in the test room. The participant was first briefed about the objectives of the study and shown around the two rooms for obtaining of consent. Then the profile of the participant was collected. After that, the think-aloud protocol was introduced and rehearsed by the participant with a simple task. After this, the participant was instructed to complete the predefined tasks with the think-aloud protocol. The whole test session was recorded with a video camera for further analysis. After a task was completed, the participants were asked to evaluate its difficulty level. Finally, the participant was presented with a gift.

Table 3.3: The three tasks

Task 1	*任务 1：请在商务通的记事本里面输入下面的文字：* 据说，VeRi 公司有 69 名员工。
Translation of Task 1	*Task 1: Please enter the following text in Notebook on the MBA 998:* It is said that the VeRi company has 69 employees.
Task 2	*任务 2：请在商务通的记事本里面输入下面的文字：* 到火车站接货 到水利局检查
Translation of Task 2	*Task 2: Please enter the following text in Notebook on the MBA 998:* I will go to the railway station to receive the goods and then go to the Water Conservancy Bureau to check their work status.
Task 3	*任务：您的记事本中已经有如下内容：* 西湖博览会上看家具 *请继续添加如下内容：* 新 WsCh 家具：全套¥90 万 *请删除"全套¥"这几个字。*
Translation of Task 3	*Task 3: Please edit the following notes with the MBA 998:* *There is already a message in your notebook on the MBA 998:* Go to the West Lake Expo to see the designs of furniture *Please add the following content:* New furniture from WsCh: ¥900 thousand for a whole package *Please delete* "¥ for a whole package" *from the content.*

3.4.3 Results and Discussion

The box designs were preferred over the full-screen designs. Of the 16 participants, 12, including five experienced and seven novice participants, preferred the box designs over the full-screen designs. A few interaction problems were observed with the full-screen designs: First, novice users didn't know how to delete text in the full-screen design without any virtual buttons. Second, the pen traces, which provide the visual feedback while the participants write a character, sometimes overlapped with the text entered in the editor, which was annoying to some participants. Third, another problem noticeable with the full-screen designs was that some pen click actions were often mixed up with character writing actions, which resulted in errors. Short strokes such as a dot are quite common in Chinese characters. However, they are not easily differentiated from pen click actions.

Continuous input was preferred. With both full-screen and box designs, 14 out of the 16 participants preferred the designs supporting continuous input (designs 2 and 3). All participants believed that these were more efficient than the discrete character input. However, some problems were identified with the continuous input. First, the continuous handwriting recognition was not exactly the same as what people usually do with pen and paper. The designs support continuous input by providing more than one writing area. Users can write characters continuously, via alternate areas, and writing the next character in the other area will automatically initiate the recognition of the previous character. In design solution 3, two input boxes were provided to end users. In design solution 2, the whole display was divided into two areas though they were not indicated with clear frames. Second, it was observed that error corrections with the continuous input cost more than error corrections with the discrete input. When users write characters continuously, they usually detect the recognition errors after writing a few more characters. If they want to correct the errors, they have to move cursors back and write the characters again. This process requires both extra cognitive processing and additional motor actions.

Participants expected soft buttons for different types of characters. Table 3.4 shows the input modes supported by each design solution and how to switch input modes. Twelve participants preferred to switch input modes with soft buttons. According to participants, all input modes, including those for Chinese characters, Roman letters, numbers, and punctuation, should be presented with soft buttons with the active mode highlighted. Thirteen participants believed that at least some other characters, especially numbers and punctuation, should be entered with soft keyboards. With design 3, the participants had to switch to a soft keyboard to enter English characters and most participants considered it acceptable. It was also observed that some participants were likely to make errors while entering simple punctuation with HWR, which implies that some other ways of entering punctuation should be supported.

Table 3.4: The modes and mode switch methods supported by each solution

	Chinese	Latin alphabet types	Numbers	Punctuation	Switch method
Solution 1	Handwriting	Handwriting	Handwriting	Handwriting	Only HWR supported
Solution 2	Handwriting	Handwriting	Handwriting	Handwriting	Virtual keys for character modes
Solution 3	Handwriting	Virtual keyboards	Virtual keys	Virtual keys	Virtual keys for modes
Solution 4	Handwriting	Handwriting	Handwriting	Virtual keys	Virtual keys for modes

Some extra editing functions were also evaluated in the study. Table 3.5 shows the function details for each solution. Fourteen participants preferred the I-bar cursor over the highlighted characters. Moreover, the participants were expecting real-time pen traces for written strokes, so delays were not acceptable. Twelve preferred soft buttons for the deletion function; this was a particularly strong preference with novice users. Some experienced users were efficient with the Graffiti gesture of deleting: a stroke drawn from right to left. Ten participants preferred the timeout setting in solution 4, which ranges from 0.4 to 1.0 seconds and enables personalized settings. Cui and Lantz (2005) suggested that the timeout should be set to 350 ms, 500 ms, and 700 ms for slow, medium, and faster writers, respectively. Fourteen participants preferred a horizontal list, which was consistent with results from a previous study (Heloisa & Ravindra, 1998).

Table 3.5: The editing functions supported by each solution

	Shape of cursor	Pen traces	Delete function	Timeout for recognition
Solution 1	I-bar	Some delay	Graffiti gesture of "←"	Not known
Solution 2	I-bar	Some delay	Graffiti gesture of "←" and a soft key	0.4 seconds
Solution 3	Highlighted character	Some delay	Soft key	1.0 second
Solution 4	I-bar	Real-time	Soft key	Between 0.4 and 1.0 seconds

The key finding of the study was that the box solution with virtual buttons for different modes and editing functions was preferable to the full-screen designs without any virtual buttons.

3.5 Effects of User Type and UI Style on User Performance

3.5.1 Objectives

Since handwriting recognition was viewed as the method of choice for those having difficulty with other methods, it was important that the solution provided be able to achieve a sufficient level of efficiency and ease of use. In the previous study, it was found that the participants preferred box designs and continuous Chinese HWR over full-screen designs and the discrete character input. The objective in this study was to understand the effects of user type and UI style (full-screen vs. three-box designs) on user performance and user behaviors in the interaction processes. This study was part of the E2 module in the UCD paradigm shown in Figure 3.2.

3.5.2 Method

Design

The experiment was a 2×2 mixed factorial design. The two factors were *user group*, with novice and experienced users of Chinese HWR, and *UI* *style*, with a full-screen design and a three-box design. The user group and UI style are the between-subjects and within-subject factor, respectively.

Participants

Twenty participants, half male and half female, were recruited for the study. They were all local student interns or members of staff working at the Nokia Research Center in Beijing. Their ages ranged from 22 to 34 years, with a mean of 27.8 years (*SD* = 3.12). All participants were right-handed. Eleven of the participants were novice users, and the others were experienced users of Chinese HWR. Novice users were defined as those with no prior experience of using Chinese HWR. Trained users were defined and recruited as those who were daily users of Chinese HWR on mobile devices for the month before this study (four of the experienced users worked in the field of Chinese HWR as developers or UI designers and used Chinese HWR every day). For the 20 users, testing orders with the two UI designs were counterbalanced.

Tasks and Materials

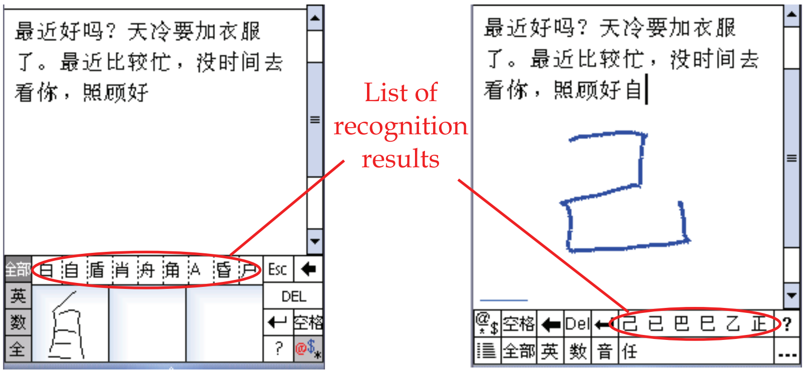
The experimental task for participants to complete was to copy a Chinese text message having 32 characters with both a full-screen and a three-box design (see Table 3.6). The 32 characters included five punctuation marks (two commas, two periods, and one question mark) and 27 Chinese characters. The punctuation marks also served as segmentation in the text message so that participants could remember the message easily.

Table 3.6: The editing functions supported by each solution

Text message to be entered by participants	最近好吗？天冷要加衣服了。最近比较忙，没时间去看你，照顾好自己。
Meaning	How are you recently? It becomes cold so be sure to wear enough clothes. I was too busy recently to take care of you, so take good care of yourself.

UI Solutions and Apparatus

The Dopod P800 was the device used in the experiment. Figure 3.4 (sections a and b) shows the full-screen and the three-box design for Chinese HWR, respectively. The full-screen design just allows discrete character input, although participants can write characters anywhere on the display. Once a character was written, users had to wait for a timeout (500 ms) until recognition was initiated and then write the next character. In the full-screen design, the top six matching characters would be listed at the bottom of the screen, for users to choose the target. The three-box design enabled continuous writing in that users could write characters in alternate boxes without waiting for a timeout. Although continuous Chinese HWR is supported in the three-box design, the participants were instructed to enter characters naturally in their own way without an obligation to use continuous input all the time. I believed that in this way, I could observe user behaviors and effects of the UI styles on user performance that are similar to those in real contexts. Other function keys were almost equally provided in both designs, except that the frames listing results were in different places and the number of characters listed differed (six for the full-screen design and nine for the three-box design).



a) The three-box design   b) The full-screen design

Figure 3.4. The full-screen design and the three-box design for discrete Chinese HWR.

Procedure

The experiment was conducted in a lab environment with a coordinator and a participant present. The participant was first briefed about the objective of the study, which was to compare different interaction designs rather than to evaluate the capability of the participants. Then a practice session followed, in which the participants could rehearse the test tasks. Then all necessary features and operations were explained, to make sure that everyone started with the same level of understanding. All participants were instructed to enter the text presented, as *quickly* and *accurately* as they could during the data collection session. However, users were not obliged to write continuously with the three-box design but asked to write in their own way. During the practice and test sessions, all participants wrote characters with their dominant hand and held the device in the other hand. The experiment sessions were recorded with a video camera for further analysis. Small gifts were presented to all participants afterward.

3.5.3 Results and Discussion

I counted all deletion actions during the text entry process and divided the resulting figure by the number of characters to calculate error rates. Here the deletion actions usually took place when the participants found that they had entered a Chinese character wrongly. The participants had to click the Delete key to cancel the writing of a Chinese character. For Chinese HWR, users cannot revise a handwritten character on the stroke level or radical level. Figure 3.5 shows the results for error rates. A single punctuation mark was counted as one character because these were entered with HWR as well. ANOVA tests showed that none of the effects were significant (user type: $F_{1,18}$ = 2.000, ns; UI style: $F_{1,18}$ = 0.097, ns; interaction between user type and UI style: $F_{1,18}$ = 2.224, ns).

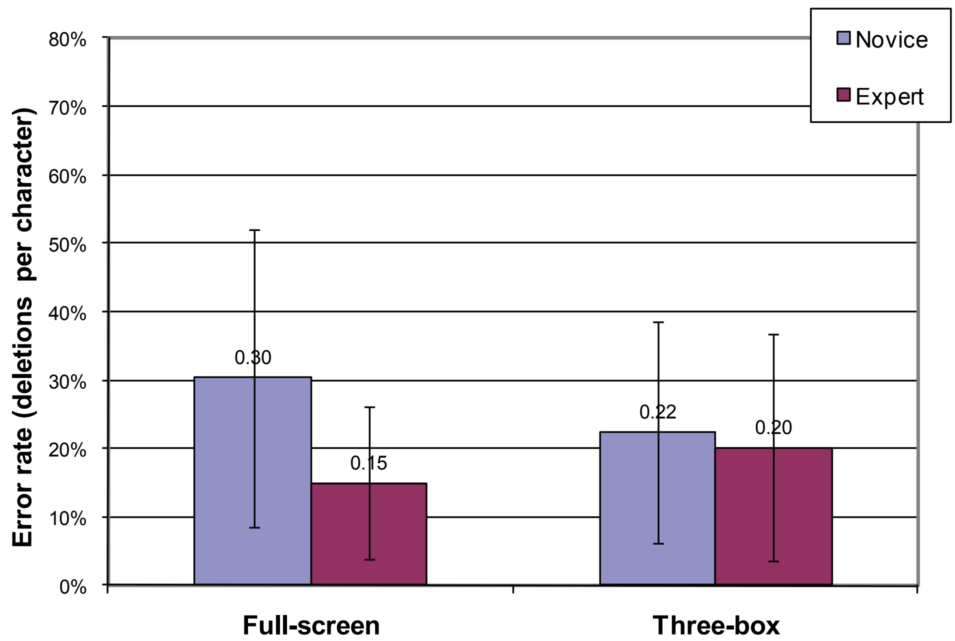


Figure 3.5. Average error rate results.

The five punctuation marks were the key sources of errors. The errors that occurred with punctuation with the full-screen and three-box UI accounted for 24.97% and 27.44% of the total number of errors, respectively. However, the five punctuation marks represented only 15.63% of all characters.

Figure 3.6 shows the results for text entry rate in CCPM. The average text entry rates of expert users were 19.33 CCPM with the full-screen UI and 18.26 CCPM with the three-box UI. For novice users, the average text entry rates were, correspondingly, 14.84 CCPM with the full-screen UI and 14.07 CCPM with the three-box UI. ANOVA analysis indicated that the main effect of user type is significant ($F_{1,18}$ = 10.532, p < .05). Expert users were faster than novice users with both UIs for Chinese HWR. None of the other effects are significant (UI style: $F_{1,18}$ = 1.105, ns; user type × UI style: $F_{1,18}$ = 0.029, ns).

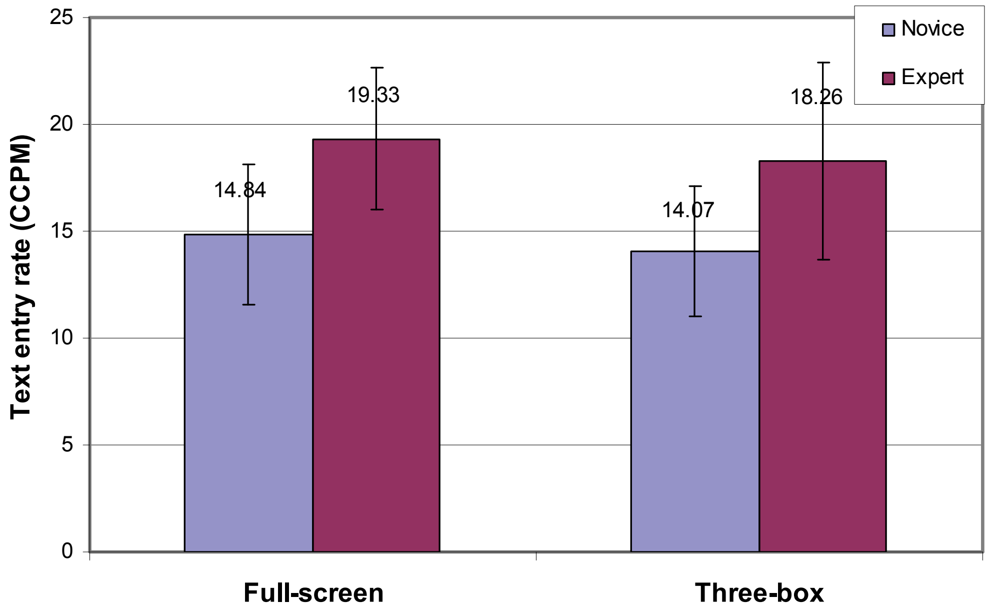


Figure 3.6. Average text entry rates.

I assumed that users could achieve better text entry rates with the three-box solution since it avoided the timeouts; however, this was not the case for either novice or expert users. To understand the reasons behind this, I analyzed users' behavior while they entered text with the three-box UI. I defined any two sequential characters that were written continuously without seeing of the recognition result for the previous one as a case of continuous writing. Then I counted all continuous writing cases for each participant when he or she completed the task with the three-box design and calculated the continuous writing rate by dividing the continuous writing count by 31 (the number of characters entered minus one). Figure 3.7 shows the average usage rate of continuous writing with the three-box design for both experienced and novice users.

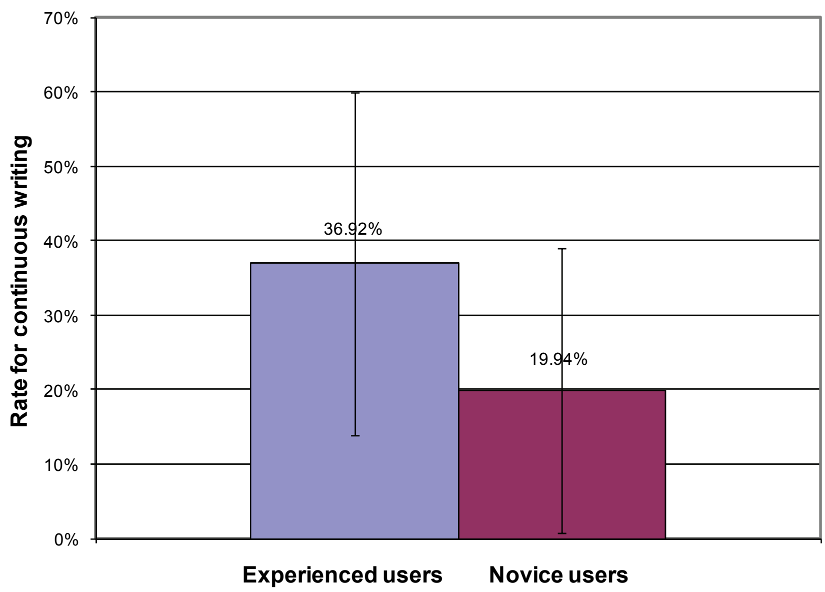


Figure 3.7. The average rate of continuous input.

The results indicated that no matter which group the users belonged to, they did not fully make use of the continuous writing supported by the three-box design, which resulted in the non-significance of the difference in user performance between the full-screen and the three-box design.

Further non-parametric analysis (Mann-Whitney test) of the data indicated that experienced users are more inclined to use continuous writing than novice users are (U = 71.5, p = .049). Observation also showed that there were two types of user strategies when the participants entered Chinese characters with the three-box UI design: one type of user seldom utilized continuous writing and always entered the second character after the previous one was correctly recognized; the other type of user employed continuous writing frequently but often stopped to correct errors noticed for previous recognitions. The former strategy is more common with novice users, and the latter is more common among experienced users. It was also observed that recognition errors noticed by users were the key factor stopping continuous writing with the three-box UI design.

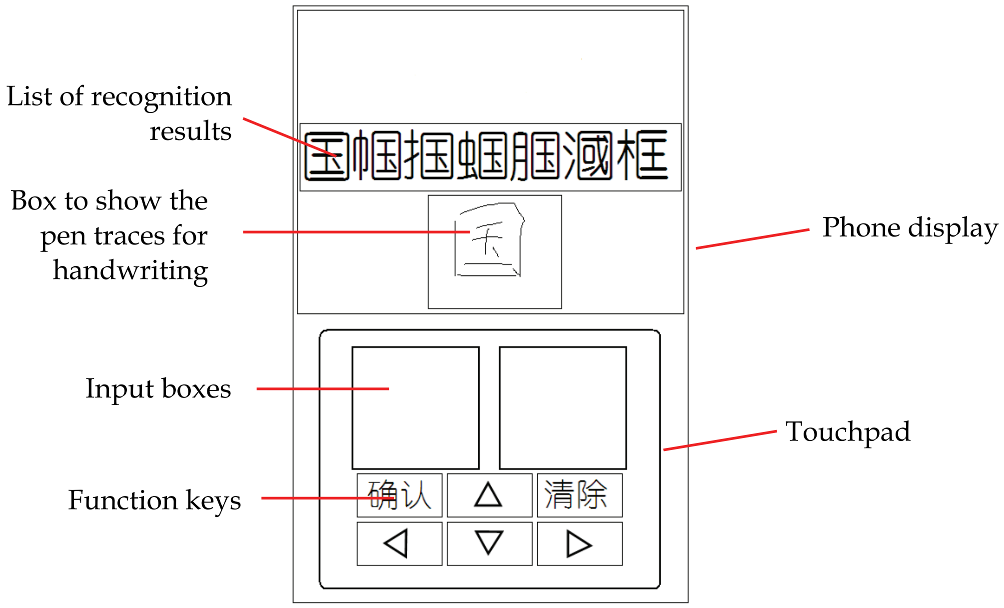
I also observed that, despite the fact that I offered three input boxes, the third box, an input box placed at the far right, was seldom used, throughout the test. This may have had something to do with the ergonomic setup of the handheld devices. When participants used the device, they often placed the bottom of their writing hand on a surface of the device. This contact point between the device and the hand became an axis as they wrote the characters. While it stabilized the writing hand against the device, it also restricted the angle at which the wrist could move comfortably as they wrote. The further they went towards the right, the more uncomfortable it was to continue writing. Moreover, participants unanimously complained that each input box was too small to write in. Considering all the results and conditions, I concluded that it is more important to provide a sufficiently large input box than to squeeze in multiple small input boxes for continuous writing.

3.6 Evaluation of the Touchpad-based Chinese HWR UI

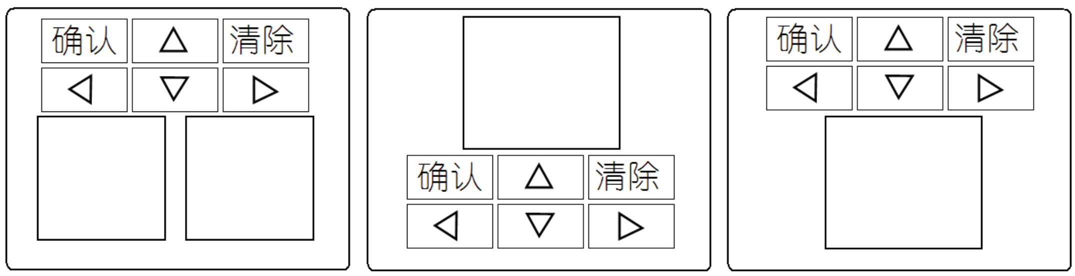
3.6.1 The Initial Designs

I designed the initial user interface layout for the touchpad-based Chinese HWR on the basis of the previous two rounds of studies. The user interface has two parts: UI layouts for the phone display and the touchpad. Figure 3.8 (pane a) shows the UI design. The phone display is divided into three parts: the top area for user-entered text, the middle area with a frame for displaying the seven most probable recognition results, and the box area at the bottom of the phone display for displaying pen traces.

Figure 3.8 (bottom area in a, b, c, and d) also shows four UI layouts for the touchpad. All layouts include one or two input boxes, four-way navigation keys, a key to clear the display, and an "OK" key. The "OK" key and the Clear key were labeled with proper Chinese phrases (OK: 确认, Clear: 清除). Here, the users must write Chinese characters in input boxes. When users write a character, the stroke traces are displayed on the phone display and audio feedback, a short tick sound, is given for writing of each stroke. After the recognition results are displayed, the user can press the navigation keys to highlight the target and click the "OK" key to select it. The left and right keys were to move the "highlighting" in the horizontal list and the up and down keys were to move to the next page of recognition results. Users can also delete incorrect characters by clicking the Clear key. The differences among the four UI layouts are in the numbers of input boxes (either one or two) and the relative positions of the input boxes and the other function keys. Solutions 1 and 2 have two input boxes each, and solutions 3 and 4 have only one input box. In solutions 1 and 3, the input boxes are above the other function keys. In solutions 2 and 4, they are below the other function keys.



a) UI for the phone display (the upper half) and UI design 1 for the touchpad (the bottom half): two smaller boxes and function keys below



b) UI design 2 for the touchpad: two smaller boxes and function keys above

c) UI design 3 for the touchpad: one bigger box and function keys underneath

d) UI design 4 for the touchpad: one bigger box and function keys above

Figure 3.8. The initial UI designs for touchpad Chinese HWR on mobile phones.

All input boxes are in the form of a square. For the two-box UI, the width and height of the boxes is 1.25 cm, because the active size of the touchpad is only 2.55 × 2.55 cm, while for the one-box UI, the height and width are 1.5 cm.

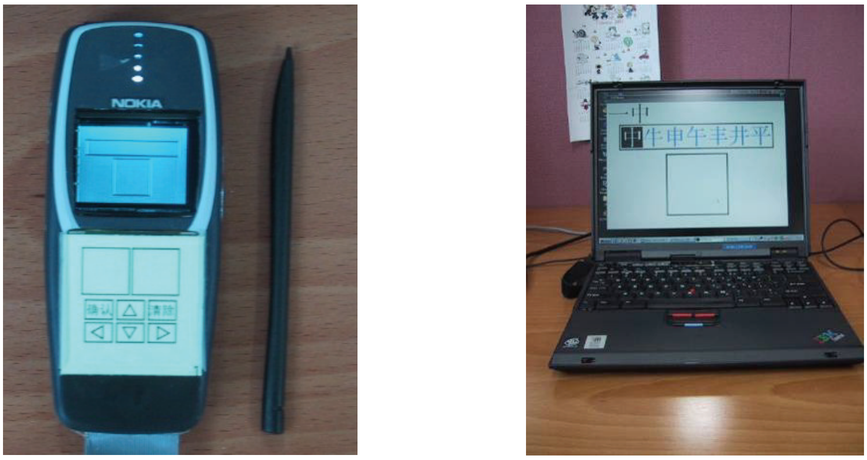
3.6.2 Objectives of the Study

Via the series of previous studies, I had collected a great deal of useful information on interaction designs for Chinese HWR. However, since the in-depth studies reported on in the previous sections were comparatively independent, focusing on different topics, a further study was still needed to check whether they can work well together in the user's mind. The objective of the prototype-based evaluation was to see whether the concept and the designs were satisfactory for users and acceptable. Moreover, the results would also provide more feedback for the design in the next round. This study was part of the E3 module in the UCD paradigm shown in Figure 3.2.

3.6.3 Method

Participants

Eight participants, four of them male and the other four female, took part in the study. Their ages ranged from 24 to 29 years, with an average of 27.0 years (*SD* = 1.85). All were right-handed. Two of them were experienced users of Chinese HWR on small touchscreen devices.



a) The terminal with a touchpad

b) The laptop running the software

Figure 3.9. The prototype used in this study: the phone prototype shown in the left figure was the input and output device and users held it to complete the experimental tasks; the laptop computer was connected with the phone prototype and the Chinese HWR engine was running on it.

Apparatus

A prototype was built for the study. The prototype consisted of a phone terminal made from a Nokia 3310 (see Figure 3.9a) and a laptop computer (see Figure 3.9b). The iPAQ stylus was used in this study (Figure 3.9a). The phone terminal served as an input and output device that was held by participants to complete the assigned tasks in this experiment. A touchpad directly replaced the keypad of the phone terminal. The physical size of the touchpad was 3.5 × 3.5 cm, and the active area was smaller, with a size of 2.55 × 2.55 cm. The phone terminal was connected with the laptop computer (shown in Figure 3.9b), on which the Chinese HWR engine was running. The UI on the phone display was also synchronized with a UI for the laptop computer. During the study, paper covers were used to change the UI layouts of the touchpad on the phone terminal.

Tasks

The participants needed to complete two tasks in the study. One was a text copying task in which the participants were asked to enter messages presented to them literally in printed form. The task helped us to focus on the interaction process and specify potential problems in it. The other was a message composition task. The participants were presented with a scenario and asked to compose a message and enter it. The task was more similar to a real usage scenario, which could help us identify potential problems in a relatively real context. While entering text with the prototype, the participants were asked to hold the device in one hand and write with the other hand, which was regarded as a common way of using it. After the participants completed the tasks, they were asked to give subjective scores to the different UI options, with a five-point Likert scale. The higher the score for a UI option, the more people felt satisfied with it.

Design and Procedure

The study was a within-subject design. All participants needed to complete the two text entry tasks with all four UI layouts. Testing orders for the UI layouts were counterbalanced with the Latin-square technique.

The study was conducted in a lab environment, where a participant and a coordinator together completed the whole process. The participant's profile was collected, and the research goals were outlined before each evaluation. Participants could ask questions if they had any. After that, the think-aloud protocol was introduced and rehearsed by the participant. Then the participants were instructed to practice with the four UI solutions to such an extent that they thought they could start with the evaluation tasks. After the practice, they started to complete the tasks. The testing processes were recorded in full with a video camera for analysis later. After the participants completed the tasks, they were instructed to give subjective scores, with a five-point Likert scale, to the UI options for the touchpad-based Chinese HWR. The subjective scores can help to show people's relative preference for the various UI options – for example, between one-box and two-box designs and between the layouts where keys were placed above or below the writing box(es). Finally, a small gift was presented to all participants after the study.

3.6.4 Results and Discussion

All participants accepted the concept very well. Although six of them misunderstood the prototype as a touchscreen product and intended to write characters in the box for showing pen traces on the display, all participants explicitly expressed that they liked the concept and thought that it was innovative. All participants successfully completed the text entry tasks with the prototype.

The visual and audio feedback and the timeout of 0.5 s were well accepted by the participants. Although the pen traces were not in exactly the same size as the written characters, no participants complained about this. I observed that the participants referred to the pen traces and adjusted their writing behavior to achieve better accuracy. The pen traces needed to be accurately displayed in real time. Inaccurate and delayed pen traces were misleading for participants. For example, it was observed that when there was a missing stroke for a pen trace, participants set out to add an extra stroke, which caused more errors.

Different types of characters (letters, numbers, and punctuation) should also be supported. Easy switching between the different character modes should also be enabled. In this study, I observed that some punctuation can seldom be recognized correctly; accordingly, participants wanted to switch to soft keyboards to enter these, but, unfortunately, the prototype did not support that function. This observation implies that our touch UI should probably support users entering punctuation with virtual keyboards. The left and right navigation keys were used often, while the up and down navigation keys were seldom used: participants seemed to prefer to write characters over again rather than go to the next page of recognition options for the target character. This was easy to understand, since user expectations of the chances of finding the target character on the other pages when it was not on the first page were fairly low. Even if merely the first recognition result was not the target character, the participants were still inclined to rewrite the character rather than select it from the list of recognition results. In this context, presenting more recognition options would not help users to access the target character better. Therefore, presentation of one page of recognition options – in our case, seven characters – seems to be enough for users.

The two boxes were not that useful when the editing functions were not optimized for error correction. The two boxes enabled the participants to write characters continuously. However, when an error was found with previous recognitions, extra actions were always needed for the error correction, including moving the cursor to the problem character, clearing it, and rewriting. Moreover, these actions were not easy with a touchpad-based UI, because the users could not operate directly on the text entered as they can with a touchscreen. Figure 3.10 shows the participants' subjective scores for the one-box and two-box designs, respectively. The results slightly favored the one-box UI: five participants gave a score of 4 to the one-box UI, while four participants gave a 4 or higher to the two-box UI; two participants would not accept the two-box UI, while only one would not accept the one-box UI. But Kolmogorov-Smirnov tests indicated that there was no significant difference in user preference for the two solutions (p = .67).

Figure 3.11 shows the subjective scores for the two UI layouts in which the navigation keys were either above or below the writing box(es). More participants preferred having the navigation keys under the input box. But the Kolmogorov-Smirnov test showed there to be no significant difference in the subjective scores between the two options (p = 1). This was probably because participants wanted to write in the area of the touchpad above so that their writing hand could get more support from the device.

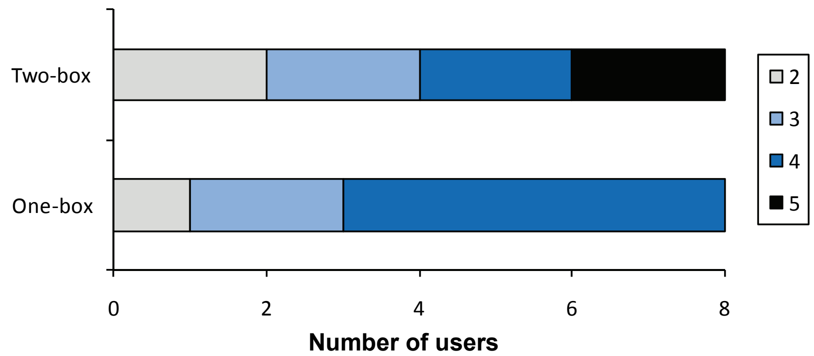


Figure 3.10. The subjective scores for the one-box and two-box designs.

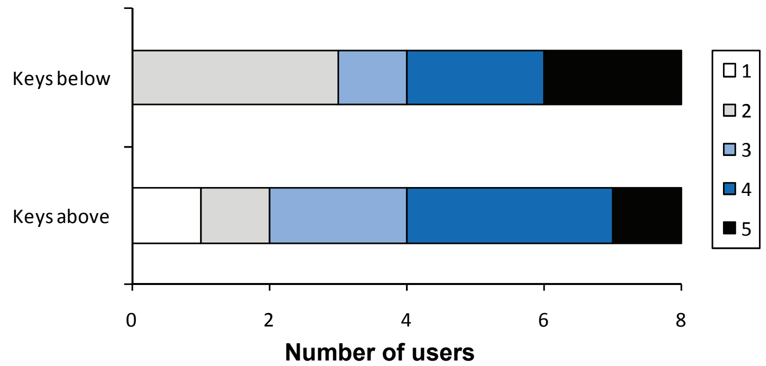


Figure 3.11. The subjective scores concerning the positions of the navigation keys.

Though user performance was not the focus of the study, I analyzed the text entry rates of the participants. The results for text entry rates varied, depending on the participants, with a range of about 6 CCPM to 12 CCPM. Further statistical analysis or comparisons were not conducted, because users were not instructed to enter text as quickly and accurately as they could in this study. Furthermore, using think-aloud in this study slowed down performance. Another study we carried out, based on the final design of the touchpad Chinese HWR, showed that the average user speed was about 11.2 CCPM (Liu & Liu, 2008).

Participants expected some advanced features in the touchpad-based Chinese HWR. For example, they expected word-based predictions: once a character was entered, the system should present predictions for the next character. In this study, only one participant complained that the stylus was too slim to hold. Some participants complained about the fatigue caused by the lack of support to the writing hand, which probably could be improved a little bit with a keypad flip added and kept open.

3.7 Other Results and Discussion

3.7.1 Results from the Other User Studies

In this section, the conclusions from our other (unpublished) user studies are briefly summarized in Table 3.7 to provide design guidelines for future development of interaction in Chinese HWR UI solutions for mobile devices.

Table 3.7: Main conclusions from other studies

Topic	Main conclusions
Stylus design	1. The length of a stylus for Chinese HWR is best between 9 cm and 11 cm, with 10 cm preferred. (Takahashi et al., 2005) 2. There are two grip points when users hold a stylus. The first grip point, where the pen is held by users with the thumb and index finger together, ranges from 12 mm to 41 mm from the stylus tip. Good stylus design assists users in holding the stylus comfortably at the grip point by, for example, providing cuts in this area. The second grip point is located from 60 mm to 91 mm from the stylus tip. Obstacles (such as protrusions) in this area should be avoided.
Visual feedback	3. Pen traces, even displayed at a distance from where a character is written and in a different size, are preferred over no visual feedback for Chinese HWR.
Audio feedback	4. Users prefer a short "tick" sound of audio feedback.
Size of an input box	5. The size of an input box for Chinese HWR should not be less than 1.5 cm × 1.5 cm.
Display of recognition results	6. Users are faster in selecting a target from a list of Chinese characters that is presented horizontally as opposed to vertically.

3.7.2 Discussion of the Design Process

Besides optimized user performance, some issues such as affordability should be taken into account in the design of new UI solutions for consumers in an emerging market. Moreover, discovering and describing the problems are critical for creation of concepts. Many times, if a problem is clearly defined, the definition itself already implies the design solutions.

The design process presented also shows a solid example of how the UCD process can be applied when a concept idea is developed into concrete interaction designs. It has been mentioned by numerous studies that involving users early is critical for development of successful user interfaces (Grudin, 1991; Vredenburg, Isensee, & Righi, 2002; Kujala, 2003). In our design process, I involved users in three modules of empirical studies: benchmark studies of competitive user interfaces; comparative studies of different UI solutions for details; and, finally, prototype-based studies. The first two modules of empirical studies were conducted before or alongside the interaction design. These studies can aid designers in making correct decisions in the interaction design process. On the basis of our design process, I summarize five functions of empirical studies in interaction design processes as follows:

– to build a benchmark for the new interaction designs (S1 and S2),

– to check user acceptance of the sensitive points for a new concept (S3),

– to differentiate UI options with users (S4, S5, and S6),

– to define the necessary parameters (S7), and

– to check the overall acceptance of a specific concept and design (S8).

Empirical studies covering these five basic topics would assist designers to make correct decisions and improve user experience with new concepts and designs.

But at the same time, I noticed that the empirical study results sometimes cannot be fully utilized in the design phases, especially when the people conducting the empirical studies and the interaction designers are different. Communicating the empirical study results to interaction designers along with the implications for design may help with the issue of communication between researchers and interaction designers. Moreover, our design process also showed the necessity of design iteration based on the prototype-based interaction design.

3.8 SUMMARY

In this chapter, an affordable Chinese HWR solution based on a touchpad and its interaction design process were presented. Via an analysis of the design space defined by immediate usability and affordability, I found that there was a lack of affordable Chinese text entry solutions that accommodate more users. We therefore proposed a touchpad-based Chinese HWR solution that is potentially good in its immediate usability. Three modules of user studies were conducted to develop the concept to the stage of UI designs. I chose three representative studies from each module and presented their details. The three studies serve as examples of how different design and research questions were addressed through involvement of users. The final prototype-based evaluation showed that both the concept and the interaction designs were well accepted by users. Results from other studies were also presented, to assist in future interaction design related to Chinese HWR for mobile devices. All efforts presented in this chapter led to the launch of the Nokia 6108 in China.

4 A Rotator Pinyin Input Solution

4.1 INTRODUCTION

In this chapter, I present the design and evaluation of Chinese pinyin text entry using as the input device a rotator, also called a wheel in some other studies (Proschowsky, Schultz, & Jacobsen, 2006). A rotator is a ring or round device that can be rotated in both clockwise and counterclockwise direction. The rotation is mapped to operations on displayed objects. The rotator enables two operations: rotating and clicking. The corresponding operations on displayed objects are *scrolling* and *selection*. The rotator has been implemented as a key input and navigation device in some mobile products, including the Nokia 7280 and Dopod P800W. Figure 4.1 shows the rotator on the Nokia 7280 as the main input device.

Figure 4.1. The rotator of the Nokia 7280.

To enter text in languages that use the Roman alphabet, users can first scroll in a list of characters displayed on the screen by rotating the device and then select the target character by clicking it. Clicking can be implemented with the rotator itself (for example, the center of the rotator in Figure 4.1 or the edge of it can be pressed) or by means of a separate dedicated key.

Several past studies have examined selection-based text entry solutions that use screen real estate to display the selectable objects. Various devices have been utilized in those studies. MacKenzie (2002b) proposed a three-key (left and right arrow keys and a Select key) date stamp method (see description below) and found its input speed to be about 9–10 words per minute with novice users. MacKenzie (2009) explored design and user performance with a one-key text entry method. It was found that text entry rates reached 5.11 WPM with a 99% accuracy rate or 7.03 WPM for error-free text. Tarasewich (2003) suggested a similar method but used the thumbwheel at the top of the left-hand side of the Sony CLIÉ PEG-S320. Wobbrock, Myers, and Aung (2004) studied the use of a joystick with the EdgeWrite method and compared its user performance with two selection-based methods. Proschowsky et al. (2006) designed a text entry method called TUP with touch-sensitive wheels and compared its user performance with that for a date stamp method with a rotator. They found that the input speed with TUP was about 6–7 WPM, which was about 30% higher than that for the date stamp method with a wheel. On the other hand, the subjective evaluation results indicated that the date stamp method was easier to understand and learn. In the selection-based text entry solutions described above, a common layout for characters was the so-called date stamp method, where all characters from "a" to "z" are listed either in alphabetical order or in a changing optimal character layout (Bellman & MacKenzie, 1998; MacKenzie, 2002). In our initial design for Chinese pinyin input with a rotator, I also used the date stamp layout.

A rotator is not a device enabling direct input as a keyboard is; therefore, performance of the rotator partly relies on objects on the display. However, the feasibility and potential for it as a good device for text entry and navigation was worth exploring. In particular, because the rotator seems good for navigation tasks and the selection of target pinyin marks and selection of Chinese characters from options provided by the system are necessary steps in the pinyin text entry solutions, it might work better in connection with Chinese pinyin input than with Roman languages. Wang, Zhai, and Su (2001) conducted an anatomical study of the Chinese pinyin text entry process with the QWERTY keyboard for the PC and found that the character selection task takes about 52% of the time in the full process. My study on pinyin character input based on the 12-key keypad (see Chapter 6) indicated that the selection process on mobile phones requires more than 65% of the time for the whole input process. Therefore, in this chapter, I explore the design possibilities with a rotator for entering Chinese text.

The rest of the chapter is organized as follows: I first describe the initial design of the pinyin input method with a rotator (called "rotator pinyin input" below, for short). I also present the findings of a quick user evaluation. Second, I describe two variations of the initial design, expected to yield improved usability, and their design principles. Third, I present an empirical study to compare the three designs in terms of performance and subjective preference. Finally, I discuss the findings, present the conclusions, and outline avenues for future work.

4.2 The Initial Design and Quick User Evaluation

The initial design was a combination of a rotator and the alphabetical layout of characters in the date stamp method. Figure 4.2 shows the user interface for rotator pinyin input. The rotator can be turned in either direction and is also clickable. To start text entry with the rotator, users need to click the rotator (or the pen key on Symbian S60 phones – for example, the Nokia 3650) to activate the input frames on the phone display. There are three input frames on the display:

- The letter frame, where the 26 Roman letters, some punctuation marks (comma, period, question mark, and exclamation mark), a space for focus change, and input modes (pinyin, stroke, English, number, and symbol) are listed.
- The pinyin frame, where the Roman letters entered to form a pinyin mark are listed.
- The Chinese character frame, where all matching Chinese characters are presented as options.

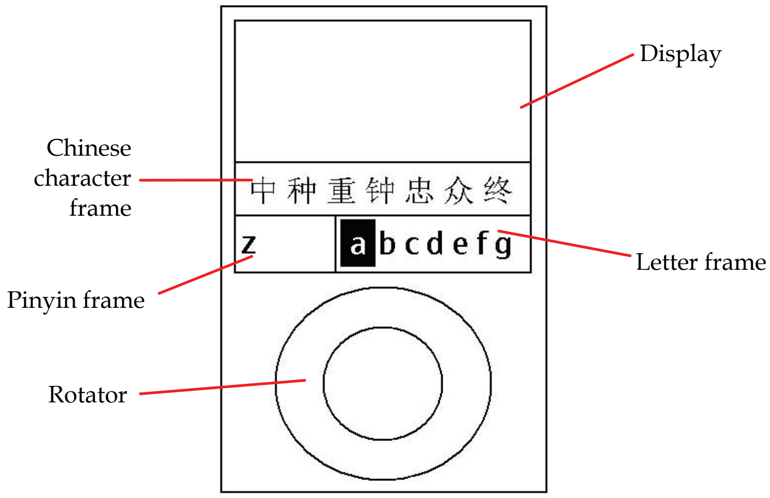


Figure 4.2. The user interface for rotator pinyin input.

In the initial design, users scroll in the letter frame with the rotator and select the target letter by clicking it. The selected letters are displayed in the pinyin frame in real time. After a pinyin mark is complete, users can select the space key in the letter frame to move the display focus from the letter frame to the Chinese character frame so that they can scroll in the latter to select the target character. To switch the focus back and make corrections to the pinyin marks entered (if necessary), users can click the Clear key. To input text in other than Chinese characters, such as characters in the Latin alphabet, punctuation marks, Arabic numerals, or similar symbols, users just need to scroll in the letter frame and select the correct input mode.

The initial design also integrates predictive input on both the pinyin level and the character level to make input faster. After users enter a letter of a pinyin mark, the input engine will predict and list possible subsequent letters in the letter frame automatically. On character level, once users enter a Chinese character, the input engine predicts the next possible character on the basis of the phrase corpus embedded in it. All predicted Chinese characters are presented in the Chinese character frame, and users can select the target one right away. The initial design for rotator pinyin input copied the user interface for rotator English input in many respects.

I conducted a quick user evaluation of the initial design to check whether the user interface was easy to learn. I found that almost no users could ascertain how to switch the focus between the letter frame and the character frame. Although some users could discover this, the means for the focus change dramatically hindered the input process and made it less smooth. Improving the method for focus change can enhance the usability of the rotator pinyin method.

I also observed that users did not realize that the rotator was clickable, but once they knew it, they accepted it well. Another common problem observed was that a click of the rotator sometimes did not result in selection of the item in focus; instead, focus had moved to the following item at the time of clicking. This mix of clicking and rotation brought to the fore the question of whether the rotator is an appropriate device for a selection task. It might be safer to map the selection function to another key. These considerations led us to the new designs described in the next section.

4.3 THE NEW DESIGNS

The key motivation for the new design ideas was that the focus change was too difficult in the initial design in terms of both discoverability and input efficiency. Would it be possible to change focus automatically between the letter frame and the Chinese character frame? As described in Section 2.3 of this thesis, most pinyin marks consist of two parts: a consonant and a vowel. Moreover, pinyin marks always end with a vowel. Hence, another natural solution is to present consonants and vowels in two lists and let the users complete a pinyin mark by simply combining the two selections. What is more, focus can automatically change from the letter frame to the Chinese character frame once a vowel is entered. I will call such a design the consonant-plus-vowel solution.

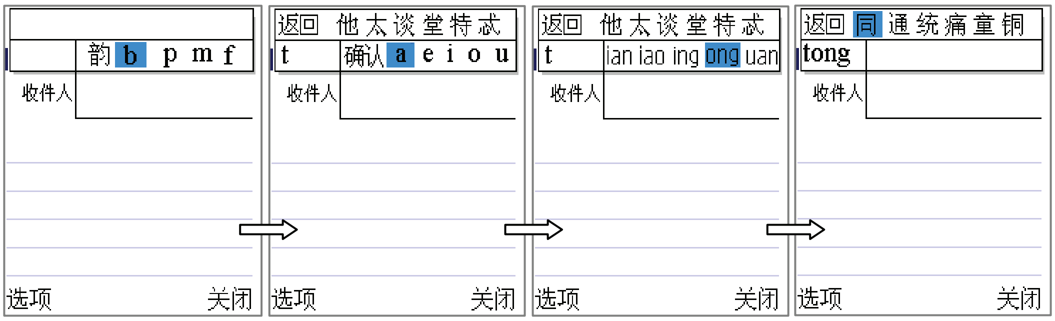
The new design may also bring extra benefits by improving input performance. Wang, Zhai, and Su (2001) mentioned that on average, each Chinese character's pinyin has 4.2 Roman characters. The result means that, on average, with the initial design users need to rotate and select in the letter frame more than *four* times to complete a pinyin mark. However, the result might not be applicable to Chinese text messages. As discussed in Section 2.2 of this thesis, the average number of Roman characters per pinyin mark is considerably smaller for Chinese text messages: only 3.24. Moreover, if one takes into account the frequencies of the Chinese characters, the figure drops to 2.88. Nevertheless, with the new designs users just need to rotate and select *twice* to complete a pinyin mark: once for its consonant and the other for its vowel. The decrease in navigation and selection time could help to improve user performance. On the other hand, the consonant-and-vowel solution has a longer vowel list, about twice as long (in screen size) as the normal Roman alphabet. All consonants and vowels of pinyin were shown in two different orders (see Figures 4.4 and 4.5). An empirical study is needed to verify the potential improvement brought by the new design.

In the new designs, the rotator is still clickable for selection. There are still only two frames (letter frame and Chinese character frame) on the display, but the content of the letter frame alternates between phonetic consonants and vowels, depending on the phase in creation of the pinyin mark.

I provided two ways to switch focus between the letter frame and the Chinese character frame. One is the automatic focus switch. Once users input a vowel to complete a pinyin mark, the focus is automatically moved from the letter frame to the Chinese character frame. Similarly, once users select the desired Chinese character, the focus is automatically switched from the Chinese character frame to the letter frame for entry of the next character. Another way is to select the functional items in the frames. To change focus from the letter frame to the Chinese character frame, users can select the "confirm" ("确认") item in the letter frame. To switch the focus from the Chinese character frame to the letter frame, users can select the "go back" ("返回") item at the very start of the Chinese character frame.

For those pinyin marks consisting of a single vowel, I added a function item as "vowel" ("韵") in the letter frame, and users can directly go to vowels by selecting it. The frequently used punctuation marks (comma, period, question mark, and exclamation mark) and the input modes (pinyin, stroke, English, number, and special characters) are also listed together with the consonants in the letter frame.

Figure 4.3 illustrates the input process step by step. In the first stage, shown on the left, the user begins entering the pinyin mark. The letter frame shows the phonetic consonants in phonetic order. The user scrolls to the right and eventually selects the letter "t." Then the letter frame automatically changes to display the phonetic vowels, in order of increasing length. In the next stage, the user has scrolled in the vowel list far enough to be able to select the ending ("ong") of the pinyin mark ("tong"). After selection (on the right), focus switches to the Chinese character frame, where the user can again scroll to select the desired character.



1) home view showing the consonants in the phonetic order

2) after the consonant of "t" is selected, all vowels are shown according to their length

3) users can rotate in the list of vowels to the target "ong"

4) users select "ong" and the highlight automatically moves to the first matching Chinese character

Figure 4.3. Text entry, step by step.

Predictive input on both pinyin mark and Chinese character level is applicable in the new design. In many pinyin marks, some vowels can never come after some consonants. Also, predictive input at the pinyin mark level can increase the input speed of the new design because the vowel list would not be that long. However, for implementation reasons, predictive input on the pinyin mark level was not included in the versions that were tested empirically. Predictive input on character level works such that after a character is entered, the most likely characters that could form a phrase in combination with it would be presented for users to select. In that way, users would not need to input pinyin marks for the later characters but just select them. However, the predictive input at character level brings an obstacle to fully automatic focus-switching. For example, if users could not find a target in the predicted character list, they would have to switch the focus manually from the Chinese character frame to the letter frame to enter it. Therefore, two methods for a focus switch are provided for the new design.

There are, in essence, two ways to present the consonants in a list: alphabetical order and phonetic order. Alphabetical order lists consonants according to the order of their first letter in the Roman alphabet. Figure 4.4 (a) shows the pinyin consonants in alphabetical order. The phonetic order is the one Chinese primary-school students are taught when they start to learn pinyin and Chinese characters. Figure 4.5 (a) shows the pinyin consonants in their phonetic order. There are also two ways to list vowels: alphabetical order and length order. With listing in alphabetical order, all vowels are listed according to the order of their first letter in the Roman alphabet. Figure 4.4 (b) shows the vowels in alphabetical order. Listing in length order means the vowels are listed according to their length. For vowels of the same length, the order is decided by the alphabetical order of their first letter. Figure 4.5 (b) shows the length order. For user evaluation, I built the following two prototypes.

4.3.1 Alpha: New Design 1

In this design, both consonants and vowels are listed in their alphabetical order, as shown in Figure 4.4. The alphabetical order was assumed to be familiar to many users, and the consistency between the consonant list and the vowel list should help users to understand the design.

b c ch d f g h j k l m n p q r s sh t w x y z zh

a) Consonants in alphabetical order.

a ai an ang ao e ei en eng er i ia ian iang iao ie in ing iong iu o ong ou u ua uai uan uang ue ui un uo v

b) Vowels in alphabetical order.

Figure 4.4. Consonants and vowels in alphabetical order.

4.3.2 Beta: New Design 2

In the Beta solution, the consonants are listed in phonetic order and the vowels are listed according to their length (Figure 4.5). I implemented this design for two reasons: 1) I assumed the phonetic order could remind users about the consonants and vowels of pinyin marks so as to help them understand the design, and 2) I assumed the length order of vowels could provide a good visual cue on where to find the needed vowel.

b p m f d t n l g k h j q x zh ch sh r z c s y w

a) Consonants in phonetic order

a e i o u v ai an ao ei en er ia ie in iu ou ua ui un uo ang eng ian iao ing ong uai uan iang iong uang

b) Vowels ordered by length

Figure 4.5. Consonants in phonetic order and vowels ordered by length.

4.4 User Evaluation

4.4.1 Objectives

Both the Alpha and the Beta solution had some performance advantages over the initial design for the following reasons:

– On average, users need to scroll and select approximately four (2.88 Roman characters plus a confirmation) times to complete a pinyin mark with the initial design, but with the new designs, the number is 2 for a consonant and then a vowel.

– Focus switches are automatic, which also decreases the number of scrolling and selection actions.

However, there are also factors that might offset the advantages brought by the two factors mentioned above. First, predictive input on pinyin level was missing for both Alpha and Beta while predictive input was fully implemented in the initial design. Second, the vowel list had to be longer (in screen size) than the list in the initial design, because of the absence of predictive input on pinyin level. On the other hand, the dimensions of a character on the screen are not really important – the length of the list that needs to be scrolled is, because the number of items directly corresponds to the motor movement of the rotator. I will discuss the effect of missing prediction capability after presenting the test results.

To verify the assumptions with facts, I designed an empirical study whose objectives were:

– to check whether the automatic switching of focus helps;

– to collect further usability problems of the initial design, Alpha, and Beta as Chinese rotator pinyin input solutions;

– to compare user performance with the three designs; and

– to collect information on user preference and comments on the three design solutions.

4.4.2 Method

Participants

Twelve volunteers, seven of them males and five females, ranging in age from 21 to 35, participated in the user evaluation. Five participants were familiar with the Symbian S60 user interface. All were right-handed and daily users of the keypad pinyin input method (T9) on their mobile phone. All were first-time users of the rotator.

Apparatus

Figure 4.6 shows the prototype built on the Nokia 3650 and the testing environment. Participants just needed to hold the phone prototype to complete the evaluation sessions. The phone prototype was connected to a laptop computer via a sharp box. All inputs via the rotator from the phone were transferred to the computer via the box. The computer, with the input engine, fed back the proper output to both the computer display and the phone prototype.

I recorded the test sessions with two video recorders. One camera was attached to the phone prototype to record the phone display and the user interactions. The other video camera was used to record the user.

Figure 4.6. The phone prototype and the test environment.

Tasks and Materials

The participants completed two text entry tasks with all three designs. The first task was for diagnostic purposes, and the second one was to collect user performance data. In the second task, the participants were asked to enter the messages exactly as they were presented on paper, which meant that they needed to correct any errors they made in the input process. I also instructed the participants to enter the text in the second task as accurately and quickly as they could. The messages used in both tasks were really short messages that I had collected from end users in past studies (see Table 4.1). All characters in the messages belonged to the 500 most frequently used Chinese characters. For task 2, the average pinyin lengths for the two sentences are, respectively, 3.1 and 2.9 letters. The single-letter correlation (MacKenzie & Soukoreff, 2003) of the two messages in task 2 with the short message corpus presented in Section 2.2 was 0.917.

Experiment Design, Procedure, and Measures

Every participant needed to complete both tasks with all three design solutions. The order of the input methods was counterbalanced. Before the evaluation started, I introduced the objectives of the evaluation, emphasizing that the evaluation was of our design solutions, not the participants themselves. Then participants were instructed to fill in a pre-evaluation questionnaire for profile collection. After that, the test session started, in which the participants entered the messages with the three methods. After completing the tasks with each method, the participant gave a subjective evaluation with a five-point Likert-scale questionnaire. Finally, participants were asked to choose the one design that they would like to have in their own mobile phone.

When analyzing the results, I counted the number of "Clear" operations (the number of times the "C" key on the mobile phone was pressed) by each participant in the second task and computed their input speed in Chinese characters per minute. I also collected subjective scores for understandability and perceived performance of the design solutions. Understandability scores indicate the walk-up usability of the design solutions, and the perceived performance scores indicate the subjective perceptions and "feel" of system performance, which is relevant in terms of feasibility of take-up of the method in the long term.

Table 4.1: Tasks and materials

Task 1	**Message**	Please input the following message: 我乘坐 MAS 613 航班 9：48 到达。
	Pinyin	wo cheng zuo MAS 613 hang ban 9:48 dao da *(average pinyin length: 3.1)*
	Meaning	I will arrive at 9:48 on flight MAS 613.
Task 2	**Messages**	Please input the following messages: 1.明天下午两点同学聚会，你能否和我一同参加？ 2.今晚不回家吃饭了，十点左右到家。
	Pinyin	1. ming tian xia wu liang dian tong xue jv hui, ni neng fou he wo yi tong can jia? *(average pinyin length: 3.1)* 2. jin wan bu hui jia chi fan le, shi dian zuo you dao jia. *(average pinyin length: 2.9)*
	Meaning	1. There is a party for classmates at 2pm tomorrow. Could you go to it with me? 2. I cannot have dinner at home today, and I will arrive home at about 10pm.

4.4.3 Results

Observation Results

In general, the participants had no problem in understanding the initial design. All of them also quickly got the idea of "consonant plus vowel" with Alpha. But since in Beta the vowels visible on the first screen (see the second step in Figure 4.3) are the ones with only one letter, it took a while for all four participants who started the evaluation with it to get the idea of the consonant-plus-vowel approach.

Eight participants discovered that the rotator was clickable. All participants mastered it quickly after they knew this. No participant discovered that pressing the pen key can also activate the frames on the display. I informed the participants of this option, but they forgot it again during the input process.

All participants readily accepted that pressing the rotator can select an item in the frames and mastered this action. No participants had problems in changing input mode with the rotator. Some participants noticed that the list in the letter frame was a looping one, and they rotated both clockwise and counterclockwise, to enter text more rapidly.

Automatic focus switch was implemented in all designs. Some individual letters – for example, "g" – can also indicate the ending of a pinyin mark, and I applied this principle, too, in the initial design. All participants used both automatic and manual focus switching. However, I observed clear hesitation before participants rotated to the function items for focus switch and selected them: the participants seemed more attentive in such cases. Moreover, since sometimes automatic focus switch was on and sometimes it was off, depending on the input conditions, participants were observed to be slightly confused with the inconsistency.

Performance Results

I analyzed performance results based on the second user task. Table 4.2 shows the number of Clear operations in the second task from all participants for error correction.

Table 4.2: The number of Clear operations performed by each participant

User no.	Initial	Alpha	Beta	User no.	Initial	Alpha	Beta
1	2	0	4	**7**	0	0	0
2	3	4	0	**8**	2	0	2
3	2	3	2	**9**	0	4	0
4	0	2	0	**10**	2	4	2
5	0	2	2	**11**	2	1	0
6	0	2	0	**12**	5	2	0

Participants seldom made serious errors with rotator input: on the average, there were two clear operations per user for Alpha, were one and a half for the initial design, and was only one for Beta. I observed that participants sometimes passed targets and had to rotate back, but that could be corrected by navigation and did not introduce any errors or Clear operations. The ANOVA test indicated that the main effect of design solution on error frequency was not significant ($F_{2,35} = 1.36$, ns).

Table 4.3 and Figure 4.7 show results concerning input speed with all design solutions. As Figure 4.7 shows, there was no uniform pattern for all participants. Some participants (3, 7, 9, and 12) were faster with the initial design, while others achieved higher input speed with Alpha (1, 2, 4, 8, and 10) and the rest were faster with Beta (5, 6, and 11). On average, participants achieved the highest input speed with Alpha, followed by the initial design and Beta. ANOVA indicated that the difference was not significant ($F_{2,35}$ = 0.37, ns).

Table 4.3: Input speed and ANOVA test results

Variable		Design solutions			F	*p*
		Initial	Alpha	Beta		
Input speed (CCPM)	Mean	6.46	6.56	6.17	0.37	.70
	SD	1.05	1.28	1.11		

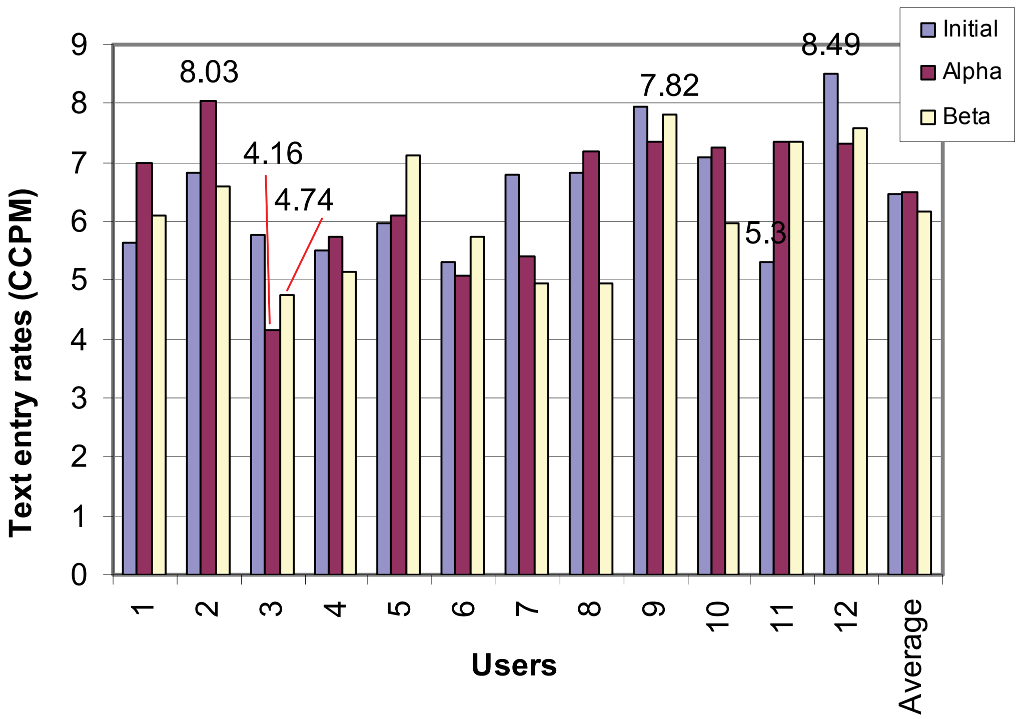


Figure 4.7. Entry speed of each user.

Results of Subjective Evaluations

Table 4.4 and figures 4.8, 4.9, and 4.10 show the results from the subjective evaluations. Participants thought the initial design was the easiest to understand, followed by Alpha and Beta. ANOVA indicated that the main effect of design solution was significant ($F_{2,35}$ = 17.64, p < .011). A further *t*-test showed that the difference between any two design solutions was significant.

Table 4.4: Subjective evaluation results for understandability and perceived performance

Variable		Design solutions			F	*p*
		Initial	Alpha	Beta		
Understand-ability	Mean	4.83	4.17	3.00	17.64	.00
	SD	0.39	0.84	0.95		
Perceived performance	Mean	2.83	3.92	3.08	4.63	.02
	SD	1.03	0.79	0.90		

As Table 4.4 shows, participants thought they could achieve the best performance with Alpha, followed by Beta and the initial design. ANOVA indicated that the main effect of design solution was significant ($F_{2,35} = 4.63$, $p < .05$), and a further *t*-test indicated that the scores for Alpha were significantly higher than those for the other two designs.

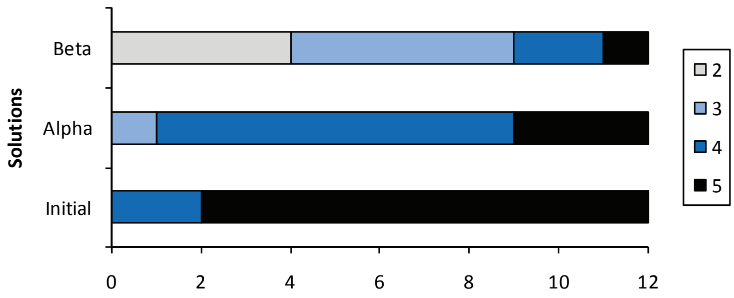


Figure 4.8. Subjective evaluation results for the understandability of all design solutions.

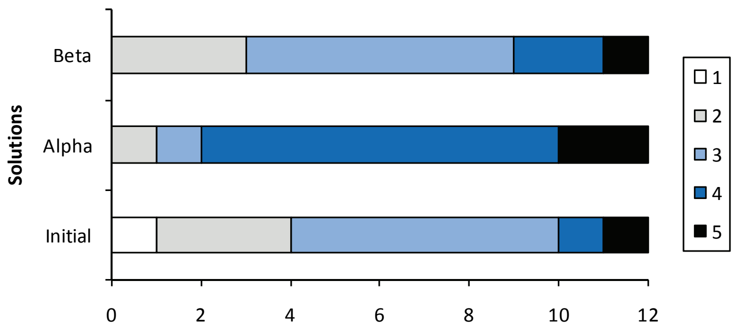


Figure 4.9. Subjective evaluation results for perceived performance.

Figure 4.10 shows which design the participants would choose for their own mobile phone. Two of the 12 participants liked both the initial design and Alpha, but not Beta. The other 10 participants chose just one as the most highly preferred solution. Seven of them preferred Alpha, one chose the initial design, and two chose Beta.

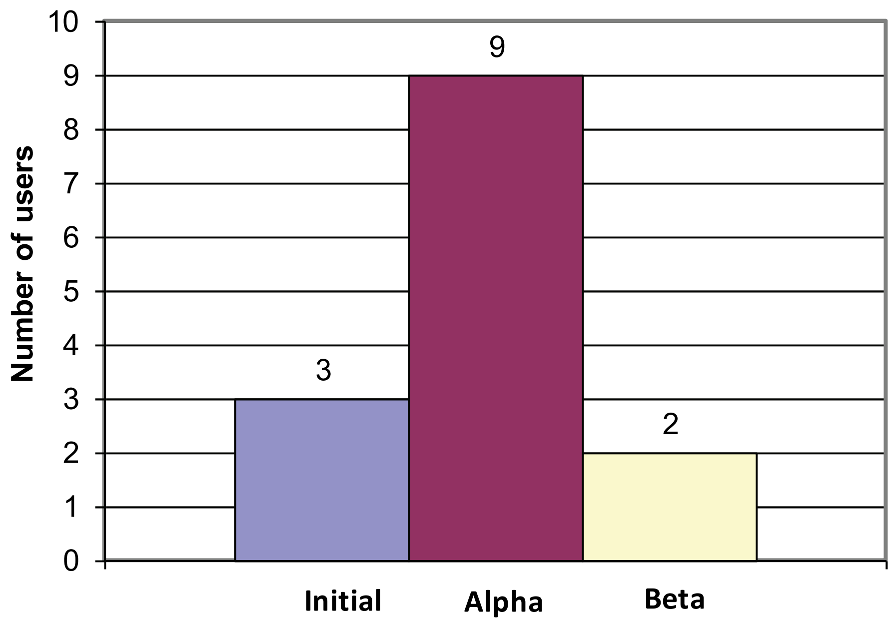


Figure 4.10. User preferences regarding the design solutions.

4.5 Discussion

Although a rotator is not a direct input device as a keyboard is, it has some advantages. First and foremost, it is very good for navigation, and it supports selection as well. This means that the rotator is sufficient in itself for some simple interaction tasks. In particular, it is a tempting choice for text entry in languages that have a larger character set than the keyboard can accommodate, and for which the production of each character therefore inherently requires navigation and selection. Moreover, that users do not need to move their fingers among different components of the input device is a potential source of improvements in operation efficiency. Finally, the rotator can be used on devices that are so small that they cannot accommodate even a normal ITU-T keypad.

Proceeding from the initial design of Chinese pinyin input with a rotator, I provided two new design solutions, based on the idea of "consonant plus vowel." The empirical research results showed that there is no significant difference among the three designs in user performance. An interesting question is how much the prediction aided in the initial design. The time per character entered consists of two parts: length of the list to be scrolled, and number of selections. Prediction does not affect the latter. Table 4.5 shows the number of each type of operations in each of the cases.

Table 4.5: Number of operations for entry of the test sentences

	Scrolling distance	Selections
Initial, without prediction	1097	119
Initial, with prediction	652	119
Alpha	979	78
Alpha, with prediction	694	78
Beta	970	78
Beta, with prediction	695	78

Thus one can see that prediction shortens the navigation by about 30% in the initial design, and in the predictive version most of the navigation time is taken by entry of the first Roman letter of the pinyin mark. In Alpha and Beta, on the other hand, the number of selection operators is only about 65% of that in the initial design. This explains well why the differences balance out, and no significant time difference was found in the test.

Table 4.5 also shows that if prediction had been implemented for Alpha and Beta, their motor performance should have been almost equal. Moreover, they do not lose much to the date stamp method in scrolling distance, and, therefore, the number of selection operations becomes the decisive factor. Here the difference is the same as without prediction – i.e., clearly favoring Alpha and Beta over the date stamp method.

The speeds published for Chinese pinyin input with the 12-key keypad vary greatly. Lin and Sears (2007) reported that the input speed of a pinyin method with the 12-key keypad is about 5.5 WPM when the participants were instructed to balance input speed and error rate as they would. Liu and Wang (2007) reported that Chinese pinyin phrasal input with the 12-key keypad can enable users to reach an average input speed of up to 34 CCPM although some other phrasal pinyin input methods can just reach an average speed of 14 CCPM. This increase in input speed may have appeared because the experiment settings were different, Chinese users are getting more and more familiar with the mobile devices and more advanced features and technologies are being developed. Therefore, I believe 6–7 CCPM for rotator-based Chinese input is already a good start for a method that does not use character-level prediction.

The two designs aim to promote usability of the rotator Chinese input by enabling automatic switching of focus. However, this did not help as much as I expected. Both automatic focus change and manual change were used with all three design solutions. In some conditions, users had to select the function items in frames manually to switch the focus. For example, if users could not find the target character in the predicted Chinese character list, they had to move the focus from the Chinese character frame to the letter frame manually. On the other hand, sometimes they did not enter the full pinyin mark when they noticed that the character needed had already appeared in the Chinese character frame. Then they would stop creating the pinyin mark and move the focus to the Chinese character frame to select the character.

New designs for focus change are still needed. Automatic switching of focus worked in some cases but not always. The inconsistency confused some users and requires more attention. I observed that the focus change possibilities provided made the input process less smooth with Alpha and Beta. Two techniques might help with this. One is to introduce dedicated keys for changing focus; using them might become automatic after a period of practice. Another possibility (only for Alpha and Beta, which use phonetic characters) is to remove the inconsistency by abandoning predictive input and ask users to input characters one by one. In such a case, the focus could always be switched automatically between the letter frame and the Chinese character frame.

Subjective evaluations are usually more sensitive than performance results, as indicated in many past studies (Darroch et al., 2005; Hornbaek, 2006). In our study, user performance results did not show improvements for Alpha and Beta in comparison to the initial design, but the subjective results showed that users believed they could perform better with Alpha than with Beta and the initial design. Many participants commented that they thought the consonant-plus-vowel approach would decrease the number of scrolling and selection actions. When comparing Alpha with Beta, many users commented that (contrary to our expectations) they were not familiar with the phonetic order of consonants in Beta, which then led to worse performance.

As the subjective results showed, the initial design was the easiest to understand for novice users. Alpha also got an average score above 4, which indicated that it, too, was easy to understand. However, Beta received a neutral score. This result was congruent with what I observed in the evaluations. In addition to the problems caused by unfamiliarity with the phonetic order of consonants, the start of the vowel list consists of only single-letter vowels, and users did not realize that the list continued beyond what was visible on the first screen.

When choosing the design they preferred overall, nine out of 12 users voted for Alpha. Users seem to value perceived performance of a design solution over immediate understandability.

4.6 Summary

I designed two new solutions, Alpha and Beta, for Chinese pinyin input with a rotator and conducted an empirical study to compare them in use by novice users. The results indicated that, although there was no significant difference among the three design solutions where user performance is concerned, users believed that they could achieve better performance with Alpha, and they chose it as their preferred input technique.

This is a promising result, because the test setup favored the traditional design. No predictive input was implemented for the new designs. Also, all test participants were novices in using the rotator as an input device; it is reasonable to expect that in a longitudinal study the advantages of the new techniques, which require a smaller number of selections, would become more pronounced.

5 A Mandarin Dictation Solution

5.1 INTRODUCTION

The text entry methods discussed above are very useful in contexts with at least one hand free. However, mobile devices are expected to be used in many different contexts, including those with no free hands – for example, when users are riding a bicycle or wearing thick gloves. Speech input provides an additional modality and solution for those contexts (Entwistle, 2003; Price & Sears, 2005; Howell, Love, & Turner, 2005; Alhonen et al., 2007). Speech recognition of natural languages is very complex and faces two general challenges: recognition of continuous speech and speaker independence (Entwistle, 2003). Although some dictation systems have been available on desktop computers for a few years (Sacher, 1998), the technologies were mostly applied for very limited functions in mobile devices, including name dialing, information retrieval, and menu navigation (Chang et al., 2002; Karpov et al., 2006; Alhonen et al., 2007). So far, only isolated speech recognition technology has been implemented for dictation of short messages on mobile devices, with which users need to enunciate word by word to the system with clear pauses between any two consecutive words (Karpov et al., 2006; Alhonen et al., 2007).

Speech recognition technologies and relevant user interfaces encounter greater challenges with Chinese (Sacher, Tng, & Loudon, 2001; Alhonen et al., 2007; Liu, Ding, & Liu, 2009). First, most Chinese characters are homophonic with many others. For example, there are only about 400 syllables in Mandarin Chinese, together corresponding to tens of thousands of Chinese characters. Therefore, even if a phonetic syllable is correctly recognized, further user interactions are still required for choosing the target Chinese character. Second, although Mandarin is the dialect with the most speakers in China, spoken Chinese falls into 10 main dialect groups. It is hard to cover potential users with a single, universal solution. Even worse, since how people speak Mandarin is affected by their native dialect, the effects of different dialects should be taken into account in Mandarin speech recognition technologies. Third, there are multiple tones for Chinese spoken languages, and most times they are semantically meaningful. For example, Mandarin has five tones. Speech recognition technology has to consider how to recognize tones for spoken Chinese and utilize them in the recognition results.

Numerous Mandarin speech recognition or dictation systems on personal computers have been built since the early 1990s, and high recognition accuracies have been reported for many of them (Gao, Zhong, & Ceng, 1995; Fu et al., 1996; Chen et al., 1997; Wang et al., 1997; Huang et al., 2000). However, studies examining human–computer interaction with Mandarin speech recognition systems are rare. Moreover, since mobile devices are used in diverse contexts, not just in offices and homes, it is worth exploring the application of speech recognition technology as a text entry solution for mobile devices.

In this chapter, human factor studies of Mandarin message dictation applications on mobile phones are presented. In the next section, I present two user studies for evaluating utility and usability of the Mandarin message dictation application with working prototypes. After that, I present a Wizard of Oz study to explore the effects of speaking style on user performance and satisfaction. Finally, the conclusions from the studies are summarized.

5.2 Evaluating the Isolated Mandarin Message Dictation Application

5.2.1 Objective

I conducted two rounds of user evaluations of the isolated Mandarin message dictation application, with a similar approach (see Figure 5.1). The main objective was to understand the utility and usability of the isolated Mandarin speech dictation application by collecting user performance data for recognition rate and text entry rate.

In both rounds of studies, I measured recognition rates with end users. Moreover, I also explored the effect of adaptation to users' voices on user performance. After the first user study, we optimized the recognition engine and made small changes in the interaction process (Alhonen et al., 2007). The second study was conducted to check whether the performance optimization of the speech recognition engine worked. In this section, the two studies are presented together, since their approaches and results are comparable.

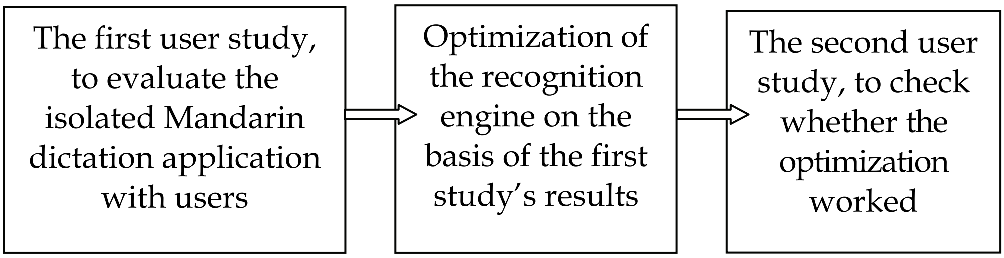


Figure 5.1. The two rounds of user studies.

5.2.2 The Isolated Mandarin Message Dictation Application

The Mandarin message dictation application works independently on mobile phones, where the recognition engine works without involving any computing ability from any other sources. The Mandarin message dictation application supports isolated dictation of Chinese characters only: users are required to leave a short pause between syllables (one syllable corresponds to one Chinese character). In the application, pressing the "OK" key can initiate a dictation process. Different indicators were designed to show stages in the dictation process. A rotating bar indicates when users can speak and the system record, whereas a rotating hourglass is shown when a syllable is under processing. The top four recognition results are given in the form of pinyin marks in a vertical frame, as Figure 5.2b shows. Also, Chinese characters that correspond to the highlighted pinyin mark are displayed horizontally at the top of the application. They are ordered from higher to lower probability on the basis of previous characters in line with a language model. The pinyin and character options would be shown to users for about a second. Users can move up/down in the selection of syllables (pinyin marks) and left/right in the selection of characters by pressing the five-way navigation key of the mobile phone. Users can also type the numeric tag in front of each character option for fast selection.

a) UI for user enrollment

b) UI for showing recognition results

c) Ability to select errors from the retained results

d) Predictive input

Figure 5.2. The screenshots for the isolated Mandarin message dictation application on a Symbian S60 phone.

Predictive input is also supported by the prototype. The application constantly predicts the most probable next character for users on the basis of the previously entered character and the language model (see the final image in Figure 5.2). This provides an option for users to bypass dictation of the next character, which often follows the previous character.

The application supports three means of error correction. The recognition options are retained in order to allow corrections later. Users can move the cursor to the incorrect character and revise it by re-selecting the target pinyin and corresponding Chinese character. Users can also delete the wrong character by pressing the "clear" key and say the syllable again. Moreover, users can switch off the dictation and correct errors with keypad input methods. The correction mechanisms may slow down the input process, but they have several advantages. First, users have full control over the recognition process: incorrect characters can be cleared immediately or afterwards, in different ways. Furthermore, the speech recognition of the next character can be predicted with the language model to select the most probable followers. Outside-vocabulary words can be handled in a convenient way: if a user cannot dictate some syllables even after several trials he or she can stop the dictation, type the character, and resume dictation, whereas in some conventional systems unknown words may corrupt the resulting sentence.

The isolated Mandarin message dictation also supports user enrollment that can adapt the system to users' voices. The enrollment can be invoked from the option menu. During enrollment, the prototype guides users through 35 phonetically rich sentences that have to be read character by character. Figure 5.2 (a) shows the screenshot for user enrollment. The green frame highlights one character at a time for users to utter. Once a syllable is detected as having been spoken, the green highlighting switches temporarily to yellow and then jumps to the next character. If a character is mispronounced, users can press the Clear key to delete the record and read it again. After all characters are uttered, the dictation application needs to be restarted, for bringing the data into use. Although the main purpose of user enrollment is to adapt the acoustic models to the speaker's voice, a side benefit is that it also assists users to learn the isolated character speaking style that is required by the application.

5.2.3 Method

Participants

Sixteen participants took part in the two rounds of user studies, with eight participants in each. However, since there were several months between the two studies, the participants in the experiments were different. Both experiments are within-subject designs. All participants were required to enter the same text messages twice, once before and once after the adaptation.

In both experiments, the native dialect of all participants was Mandarin, to ensure that the results were comparable.

Tasks and Materials

In both experiments, the participants needed to copy five text messages twice with the Mandarin SMS dictation application. The five text messages, selected from a Chinese SMS corpus, include 87 Chinese characters and 11 punctuation marks (see Table 5.1). All messages were constantly shown to the participants during the experiment. Errors in the recognition results were instructed to be corrected with the 12-key keypad-based pinyin character input method provided in Nokia E50 (see Figure 5.6). The input processes were recorded in their entirety with a video camera for data analysis.

Table 5.1: The five messages to be entered in both evaluation studies by each participant

No.	Entered messages	
1	Message	国外的饭实在是太难吃了，因此很少出去吃，多数自己做。
	Meaning	It is too hard to get used to the food overseas, so I seldom go to restaurants but normally cook myself.
2	Message	王强什么时候回家呢？
	Meaning	When will Wang Qiang (a person’s name) come back?
3	Message	爸妈不想去，谁有时间有兴趣就去吧。注意安全。
	Meaning	My parents woud not like to go for the program. If anyone else is insterested in it, please feel free to go and take care.
4	Message	完全不知道怎么表达我的谢意。
	Meaning	I don’t know how to express my appreciation.
5	Message	人总是这样，得到的东西不好好珍惜，失去了才觉得可贵。
	Meaning	People do not cherish things they already have but when they lose these things, they start to feel their value.

Apparatus

The device used in both experiments was the Nokia E50 (see Figure 5.6). There is little difference between the applications that were used in the two studies. In the first user study, when users selected a target Chinese character from the options presented, by clicking the navigation keys, the clicking sound often activated the speech recognition engine. The unnecessary recognitions interfered with the input process and decreased user performance. In the second study, I set a 500 ms timeout after each recognition so that the speech recognition would not be wrongly activated.

Procedure

Each participant took part in the studies in a quiet lab with a researcher. First, the researcher briefed the participant on the objectives. Then the participants were instructed to use the application to enter text messages and give comments on their first impression of the application. Before the data collection, the researcher explained to the participant how to use the dictation application, so that each participant had the same understanding of the input process. This was followed by a trial session wherein the participant could practice until ready to start the data collection sessions. The data collection phase included two sessions: one before the enrollment and the other after the enrollment. Finally, the participant was asked to fill in a five-point Likert-scale questionnaire to evaluate the dictation application's acceptability, utility, and so on. After the experiments, each participant received a gift.

5.2.4 Results

Recognition Rates

Figure 5.3 and Table 5.2 present the results for the 4-best recognition rate. The 4-best recognition rate refers to the rate that the target syllable is included in the top 4 recognition results. In the first study, both average 4-best recognition rates before and after enrollment were around 70%. But in the second study, both 4-best average recognition rates were above 95%. Recognition rates increased after enrollment, in both studies, but the increases were negligible.

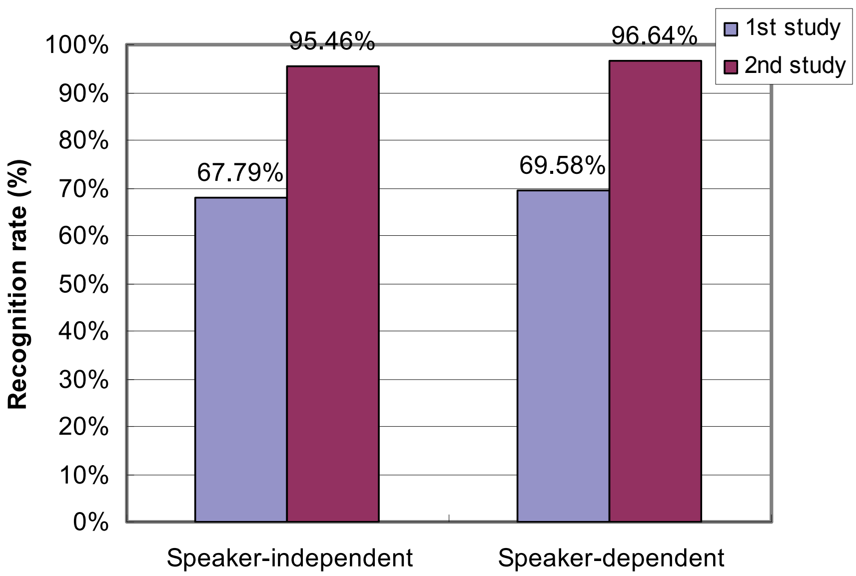


Figure 5.3. A comparison of recognition rates.

A two-way mixed-design ANOVA analysis indicated that the recognition rates in the second study were significantly higher than those in the first study ($F_{1,15} = 35.28$, $p < .001$). The main effect of enrollment is not significant ($F_{1,15} = 0.20$, ns). The interaction is not significant either ($F_{1,15} = 0.009$, ns).

Table 5.2: Recognition rates and standard deviations

Recognition rates (4- best)	**Speaker-independent**		**Speaker-dependent**	
	Mean	*SD*	Mean	*SD*
1st study	67.8%	17.5%	69.6%	1.2%
2nd study	95.5%	14.1%	96.6%	2.4%

Text Entry Rates

Figure 5.4 and Table 5.3 show the results concerning text entry rates when error corrections were also taken into account in the analysis. The average text entry rates ranged from eight to 10 CCPM. Although the 4-best recognition rates in the second study were significantly higher than those in the first study, text entry rates in the second study do not show a great increase in comparison with the first study.

A two-way mixed-design ANOVA analysis was conducted, and the results indicated that there were no significant differences between the two studies for text entry rate ($F_{1,15} = 0.0001$, ns). However, the main effect of enrollment was significant ($F_{1,15} = 8.60$, $p < .05$), and the interaction was significant ($F_{1,15} = 7.88$, $p < .05$). Further paired *t* tests were conducted to analyze the interaction effect. The results indicated that the enrollment significantly increased text entry speeds in the first study ($t = 0.036$, $p < .05$) but not in the second study ($t = 0.36$, ns).

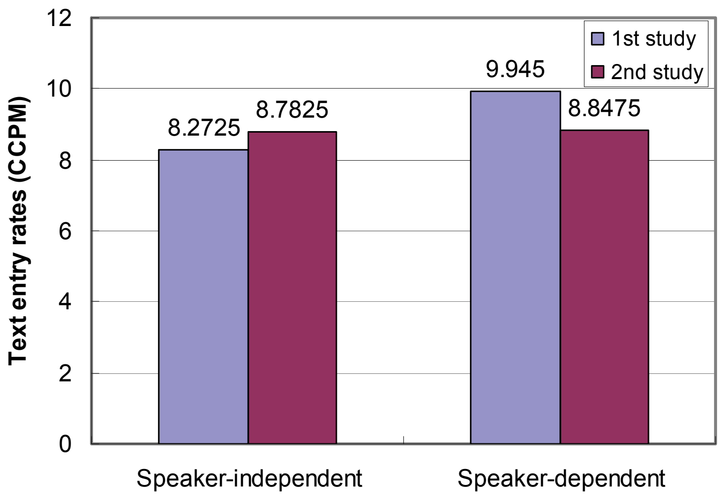


Figure 5.4. A comparison of text entry rates.

Table 5.3: Text entry rates and standard deviations

Text entry rates (CCPM)	Speaker-independent		Speaker-dependent	
	Mean	*SD*	Mean	*SD*
1st study	8.27	1.47	9.95	1.71
2nd study	8.78	1.32	8.85	1.67

Observed Results and Subjective Scores

Figure 5.5 shows the subjective scores for three aspects of the application: easy of use, perceived performance, and usefulness. The scores in the second study were slightly better than the scores in the first study. The results indicated that participants believed that speech dictation is a useful tool for them; however, they are neutral as to its usability and perceived performance.

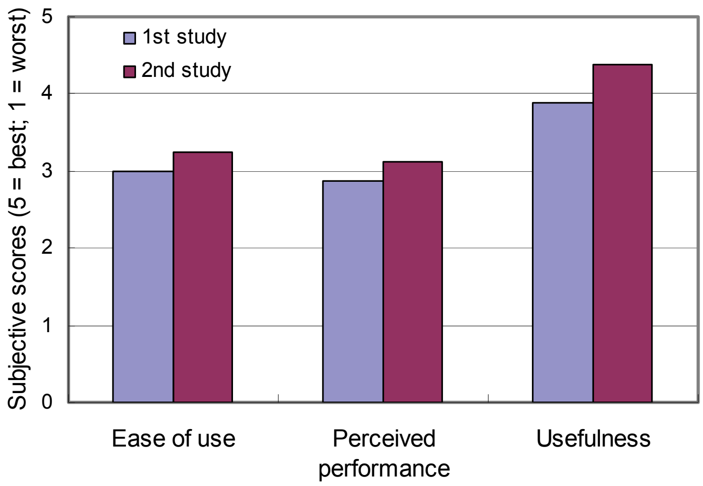


Figure 5.5. The results for text entry rate.

During the study, many participants complained about the isolated speaking style and believed that it greatly affected the input performance. Most of them expected that the application would support continuous dictation, enabling them to enter text in a more natural way without selecting the target pinyin and characters.

5.2.5 Discussion

The results indicated that the 4-best recognition rates in the second study were significantly higher than those in the first. This is largely because of the optimization of the speech recognition engine between the two rounds of user studies. Three major updates of the engine had been done: first, the acoustic model was optimized with more training samples, covering more accents and samples from mobile phones; second, an adaptation algorithm was added for recording channels, and the adaptation helps to fine tune the recordings of people's voices from different pieces of hardware; and, third, the language model was updated with a larger corpus. However, the participants were not significantly faster in their text entry in the second study. The main reason might be the timeouts (500 ms) that I added after each recognition with the second prototype. The timeout was added to avoid mis-activation of the speech recognition engine; however, adding a timeout after each click clearly decreased the text entry rates.

Although enrollment did not help to increase recognition rates in the studies, participants in the first study achieved higher speeds after the enrollment. It is believed that the difference in text entry rates in the first study did not result from the enrollment, since enrollment did not help to increase recognition rates. Instead, since the recognition rates in the first study (around 70%) were much lower than those in the second study (around 95%), the participants needed to spend more time on error correction in the first study. Thus, it is assumed that the increase in text entry rates in the first study was mainly a result of the practice effect on the error correction tasks. Since the recognition rates were very high in the second study, error correction would not require much time or effort. Thus, little practice effect was involved to cause the increase in text entry rates.

The average 4-best recognition rate in the second study reached around 96%. However, it did not result in a comparable increase of subjective scores on usability and perceived performance compared with the results in the first study. This is partly because people don't like the isolated speaking style. Participants in both studies complained about the speaking style and thought it was not so natural to use. Moreover, it was also noticed that when users entered text with the SMS dictation application, they needed to switch between the oral modality and manual modality for different tasks. This was not driven by users but the application. For example, when starting with the recording of voice, users need to press a key to initiate the recording process. Moreover, candidate selection and error corrections were fully done through manual operations of the keys. In the two rounds of evaluation studies, I did not take a close look at the effects of modality switch in the dictation process, and it could be covered in my future work.

5.3 Effects of Speaking Style on Dictation of Mandarin Text Messages

5.3.1 Objectives

In the isolated Mandarin message dictation application, users need to enter messages by speaking to the phone character by character with clear pauses between characters. The discontinuous input style impaired usability of the application. Users expected to speak continuously to the application as they usually do in daily life or at least to avoid the selection of pinyin marks.

There is still more than one way to implement a continuous or quasi-continuous Mandarin message dictation application. This is because there is a middle level between sentence and character in Chinese: the phrases. Hence we have two options for our future work: we can develop either a phrase- or a sentence-based Mandarin message dictation system on mobile phones (Alhonen et al., 2007). The input process with a phrasal dictation system can be as follows: first, users articulate a phrase to the application; second, the system recognizes it and provides a few recognition options for users to select. Since most phrases in Chinese (see Figure 2.4) consist of two to three characters, the advantage of a phrase-based dictation system would be that it has comparatively low requirements for the performance of the mobile device, including its processors and memory. Moreover, it should not require users to select the syllables recognized. Instead, users should be able to select the target phrases directly. However, when compared with sentence input, phrase-based dictation still requires clear pauses between phrases. In that sense, phrase-based dictation is still an isolated recognition solution.

To gain better understanding of users' preferences in relation to speaking styles, I conducted this study exploring effects of speaking styles (phrase-based or continuous speaking styles) on user performance and preference. However, I cannot study the effects of speaking styles without taking some other significant factors into account. Text length and recognition rate, which have already been shown to be able to affect user performance with speech recognition systems (Price & Sears, 2005), were explored together in this study. The effect of recognition rate on user performance and satisfaction with speech input is so apparent that it seems not necessary to cover it in this study anymore. However, it is still included in this study as a key independent variable since I want to understand if interaction effects exist among the three variables on user performance and satisfaction. It is believed that the study's results, especially if there are interactions among the three independent variables, can assist researchers in decisions on how to develop a more useful dictation application for mobile phones.

Because of the lack of working prototypes, I applied the Wizard of Oz method in this study. Wizard of Oz experiments have proven to be an effective method to collect data on interactions between a user and complex systems (Dahlbäck, Jönsson, & Ahrenberg, 1993; Bernsen & Dybkjær, 1998; Fiedler, Gabsdil, & Horacek, 2004). The method has been applied in the design of speech- or pen-based systems (Klemmer et al., 2000; Sinha, Shilman, & Shah, 2001; Lyons, Skeels, & Starner, 2005), context awareness solutions (Davis et al., 2007), augmented reality (Dow et al., 2005), and medical systems (Molin, 2004). Wizard of Oz is usually applied in early stages of system design to gain understanding of the user's mental model and to compare interaction design options, especially when there is no working prototype available. The technique helps to save costs and enables user involvement at an early stage of system design. The technique fits well with our study objectives for the following two reasons.

The study is conducted in an early stage of design of the next version of the Mandarin message dictation application for mobile phones. There is no working prototype available for a true user study. However, understanding user behavior and preferences concerning speaking styles is critical for showing the appropriate direction for development of the system. Hence, we created a software program specifically for this study to simulate the interaction processes with different speaking styles (Liu, Jiang, & Yang, 2009). All short messages that users were to enter were predefined beforehand. Hence, when users interact with the system, all of the necessary interaction elements were there and they could hardly realize that the recognition engine was missing.

In this study, I wanted to address the following questions:

1) What are the effects of speaking style on user performance and preference in combination with recognition rate and text length?

2) What is the interaction effect of the three factors on user performance and preference? Will users have different preferences on speaking style under different circumstances, e.g., with different recognition rates or when entering messages of different length?

3) Will the effects of recognition rate and text length be comparable to past results?

4) What design implications can we find from the study's results concerning Mandarin message dictation systems?

5.3.2 Method

Design

The experiment was a 3×3×2 within-subject design with three independent variables: interaction style, recognition rate, and length of messages. The three interaction styles are phrase-based input with four candidates, phrase-based input with seven candidates, and the sentence input. Three levels of recognition rates were taken into account: 50%, 70%, and 90%. Users needed to enter messages of two different lengths: 10 characters and 20 characters. Thus, at least 18 short messages need to be entered to cover all conditions.

The time for entering each short message was recorded, and the participants were also asked to give a score to each input case with a questionnaire using a five-point Likert scale.

Participants

Twelve users, half of them male and half female, took part in the study. All were users of SMS and the Chinese pinyin input method. All were familiar with the user interfaces of the Symbian S60 UI and the application. All were right-handed.

a) User interface for phrasal input style with four candidates

b) User interface for phrasal input style with seven candidates

c) Sentence input: after users speak to the phone, the whole sentence appears

d) The Nokia E50

Figure 5.6. The interfaces for different speaking styles and the phone used in the experiment: a) phrasal input with four options, b) phrasal input with seven options, c) sentence input, and d) the Nokia E50.

Tasks

The task that participants needed to complete in the experiment was to copy 18 presented messages by speaking to the mobile phone in predefined ways. The outcome messages for each dictation were also predefined to indicate the recognition rate as if there was a functioning speech recognition engine as the back end. For example, if a message with 10 Chinese characters is pre-defined to be 50% correctly recognized, two phrases and one character might be shown to be wrongly recognized. Where the errors appeared in the message would be randomly decided by the program. And the mistaken phrases or characters would be chosen from phrases or characters with similar pronuciations that were prepared beforehand by me. The participants were also instructed to correct all errors with the Chinese pinyin input method provided by the Symbian S60 platform in Nokia E50. It was up to participants whether they would correct an error immediately after it appeared or after a whole message was entered in the phrasal speaking mode. Time spent on error corrections was taken into account in data analysis.

During the task completion process, all messages were shown to the participants with clear indications of the input style to be used for each message. Table 5.4 shows how I presented messages to users. Messages to be entered with phrasal speaking style were shown with an underline between any two consecutive phrases, and messages to be entered with continuous speaking were shown just as they are to participants.

Materials

Eighteen short messages were selected from a real short message corpus for this study. For each message, the incorrectly recognized characters were randomly arranged. Table 5.4 shows examples of the short messages and their form of presentation, indicating different input styles.

Table 5.4: Examples of the presentation of messages

Speaking style	Message with 20 characters	Message with 10 characters
Phrase	每天__高兴__时候__感觉__不错__但是__郁闷__时候__非常__难受	我们__现在__正在__教室__上课
Sentence or continuous speaking	天气温度过高 注意孩子防暑降温多给孩子喝水	衷心希望大家每天快乐

Apparatus

A Nokia E50, with a software program designed specifically for the experiment, was used in the study. Results for task completion time were automatically logged by the program for data analysis.

Procedure

Each participant took part in the experiment individually with a researcher in a quiet lab. First, the researcher briefed the participant on the objectives of the study. Then the researcher explained to the participant how to use the dictation application to enter messages. There was also a trial session in which the participant could practice until ready to start the data collection sessions. The data collection phase included two sessions: the 10-character message session and the 20-character message session. To balance out any effects of testing order, half of the participants started with the 10-character message session and half started with the other one. The Latin-square experiment technique was applied to offset any possible effects caused by testing orders of speaking style and recognition rate.

When entering a message, a participant needed to speak to the mobile phone in the predefined way and make sure that the message was entered exactly like the one presented. After entering a message, participants gave a score to the task difficulty, with a five-point Likert-scale questionnaire. The process was repeated until the participant had entered all 18 messages. Each participant received a gift after the experiment.

5.3.3 Results

Results for Task Completion Time

Results related to task completion times are shown in Table 5.5 and Figure 5.7. A three-factor repeated ANOVA test was applied to analyze the data.

Main effects of all three factors are significant (recognition rate: $F_{2,22}$ = 46.88, p < .001; interaction style: $F_{2,22}$ = 22.73, p < .001; SMS length: $F_{1,11}$ = 27.61, p < .001). None of the interactions were significant. Further analysis of the effect of recognition rate indicated that there were significant differences between any two of the three recognition rates. The higher the recognition rate, the less time was needed to enter a message. Further analysis of speaking styles found that completion times with sentence input were significantly shorter than with either of the phrasal input styles. No difference was found between the two phrasal input styles. It was also found that participants needed to spend significantly longer to enter the 20-character messages than the 10-character messages.

Table 5.5: Task completion times, with standard deviations

Task completion time (s)		Speech recognition rates (%)					
		50%		70%		90%	
		Message lengths (in characters)					
		10	20	10	20	10	20
Four-phrase input	Mean	74.33	117.24	69.99	98.05	42.66	63.83
	SD	18.152	17.637	32.248	36.596	10.475	12.486
Seven-phrase input	Mean	94.25	117.78	61.62	103.88	42.52	67.64
	SD	22.945	20.053	21.719	43.302	14.189	11.084
Sentence input	Mean	65.34	100.76	58.383	73.78	29.70	42.85
	SD	16.773	34.463	28.972	17.355	15.132	10.816

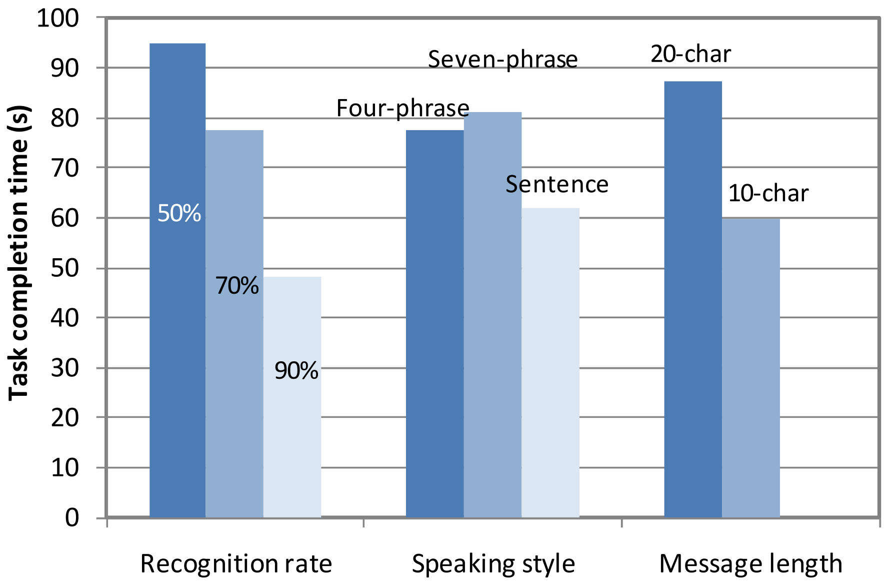


Figure 5.7. Comparison of task completion times.

Results for Text Entry Rate

I converted the task completion time into text entry rate in CCPM and explored the effects of recognition rate, speaking style, and message length on it. How I calculated CCPM from the task completion time is shown in

Equation 5-1, where N is the number of characters that people entered (with punctuation) and T is the task completion time in seconds.

$$\text{CCPM} = N/T \times 60 \tag{5-1}$$

Table 5.6 shows the results for text entry rates. Recognition rate, interaction style, and message length all affected text entry rates as shown in the table. Text entry rates increased as the recognition rate increased. The text entry rates with sentence input were all higher than those with phrasal inputs. On the other hand, text entry rates with the two phrasal inputs were similar, and there was no clear pattern in their differences. Text entry rates in dictation of long messages were higher than those for dictation of short messages.

Table 5.6: Means and standard deviations for input speed

Input speed (characters per minute)		Speech recognition rates (%)					
		50%		70%		90%	
		Message lengths (in characters)					
		10	20	10	20	10	20
Four-phrase input	Mean	8.75	10.44	9.73	13.41	14.83	19.39
	SD	3.254	1.542	2.964	3.510	3.514	3.350
Seven-phrase input	Mean	6.67	10.48	10.61	13.18	15.29	18.17
	SD	1.408	1.886	2.806	4.490	4.069	2.896
Sentence input	Mean	9.79	12.91	11.84	17.28	25.69	29.94
	SD	2.729	3.342	3.851	4.880	14.239	8.843

Table 5.7: Effect of recognition rate, input style, and SMS length on input speed

Source	*F*	*p*
Recognition rate	87.884	< .001
Speaking style	27.088	< .001
Message length	27.053	< .001
Recognition rate × speaking style	7.494	< .001
Recognition rate × message length	0.249	.780
Speaking style × message length	0.281	.755
Recognition rate × speaking style × message length	0.342	.850

Table 5.7 shows the three-factor repeated ANOVA test results for text entry rate. Main effects of recognition rate ($F_{2,22}$ = 87.884, $p < .001$), input style ($F_{2,22}$ = 27.088, $p < .001$), and message length ($F_{1,11}$ = 27.053, $p < .001$) are all significant. Moreover, the interaction effect between recognition rate and speaking style was significant ($F_{4,44}$ = 7.494, $p < .001$). Further analysis indicated that text entry rates with different speaking styles were

significant at the 0.05 level when recognition rates were 50% and 70% (50%: $F_{2,22} = 5.517$, $p = .005$; 70%: $F_{2,22} = 3.525$, $p = .035$). Text entry rates with the three speaking styles were significantly different when the recognition rate is 90% ($F_{2,22} = 16.782$, $p < .001$). No other significant interactions were found.

Results for Subjective Scores

Results regarding subjective scores are shown in Table 5.8.

Table 5.8: Descriptive results concerning subjective evaluations

Subjective scores (1 = very dissatisfied, 3 = neutral, 5 = very satisfied)		Speech recognition rates (%)					
		50%		70%		90%	
		Message lengths (in characters)					
		10	20	10	20	10	20
Four-phrase input	Mean	1.83	2.67	2.33	3.25	4	4.33
	SD	0.835	0.778	0.888	0.754	0.603	0.492
Seven-phrase input	Mean	1.75	2.75	2.42	3.83	4	4
	SD	0.866	0.754	0.996	0.835	0.739	0.426
Sentence input	Mean	1.33	2.75	2.67	3.67	4.33	4.42
	SD	0.492	0.866	0.888	0.778	0.651	0.515

Table 5.9: Effect of recognition rate, input style, and SMS length on subjective scores

Source	***F***	***p***
Recognition rate	135.644	< .001
Speaking style	0.933	.408
Message length	38.047	< .001
Recognition rate × speaking style	3.196	.022
Recognition rate × message length	9.683	.001
Speaking style × message length	0.376	.691
Recognition rate × speaking style × message length	1.883	.130

Table 5.9 shows the ANOVA test results, which indicate that the main effects of recognition rate ($F_{2,22} = 135.644$, $p < .001$) and SMS length ($F_{2,22} = 38.047$, $p < .001$) were significant. Recognition rate had significant interactions with SMS length ($F_{2,22} = 9.683$, $p < 0.001$) and speaking style ($F_{4,44} = 3.196$, $p < .05$). No other significant effect was found. When the recognition rates were 50% and 70%, participants gave similar scores to the three speaking styles. When the recognition rates increased to 90%, they gave higher scores to sentence input but the difference is not statistically significant ($F_{2,22} = 2.546$, ns). When the recognition rates were

70% and 50%, users gave significantly higher scores to tasks of entering 20 characters (50%: $F_{1,11} = 35.609$, $p < .001$; 70%: $F_{1,11} = 30.075$, $p < .001$). When the recognition rate reached 90%, users were inclined to give similar scores to both types of tasks ($F_{1,11} = 1$, ns).

5.3.3 Discussion

The results of this experiment showed that text entry rates for Mandarin dictation range from 6.67 CCPM to 29.94 CCPM, depending on recognition rate, input style, and the length of the message entered. However, it is worth noting that the results were obtained from a Wizard of Oz experiment wherein time required for system processing, such as the computing process of the recognition engine, was not addressed. The results also indicated that the main effects of recognition rate, input style, and message length on text entry rates were significant, which verified existing results from relevant studies (Sinha, Shilman, & Shah, 2001; Price & Sears, 2005; Ranjan et al., 2006). Moreover, the interaction effect of recognition rate and speaking style on text entry rate was significant, indicating higher recognition rate with the sentence speaking style would significantly increase text entry speeds.

I analyzed the dictation tasks to explain the effects. The phrasal dictation task can be divided into three sub-processes: speaking process, candidate selection from phrase options, and error correction. Past linguistic studies discovered interesting results concerning the Mandarin speaking process. Pellegrino, Farinas, and Rouas (2004) reported that the average speaking rate for Mandarin is three syllables (characters) per second when people speak at a normal speed. Jeng (2005) found that the size or length of speaking units significantly affects the duration of each syllable. The duration of uttering a syllable in the monosyllable condition is greater than that in the sentence speaking condition, which means that the speed for speaking a sentence could be higher than that for a phrase or a character. I recorded and calculated participants' speaking time when they uttered a whole sentence. The average speaking times for the 20-character messages and the 10-character messages are 6.2 seconds ($SD = 0.38$) and 4.2 seconds ($SD = 0.18$), respectively, and the corresponding speaking speeds are 3.2 and 2.4 syllables per second.

I assumed that the speaking time in the sentence dictation and phrase dictation were the same. Moreover, the time used to correct errors in messages of the same length could also be assumed to be the same when the recognition rate is the same. Hence, the only difference between the phrasal and the sentence dictation is that the latter doesn't include the candidate selection process. I subtract the time for sentence dictation from the time for phrasal dictation; the results would be the time needed for the candidate selection process. What is more, if I subtract the speaking time and the candidate selection time from the total time for phrasal dictation, I can obtain the time needed for error correction. Table 5.10 and Table 5.11 present the time required for each sub-process in absolute terms and as a percentage of the dictation process, respectively. According to the tables, error corrections take the greatest percentage of the task completion time, followed by candidate selection. Speaking takes the smallest proportion of the time.

Table 5.10: Average time required by the sub-processes

		Recognition rates					
		50%		70%		90%	
		Message lengths					
		10	20	10	20	10	20
Average error correction time (seconds)		61.17	94.56	54.213	67.58	25.53	36.65
Speaking time (seconds)		4.17	6.2	4.17	6.2	4.17	6.2
Candidate selection time (seconds)	Phrasal-4	8.99	16.48	11.607	24.27	12.96	20.98
	Phrasal-7	28.91	17.02	3.237	30.1	12.82	24.79

All effects can be explained from the task analysis results. First, since the sentence dictation does not require the sub-process of candidate selection, sentence dictation is faster than phrasal dictation. Second, the higher recognition rate reduces the time used for error correction, which resulted in increased text entry rates. The results also indicated that recognition rate and interaction style interactively affected text entry rate. When the recognition rate reached 90%, the text entry rate for sentence input was much better than that of phrasal input, whereas it was just marginally better than phrasal input when recognition rates were 50% and 70%. This is because, when the recognition rate is low, error correction took the most user effort and it offset the advantages of sentence input for candidate selection. However, when the recognition rate increased to 90%, user effort and time in error correction decreased and the advantages of sentence input in terms of candidate selection and speaking speed were comparatively increased.

The results also indicated that the text entry rate for entering 20-character messages was greater than that for entering 10-character messages. The result supported theoretical models provided by Price and Sears (2005). The models were to predict effects of error correction rate, recognition accuracy, and latency on text entry rate in text dictation tasks. One of the key assumptions was that text entry rate would increase as messages became longer.

Table 5.11: The sub-processes as a proportion of the total task completion time, average

Percentage of the task completion time		Recognition rates					
		50%		70%		90%	
		Message lengths					
		10	20	10	20	10	20
Error correction	Phrasal-4	82.30%	80.66%	77.46%	68.92%	59.85%	57.42%
	Phrasal-7	64.90%	80.29%	87.98%	65.06%	60.04%	54.18%
	Sentence input	93.62%	93.85%	92.86%	91.60%	85.96%	85.53%
Speaking time	Phrasal-4	5.61%	5.29%	5.96%	6.32%	9.77%	9.71%
	Phrasal-7	4.42%	5.26%	6.77%	5.97%	9.81%	9.17%
	Sentence input	1.43%	0.93%	1.59%	1.24%	2.89%	2.00%
Candidate selection	Phrasal-4 percentage	12.09%	14.06%	16.58%	24.75%	30.38%	32.87%
	Phrasal-7 percentage	30.67%	14.45%	5.25%	28.98%	30.15%	36.65%

The main effects of recognition rate and message length are significant in subjective scores, which indicated that these two factors were the main ones considered by the participants when they gave scores. Input style also interactively affected the subjective scores in combination with recognition rate. When the recognition rate is 90%, participants preferred the sentence input over the phrasal input options. When the recognition rates were 50% and 70%, the participants gave similar scores to all three input styles. This was probably because when the recognition rate was low, people still paid more attention to the error correction process and didn't have high requirements for the input styles. However, when the recognition rate was high enough, people started to think about having a better input style.

The research findings implied certain design guidelines for a Mandarin dictation system. Recognition rate is the key factor affecting user performance and satisfaction with a Mandarin dictation system on mobile phones. This is because recognition rate determines the effort required for error correction. Error correction with keypad-based methods takes the most effort in a Mandarin message dictation task, which implies that efficient methods for error correction are critical for improving usability of speech dictation systems. When the recognition rate of a system is high, users expect to be able to speak in a more natural way. When the recognition rate of a system is not high enough, providing an optimal error correction method would be critical to improve its usability.

5.4 SUMMARY

Speech dictation is a useful text entry method especially for mobile devices that are used in diverse contexts. In this chapter, human factor studies of Mandarin message dictation solutions were presented. First, two evaluation studies of an isolated Mandarin dictation solution were presented, to explore its utility and usability. The results indicated that the top four recognition rates with the isolated Mandarin message dictation application can reach around 95%. However, the user enrollment system does not work well with native Mandarin speakers: recognition rates did not increase in either study after user enrollment. Moreover, average text entry rates with the application are in the range 8–10 CCPM. Although the recognition rate increased from around 70% to about 95% in the second study, the text entry speeds in the second study were not significantly greater than those in the first. It is believed that the results were caused by the extra 500 ms added after each recognition in the second prototype to avoid unintentional activation of the recognition engine.

The isolated Mandarin dictation was not preferred by participants; they found it not natural enough. In the second part of the chapter, I presented a Wizard of Oz study to explore effects of speaking style on user performance and satisfaction, along with two other variables: recognition rate and message length. The experiment's results indicated that all three factors affected user performance, including task completion time and text entry rate. Further task analysis results showed that error corrections took the most time in the dictation process, followed by the selection of target phrases and the speaking phase. The results can be applied to guide interaction designs for Mandarin dictation systems on mobile devices. The results related to text entry rates should be verified with a full working implementation.

6 Keypad Pinyin Solutions

6.1 INTRODUCTION

Pinyin text entry methods are the primary methods used by Chinese users on both personal computers and mobile devices (Sheng, 1985; Sacher, 1998; Sacher, Tng, & Loudon, 2001; Lin & Sears, 2005, 2007; Liu & Wang, 2007; Liu & Räihä, 2008; Liu & Räihä, 2010). Although many novel devices are designed for mobile devices, the 12-key keypad is still the dominant input device because of its familiarity, and also because its compact size is suitable for being held and used with one hand. Chinese pinyin text entry methods based on the 12-key keypad can be classed into two groups: methods supporting phrasal input and methods supporting predictive character input only. Pinyin methods supporting phrasal input enable users to enter a phrase that includes more than one character by typing pinyin marks of the corresponding characters. Moreover, the phrasal pinyin input methods also allow users to enter text character by character with the predictive feature. Hence, the phrasal pinyin input methods are not exclusive. Instead, they also support character input with the predictive feature. Pinyin methods supporting only character input allow users to enter only one character at one time, but they usually allow predictive input.

In this chapter, I present two studies, of the phrasal pinyin input methods and character pinyin input methods, respectively. The first study was aimed at providing interaction design guidelines for phrasal pinyin input methods. In this study, we compared the usability of five phrasal pinyin input methods by analyzing both objective and subjective measurements. The second study was designed for building a user model to estimate average users' error-free speeds with the two types of pinyin character input methods when the predictive input was turned on and off.

The rest of the chapter is organized as follows. First, I present the evaluation of five pinyin phrasal input methods and provide relevant design guidelines. Second, I present the predictive user model for users' error-free speeds with the character input along with four experiments: first, two for definition of parameters, then two for evaluation of the model. Finally, I summarize the key conclusions from both studies.

6.2 Evaluating Pinyin Phrasal Input on Mobile Phones

6.2.1 Objectives

The 12-key keypad brings ambiguity to the Chinese pinyin phrasal input. A sequence of key presses may still result in more than one pinyin option for different phrases. For example, if users type the "64426" key sequence, there are still multiple pinyin options, including "nihao," "migao," and "nigan" (see Figure 6.1). Two types of pinyin phrasal input methods were designed to cope with this challenge. One type requires users to select first the target pinyin mark and then the target phrase (see Figure 6.1). The other type allows users to skip the process of selecting the target pinyin mark, by utilizing integrated language models, and directly get the target phrases (see Figure 6.2). Figures 6.1 and 6.2 show the input processes with examples representing the two types of pinyin phrasal input methods.

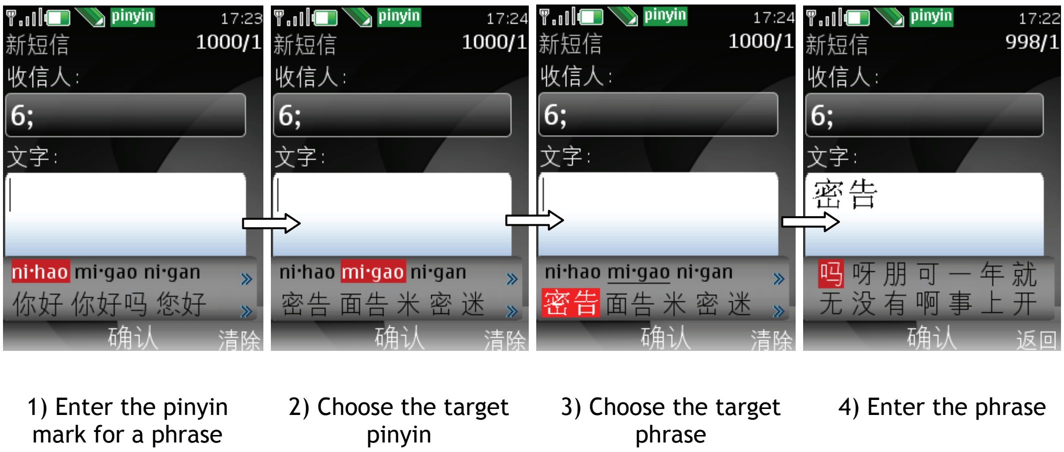
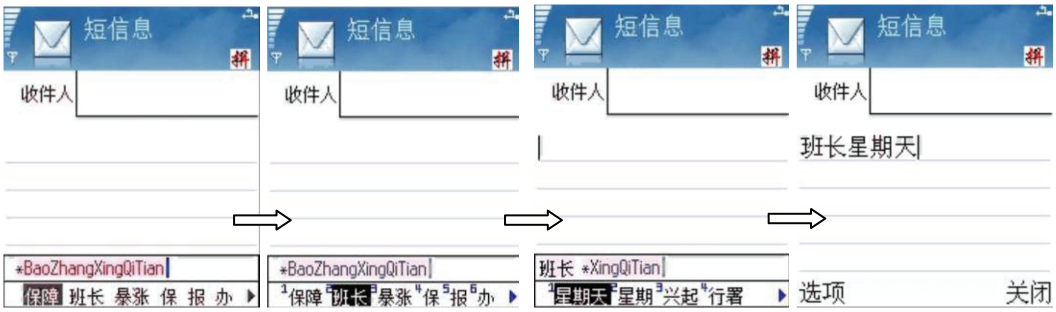


Figure 6.1. A phrasal input pinyin method requiring selection of the target pinyin.

Phrasal input is a new feature for mobile text entry methods. There are neither published studies of the interaction process nor guidelines provided to aid practitioners with interaction design. This study aimed to provide guidelines for the pinyin phrasal input feature by comparing five interaction designs, with user research. The five interaction designs represent the two types of pinyin phrasal input: Guobi and Cstar solutions, requiring no selection of the target pinyin marks, and solutions of Nokia, T9, and Zi, requiring selection of the target pinyin (the five solutions are called Guobi, Cstar, Nokia, T9, and Zi, respectively, later in the chapter for short). To understand the effects of method category (requiring pinyin selection vs. no pinyin selection) and interaction design on user performance and preference, this study covered more than one interaction design from each category of pinyin phrasal input. For example, Guobi and Cstar differ in key mappings of functions and feedback in the input process. When users enter the pinyin mark of “nihao” with Cstar, only the entered pinyin mark is shown to users. But with Guobi, both the numbers associated with the pinyin mark (i.e., “64426”) and the pinyin mark itself (i.e., “nihao”) will be shown to users.



1) Enter pinyin marks for two phrases
2) Choose the target phrase for the first phrase
3) Choose the target for the second phrase
4) Enter the phrases

Figure 6.2. A pinyin phrasal input solution requiring no selection of the target pinyin.

In the following sections, I first explain the method that was utilized in the study. Then I present the results. Finally I discuss the results and summarize design guidelines for the pinyin phrasal input feature.

6.2.2 Method

Participants

Eighteen volunteers, half male and half female, took part in the study. Their ages ranged from 20 to 32, with an average of 23.8 years (SD = 3.35). All were users of the pinyin input methods while composing SMS messages. They send about 15 (SD = 7.4) text messages per day, on average.

Six of the 18 participants (3 male and 3 female) were trained users of the phrasal input solutions. They began to use the five devices two months before the experiment, using each of the five phrasal pinyin inputs as their daily input method for approximately 10 days. The other 12 were novice users. Training orders for the five methods were balanced among all participants with an incomplete Latin square technique.

Apparatus

Five pinyin phrasal input methods were studied on four mobile phones, because the phrasal input is a new feature and there is no uniform platform properly supporting all methods. I used the Nokia 6630 with Cstar and Guobi, Nokia 6131 with Nokia, Nokia E70 with Zi, and

Samsung E728 with T9. A mini-camera, attached to the mobile phones, was used to record the phone display and user operations on the keypad. This information was later analyzed manually.

Tasks and Materials

The task was to enter six short messages that include 162 Chinese characters and 16 punctuation marks (6 full stops, 6 commas, 2 question marks and 2 exclamation points). Table 6.1 shows three examples of the six messages. All of the messages were selected from our short message corpus described in Section 2.2. I calculated the average number of letters per pinyin mark for the six short messages to be 2.88, which was identical to the result that was calculated from our short message corpus. The single-letter correlation (MacKenzie & Soukoreff, 2003) of the six messages with the short message corpus presented in Section 2.2 was 0.921.

Table 6.1: The message examples used in the experiment

No.	Messages	
1	Message	今天在实验室手机没电，现在才看到短信。多谢张老师通知！
	Pinyin	Jin tian zai shi yan shi shou ji mei dian, xian zai cai kan cao duan xin. Duo xie zhang lao shi tong zhi!
	Meaning	I just saw your message right now since my mobile phone ran out of battery when I was in the lab. Thank you for notifying me of this!
2	Message	我用心感知上帝的存在。你并不能证明上帝不存在，是吧？
	Pinyin	Wo yong xin gan zhi shang di de cun zai. Ni bing bu neng zheng ming shang di bu cun zai, shi ba?
	Meaning	I sense God with my heart. You cannot prove that God does not exist, can you?
3	Message	实验软硬件的相关信息我已经收到了，十分全面。非常感谢!
	Pinyin	Shi yan ruan ying jian de xiang guan xin xi wo yi jing shou dao le, shi fen quan mian. Fei chang gan xie!
	Meaning	The software and hardware for the experiment were received. Thank you very much!

Design and Procedure

The experiment was a 5 (the five phrasal pinyin input methods) × 2 (two user experience levels) two-factor mixed design. Each participant was instructed to complete the tasks with all five phrasal pinyin input methods.

The order of the methods evaluated and of the messages were counterbalanced across users with the Latin-square technique.

The experiment was conducted in a lab environment with a participant and a coordinator together completing the experiment process. First, the profile of each participant was collected and the research goals were explained by the coordinator. The participant could ask questions at will. A tutorial session then followed, in which use of the five phrasal input methods was explained to the participant. Then the participant was instructed to practice with the five designs until he or she was ready to start with the tasks for data collection. Participants were instructed to complete the tasks as quickly and accurately as possible with the phrasal input solutions. After the experimental task with a phrasal pinyin input method, the participant was asked to evaluate it with a five-point Likert-scale questionnaire addressing its usability. The whole experiment was recorded with a video camera for analysis later, with the agreement of the participants. A small gift was presented to the participant at the end of the experiment.

Measures

Three metrics were utilized in the data analysis phase: keystrokes per character to characterize different input solutions, Chinese characters per minute for text entry rate, and error rate.

I revised and applied KSPC in the study (MacKenzie, 2002a):

$$KSPC = \frac{C + IC + F}{N1 + N2} \tag{6-1}$$

Here *C* refers to the correct keystrokes, *IC* represents the incorrect keystrokes, *F* indicates keystrokes to fix errors, and *N1* is the total number of Chinese characters while *N2* is the total number of punctuation marks. Here characters in KSPC mean the Chinese characters instead of the Roman letters in this study.

The following equation shows how I calculated CCPM:

$$CCPM = \frac{N1 + N2}{T} \tag{6-2}$$

where *T* indicates the total amount of time (in minutes) for accomplishing the task and the meanings of *N1* and *N2* are the same as they are in the equation for KSPC. In the calculation of CCPM, I treated a punctuation mark equally with a Chinese character since punctuation marks normally require more than one key press with the 12-key keypad. Moreover, punctuation marks are only small part of the 6 messages to be entered in the task so they would not affect the CCPM results much. There are in total 16 punctuation marks and 162 Chinese characters in the 6 messages and thus punctuation marks are only about 8.99% of the total text.

The following equation shows how I calculated error rate:

$$ER = \frac{E}{N1 + N2} \tag{6-3}$$

E indicates the total number of errors in the input process, and the meanings of *N1* and *N2* remain the same as in the equations for KSPC and CCPM. Five categories of errors were taken into account in the calculation of the error rate: 1) misspelled pinyin, where participants entered wrong letters for a pinyin mark; 2) missed target selection, where participants failed to select the target character or phrase; 3) over-selection, where participants passed the target by pressing the navigation key too many times or entered the wrong number tag for a target; 4) other action errors, including pressing of the "complete SMS" key and accidentally exiting the editor, or pressing the Clear key and deleting characters accidentally, or pressing the wrong keys when switching language mode; and 5) numbers of differences between the messages presented and messages entered by the participants. Whether corrected in the procedure or not, they were included in the analysis for error rate.

6.2.3 Results

Qualitative Results

In the experiment, participants often failed to identify phrases that a phrasal pinyin input method supports: after participants entered the pinyin mark for a phrase, corresponding phrase options did not appear, because the target phrase was not covered. The five methods evaluated have different ways of handling the problem. Guobi and Cstar enable on-line creation of new phrases to enlarge their bank of supported phrases. For example, if entering a pinyin mark for a phrase including two characters and finding that the system does not support it, users can move the cursor from the end of the pinyin mark for the phrase to each end of the pinyin for the two characters (in this phrase) and choose the target characters to complete the phrase. The Guobi and Cstar methods would then automatically memorize the phrase such that when users enter them again, they can get them by entering the pinyin mark for the phrase. By contrast, with the other three methods, the Nokia, T9, and Zi, participants have to delete the entered pinyin mark for the second character and enter this phrase character by character.

The design of hardware also affected user performance. The E70 used in the test was a new phone, and its five-way navigation key was difficult to operate. I observed that the confirmation click was often detected as a scroll-down click and vice versa. Also, the keys of the Nokia 6131 (a folding phone) were a bit too flat to click for some users.

Quantitative Results

As mentioned in Section 6.2.2, I collected and calculated user data for both objective measurements (including KSPC, text entry rate in CCPM, and total error rate) and subjective measurements (of the overall usability of the five phrasal input methods with a five-point Likert scale). I then conducted a two-factor mixed-design ANOVA test to explore the effects of method and experience level on the four dependent measures: KSPC, text entry rate, error rate, and subjective score for usability. Table 6.2 shows the test results. According to it, the main effects of method on the four dependent variables are all significant (KSPC: $F_{4,64} = 16.8$, $p < .01$; CCPM: $F_{4,64} = 17.1$, $p < .01$; error rate: $F_{4,64} = 7.1$, $p < .01$; subjective score: $F_{4,64} = 17.8$, $p < .01$) and the interactions between method and the experience level of participants were all insignificant (KSPC: $F_{4,64} = 1.6$, ns; CCPM: $F_{4,64} = 1.7$, ns; error rate: $F_{4,64} = 2.5$, ns; subjective score: $F_{4,64} = 1.6$, ns). The main effects of experience level were significant for KSPC ($F_{1,16} = 5.2$, $p < .05$) and text entry rate ($F_{1,16} = 17.3$, $p < .01$) results but not significant for error rates ($F_{1,16} = 3.56$, ns) or the subjective scores ($F_{1,16} = 1.5$, ns). Since for error rates and subjective scores, only the main effect of methods was significant, I merged the data from novice and trained users and present them together.

Table 6.2: The two-factor mixed-design ANOVA test results

	KSPC		Text entry rate (CCPM)		Error rate		Subjective score	
	F	***p***	***F***	***p***	***F***	***p***	***F***	***p***
Phrasal input method	16.8	.000**	17.1	.000**	7.1	.005**	17.8	.000**
Experience level	5.2	.036*	17.3	.001**	3.56	.08	1.5	.244
Method × experience level	1.6	.21	1.7	.19	2.5	.11	1.6	.22

* $p < .05$; ** $p < .01$.

Figure 6.3 shows the average KSPC results for both novice and trained participants. The results showed that Zi cost the most keystrokes per character and Cstar the fewest keystrokes per character for both user groups. Trained participants clearly pressed fewer keys on average for entering a Chinese character than the novice participants did. A further *t*-test on the KSPC results indicated that Cstar required significantly fewer

KSPC than all other methods (t = 2.0 (Nokia), 2.7 (Guobi), 3.8 (T9), 7.2 (Zi), $p < .05$ for all) and Zi required significantly more KSPC than all other methods (t = 7.2 (Cstar), 6.0 (Nokia), 5.9 (T9), 5.3 (Guobi), $p < .05$ for all). Nokia was significantly better than T9 ($t = 2.5$, $p < .05$), but both Nokia and T9 showed no significant difference from Guobi (Nokia vs. Guobi: $t = 1.6$, ns; T9 vs. Guobi: $t = 0.4$, ns).

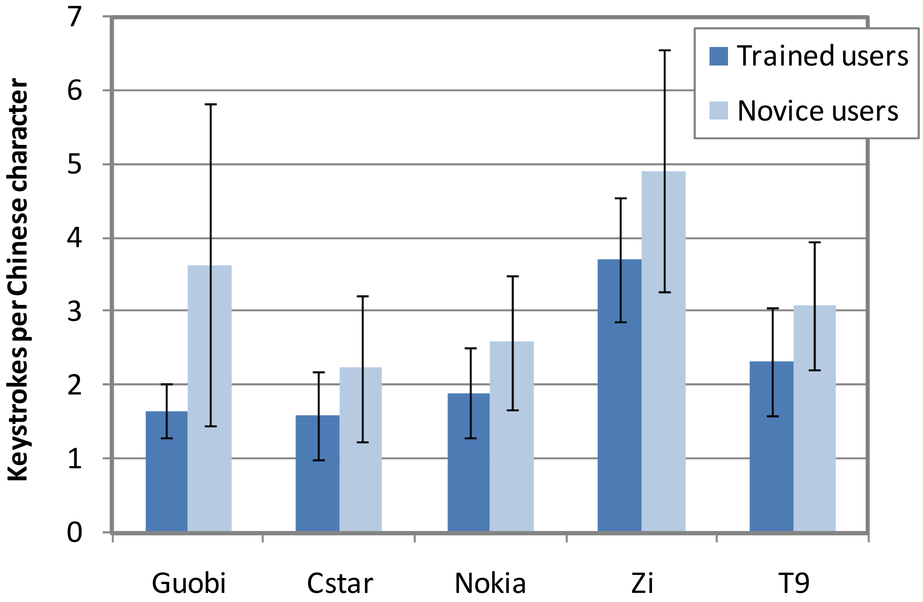


Figure 6.3. Average KSPC results for the phrasal input methods.

Figure 6.4 shows the average CCPM results for both novice and trained participants. Both groups of participants achieved the highest CCPM figure with Cstar and the lowest with Zi. Trained users were clearly faster than novice users. I conducted a t-test to understand the main effect of method. The results showed that the text entry rate with Cstar was significantly higher than that with the other designs (t = 2.0 (Nokia), 2.8 (Guobi), 3.1 (T9), 6.7 (Zi), $p < .05$ for all). The text entry rate with Zi was significantly lower than those with the other designs (t = 6.7 (Cstar), 5.9 (Nokia), 6.0 (Guobi), 5.5 (T9), $p < .05$ for all). The text entry rate with Nokia was significantly higher than that with T9 ($t = 2.3$, $p < .05$), but neither Nokia nor T9 showed a significant difference with Guobi for text entry rates (Nokia vs. Guobi: $t = 0.7$, ns; T9 vs. Guobi: $t = 1.4$, ns).

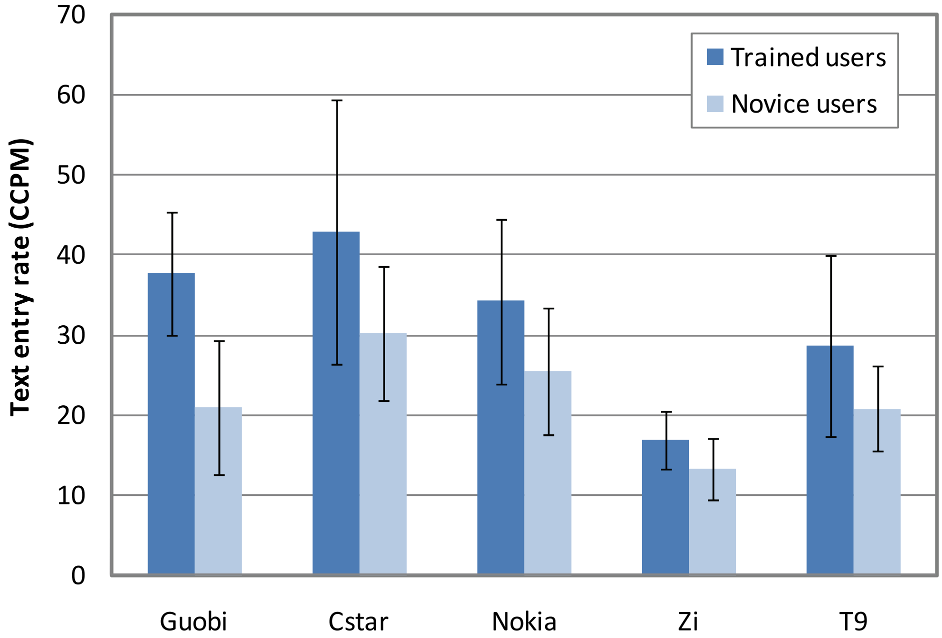


Figure 6.4. Average text entry rates for the phrasal input methods.

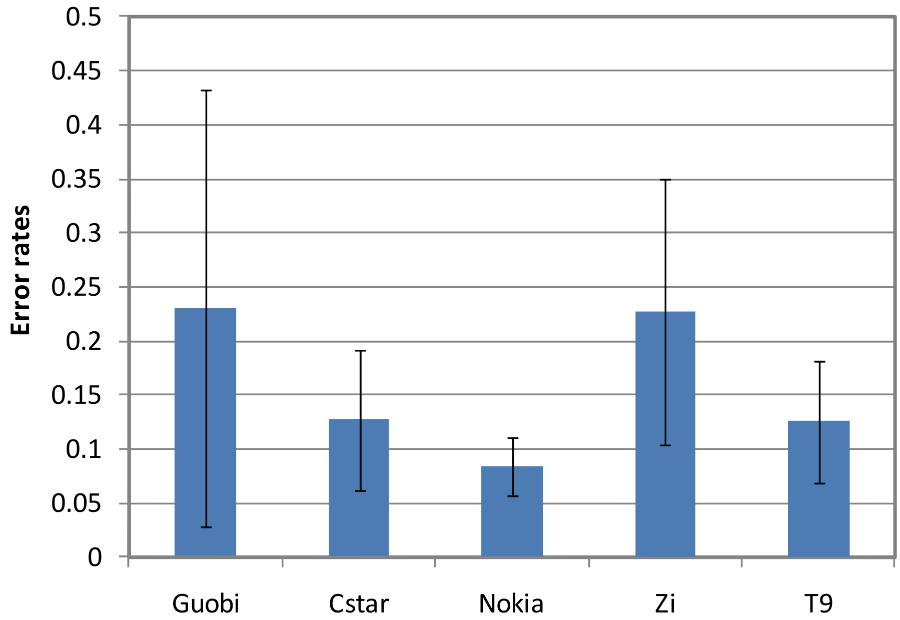


Figure 6.5. Average error rates for the phrasal input methods.

Since error rates were not significantly affected by users' experience level, the error rate data from novice and trained users were merged and analyzed together. Figure 6.5 shows the average error rate results for all participants. According to it, the participants made the fewest errors with Nokia, followed by T9 and Cstar. The participants made the greatest number of errors with Guobi, then Zi. A further *t*-test showed that the five designs can be classified into three groups: Nokia with the lowest error rate, Cstar and T9 next, and then Guobi and Zi with the highest error rate. Differences in error rate within each group were not significant, while differences between any two methods belonging to different groups were significant (Nokia vs. Cstar: $t = 2.5$, Nokia vs. T9: $t = 2.7$, Cstar vs. Guobi: $t = 2.7$, Cstar vs. Zi: $t = 6.7$, T9 vs. Guobi: $t = 2.4$, T9 vs. Zi: $t = 5.4$, with $p < .05$ for all).

Since the subjective scores were not significantly affected by participants' experience level, the results for subjective scores by novice and trained participants were also merged and analyzed together. Figure 6.6 shows the average subjective scores from all participants for the five phrasal input methods. A further *t*-test showed that the five methods could be classed into two groups: participants were more satisfied with Cstar, Guobi, and Nokia than with T9 and Zi. For methods in the same group, their subjective scores had no significant difference, but for any two methods that belonged to different groups, their subjective scores were significantly different (Cstar vs. T9: $t = 5.0$, Cstar vs. Zi: $t = 9.2$, Nokia vs. T9: $t = 4.1$, Nokia vs. Zi: $t = 3.4$, Guobi vs. T9: $t = 5.8$, Guobi vs. Zi: $t = 7.3$, with $p < .05$ for all).

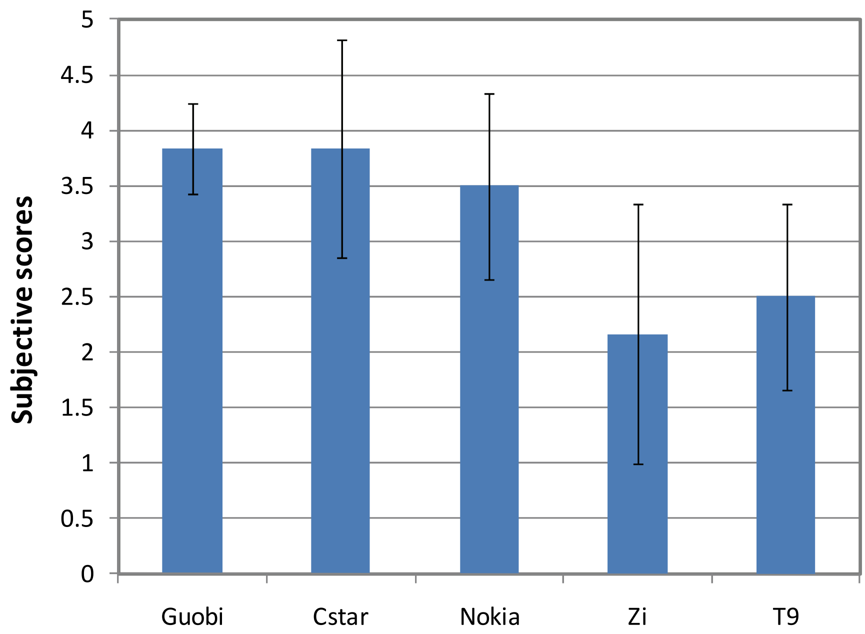


Figure 6.6. The average results for the phrasal input methods, according to subjective scores (where 5 = best and 1 = worst).

I summarize all quantitative results from this study in Table 6.3. Any changes on the gray scale indicate statistical significance. However, since Guobi had no significant differences with either Nokia or T9 for KSPC and for text entry rate, the cell for Guobi is divided into two parts, one like the gray for Nokia and the other half like the gray for T9. According to Table 6.3, the results for KSPC and text entry rate were quite consistent with each other.

Table 6.3: The ranking of the five phrasal pinyin input methods as indicated by the summary of the quantitative results and figures (1 = the highest rank, best … 5 = fifth rank, worst)

	Cstar	Nokia	Guobi	T9	Zi
KSPC	1st	2nd	4th	3rd	5th
Text entry rate (CCPM)	1st	2nd	3rd	4th	5th
Error rate	3rd	1st	5th	2nd	4th
Subjective score	1st	3rd	1st	4th	5th

6.2.4 Discussion

According to the analysis of the Chinese SMS corpus in Section 2.2, a pinyin mark on average includes 2.88 letters when frequencies of Chinese characters in the SMS corpus are taken into account. Moreover, according to Figure 2.5b, when users enter a Chinese character with the 12-key keypad-based pinyin method, they need to press at least two extra keys to respectively confirm the target pinyin and character. In this sense, the minimal KSPC for pinyin methods that are based on the 12-key keypad and enable character input only is 4.88 when errors are not taken into account. In this experiment, the messages entered by participants are also characteristic of an average 2.88 letters per pinyin mark. But according to the KSPC results of this study (see Figure 6.3), the average empirical KSPC for Chinese phrasal pinyin input methods like Cstar, Nokia, T9 and Guobi were below 4 when error rates were also covered in the calculations, which is much lower than 4.88. In this sense the phrasal input methods reduce KSPC compared with pinyin methods that support character input only. Phrasal input treats a phrase with more than one character as an entry unit. Hence selection of a target phrase will cover more than one character, which saves some keystrokes. Moreover, phrasal input methods like Cstar and Guobi require no selection of target pinyin marks but only for phrases, which can probably save some extra strokes.

By comparing the performance of the trained and novice user groups, I found that experience level affected the results for KSPC and text entry rates but not the error rate or subjective scores. Trained users entered Chinese characters with significantly fewer KSPC and at a higher speed. The higher text entry rates of trained users may have mainly resulted from the significant decrease in KSPC, since there was no significant difference between trained and novice users in error rates. This is perhaps another key difference between phrasal pinyin input methods and character pinyin input methods. With pinyin input methods supporting character input only, the average KSPC would be supposed to be rather stable with a fixed interaction design for both novice and trained users. But with pinyin input methods supporting phrasal input, the ways how trained users and novice users use it could vary a lot because of many reasons. For example, trained users could have a better estimation on the phrase sets that a phrasal input can support. Moreover, interaction designs for phrasal pinyin input methods would be more complex than methods supporting character input only. Hence, effects of training on user performance with phrasal pinyin input methods would be more apparent than with pinyin input methods supporting character input only.

Both performance and subjective evaluation results showed that Cstar and Nokia performed well. They were either no. 1 or no. 2 for three out of the four measurements. The participants made fewer errors with Nokia than Cstar, but they achieved a better text entry rate and KSPC value with Cstar than Nokia. Cstar does not require selection of pinyin marks, which may effectively decrease the average KSPC figure and then result in higher text entry rates. Subjectively, Cstar was also slightly preferred over Nokia. During the experiment, it was observed that the participants' hands fit Cstar and Nokia easily and the input processes were relatively smooth. Moreover, both methods were carefully designed and functions were mapped to keys with consistency. Cstar also supports on-line phrase creation, as mentioned in the discussion of the qualitative results, which increased its adaptiveness. Nokia provided a solution for off-line phrase creation, but it was hidden from users. Although the Nokia pinyin phrasal input method supported all kinds of phrases well (two-character phrases, three-character phrases, idioms, special names, and so on), it probably could still not compensate for the lack of on-line phrase creation.

Guobi was subjectively preferred by the participants. However, it did not perform so well by the other three metrics: it came fourth for KSPC, third in terms of the text entry rate, and fifth in relation to error rate. These results may show us that the participants in fact preferred phrasal input methods of a type that does not involve the pinyin selection process. Guobi was a method with both strong advantages and serious disadvantages. The participants made more errors with Guobi, which might be because there were too many shortcuts or, in other words, expert features in the interaction process. The overloading of functions on phone keys resulted in confusion and a lot of errors for first-time users, although trained users achieved better performance.

The T9 phrasal input came out in the middle among all methods in terms of performance results. In the experiment, users made frequent errors in switching mode. When users wanted to change the language mode from pinyin to English with T9, they needed to make a long press of the key three times and short-click it once. This resulted in user failure in the mode switching task.

As the results showed, Zi did not perform well. This result might be partly because the keys, especially the five-way navigation key of the E70, were not easy to operate. I observed that the confirmation click was easily detected as a scroll-down click and vice versa. The system of the E70 also reacted slowly, which resulted in frequent passing beyond the target candidate, negatively influencing user performance.

Based on the results I obtained in the study, I suggest the following design guidelines for the Chinese pinyin phrasal input methods for the 12-key keypad:

- Maintain consistency in UI design and find a balance between radical new designs and well-accepted ones. To make the user interface of a pinyin phrasal input method easy to learn and use, designers should make the best use of users' existing knowledge of phone interaction and keep their designs consistent with the most well-accepted key mappings. Some radical new designs are needed for expert features, but designers need to be careful to balance the radical elements and the commonplace ones such that they can coexist harmoniously, not interfering with each other.

- Optimize the corpus constantly. Users – especially novice users – find it difficult to guess what phrases are supported by a phrasal input method. An optimized corpus would increase both user performance and satisfaction. Additionally, with the rapid development of Chinese society, the language itself and how people use it also change constantly. To be adaptive to the changes, regular updating and optimizing of the corpus are important.

- Provide easy-to-use on-line phrase creation. Enabling on-line addition of new phrases to the corpus is necessary for phrasal input methods on mobile phones, and it is critical for avoiding failure in entry of phrases. The Cstar method provides a good interaction solution for this function. When users enter a phrase that is not supported by the method, Cstar allows moving the cursors in the pinyin marks entered and choosing characters one by one for the phrase. The interaction process would trigger the system's memorization of the phrase; then, when users enter it again, they could get it automatically.

- Omitting the pinyin selection process may be useful for phrasal pinyin input methods. The results showed that once a phrasal input method is carefully designed, it can result in good user performance and user satisfaction no matter the category of the method, but the participants seem to prefer the type of method requiring no selection of pinyin marks (here Cstar and Guobi) rather than the type requiring selection of the correct pinyin marks (here Nokia, T9, and Zi). The former class of method does require greater computing capabilities, less of a bottleneck for mobile devices as time passes. It is believed that omitting the pinyin selection process will be a trend for phrasal input methods based on the 12-key keypad.

6.3 Predicting Pinyin Input Speeds for Mobile Phones

6.3.1 Objective

In this section, I present a predictive model that estimates the average text entry speeds of users' error-free Chinese pinyin input on mobile phones. Two facts motivated me to build the model. First, relative to the large number of Chinese-speakers around the world, published studies of the user performance of mobile Chinese pinyin input were relatively few in number, and the results of this limited number of studies often contradicted each other. Lin and Sears (2005) reported a text entry speed of 5.46 CCPM when the participants naturally balanced between input speed and error rate. Liu and Wang reported user speeds of Chinese pinyin input supporting phrases between 15 and 34 CCPM (Liu and Wang, 2007). Second, there was no predictive model built to estimate user performance or understand the mechanisms involved in Chinese text entry tasks on mobile phones. A predictive model can assist researchers and practitioners in comparing text entry solutions without the need for tedious empirical studies and in identifying the areas for improvement and design opportunities (Card, Moran, & Newell, 1980, 1983; Soukoreff & MacKenzie, 1995; Silfverberg, MacKenzie, & Kohornen, 2000; Cockburn, Gutwin, & Greenberg, 2007; Cockburn & Gutwin, 2009; Dunlop & Crossan, 2000; Dunlop & Masters, 2008; James & Reischel, 2001; MacKenzie & Soukoreff, 2002; Zhai, Smith, & Hunter, 2002; Zhai, Sue, & Accot, 2002; Myung, 2004; Isokoski, 2004).

The rest of this section is organized as follows: first, I explain the input process for pinyin input on mobile phones; second, I explain the model and its core elements; third, I present three experiments to set parameters for the model and compare its predictions with empirical user speeds; and fourth, I present and discuss the results. Finally, I draw conclusions.

6.3.2 Models for Sub-tasks of Chinese Text Entry

According to the task analysis in the previous section, there are two sub-tasks involved in Chinese pinyin input on mobile phones: the retrieval and typing of a pinyin mark and the disambiguation task for selecting the target pinyin mark and Chinese character. Accordingly, I built the predictive model as expressed below:

$$T = T_m + T_d \tag{6-4}$$

Here, T represents the average time required to enter a Chinese character and equals the sum of T_m and T_d; T_m is the average time required for retrieving and typing a pinyin mark, which is represented by the average motor movement time required to type a pinyin mark (I assume that the cognitive retrieval of a pinyin mark would take little time for expert users); and T_d is the average time spent selecting the target pinyin mark and Chinese character.

There are two strategies that users can apply in the disambiguation process. Users can visually identify the target in the list of options first and then move the highlighting to it by pressing navigation keys, or they can go through the items one by one, both visually checking whether an item is the target and moving the cursor to it as well. No matter which strategy the users will apply, the two processes can be analyzed separately, since they consist of the same elements, only in different order. Moreover, users can be expected to start with the first items, since these are more likely to be the intended ones than the last ones are. Often it is not necessary to search the entire list before the desired mark or character is found. So I split the disambiguation process into two sub-processes: visual search and navigation. Consequently, T_d includes times for visual search (T_v) and navigation (T_n):

$$T_d = T_v + T_n \tag{6-5}$$

For Chinese characters entered with only the predictive feature, the average time for entering a character consists of T_d alone.

On the basis of this model, I applied several theories to predict the times for the sub-tasks. These theories are presented in the next three subsections. First, I present the corpus used for building the language model. Second, I present the movement model by combining Fitts' law and the language model (T_m). Then, I describe the keystroke-level model (KLM) developed to predict the average navigation time (T_n). Finally, I present the linear model to estimate the average visual search time (T_v).

Table 6.4: The digraph frequencies for the pinyin coding system based on the Chinese SMS corpus

First	Second Letter																											
Letter	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**	**Alone**	**Total**
A	-	-	-	-	-	-	-	-	186319	-	-	-	-	545926	414159	-	-	-	-	-	-	-	-	-	-	-	355794	1502198
B	84842	-	-	-	17267	-	-	-	31573	-	-	-	-	-	1637	-	-	-	-	-	100583	-	-	-	-	-	-	235902
C	12492	-	-	-	718	-	-	67381	4717	-	-	-	-	-	2973	-	-	-	-	-	4844	-	-	-	-	-	-	93125
D	88800	-	-	-	117034	-	-	-	55871	-	-	-	-	-	9693	-	-	-	-	-	51122	-	-	-	-	-	-	322520
E	-	-	-	-	-	-	-	-	112106	-	-	-	-	207008	-	-	-	18336	-	-	-	-	-	-	-	-	358384	695834
F	38064	-	-	-	11854	-	-	-	-	-	-	-	-	-	1388	-	-	-	-	-	9785	-	-	-	-	-	-	61091
G	40958	-	-	-	61700	-	-	-	-	-	-	-	-	-	15518	-	-	-	-	-	44879	-	-	-	-	-	389905	552960
H	140063	-	-	-	145117	-	-	-	168529	-	-	-	-	-	47800	-	-	-	-	-	226875	-	-	-	-	-	-	728384
I	419155	-	-	-	51513	-	-	-	-	-	-	-	-	156243	1124	-	-	-	-	-	54906	-	-	-	-	-	987643	1670584
J	-	-	-	-	-	-	-	-	185336	-	-	-	-	-	-	-	-	-	-	-	16034	-	-	-	-	-	-	201370
K	30118	-	-	-	22539	-	-	-	-	-	-	-	-	-	4546	-	-	-	-	-	17030	-	-	-	-	-	-	74233
L	56750	-	-	-	11469	-	-	-	174933	-	-	-	-	-	2890	-	-	-	-	-	6860	1169	-	-	-	-	-	254071
M	75422	-	-	-	68171	-	-	-	27254	-	-	-	-	-	2039	-	-	-	-	-	1978	-	-	-	-	-	-	174864
N	50566	-	-	-	38291	-	389905		191330	-	-	-	-	-	228	-	-	-	-	-	1453	3342	-	-	-	-	601666	1276781
O	-	-	-	-	-	-	-		-	-	-	-	-	65138	-	-	-	-	-	-	107634	-	-	-	-	-	752169	924941
P	8747	-	-	-	7791	-	-		9825	-	-	-	-	-	7842	-	-	-	-	-	757	-	-	-	-	-	-	34962
Q	-	-	-	-	-	-	-		61899	-	-	-	-	-	-	-	-	-	-	-	38236	-	-	-	-	-	-	100135
R	14257	-	-	-	24494	-	-		2661	-	-	-	-	-	2072	-	-	-	-	-	7709	-	-	-	-	-	18336	69529
S	9945	-	-	-	885	-	-	273531	14525	-	-	-	-	-	2958	-	-	-	-	-	16753	-	-	-	-	-	-	318597
T	39344	-	-	-	4712	-	-		59233	-	-	-	-	-	11506	-	-	-	-	-	3825	-	-	-	-	-	-	118620
U	121534	-	-	-	23088	-	-		89651	-	-	-	-	17256	111198	-	-	-	-	-	-	-	-	-	-	-	425012	787739
V	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4511	4511
W	40625	-	-	-	26362	-	-		-	-	-	-	-	-	213879	-	-	-	-	-	18930	-	-	-	-	-	-	299796
X	-	-	-	-	-	-	-		185387	-	-	-	-	-	-	-	-	-	-	-	12055	-	-	-	-	-	-	197442
Y	107410	-	-	-	24896	-	-		91259	-	-	-	-	-	62332	-	-	-	-	-	23007	-	-	-	-	-	-	308904
Z	77193	-	-	-	14182	-	-	128079	18176	-	-	-	-	-	7478	-	-	-	-	-	22484	-	-	-	-	-	-	267592
Total	1456285	-	-	-	672083	-	389905	468991	1670584	-	-	-	-	991571	923260	-	-	-	-	-	787739	4511	-	-	-	-	3893420	11276685

The Corpus and Language Model

In Section 2.2, I mentioned that we have collected a corpus of Chinese text messages for language models (Ma et al., 2007; Alhonen et al., 2007; Liu & Räihä, 2008, 2010). Based on the corpus, I built the language model required for building the predictive model for user speeds. First, I translated the Chinese characters to pinyin marks and analyzed the frequencies of all pinyin marks. From the frequencies of pinyin marks, I elicited the digraph probabilities for each legal pair of letters in pinyin. The language model resulted in a 26×26 matrix of letter pair frequencies (see Table 6.4). The 26 characters were the Roman letters A to Z. Each letter pair, *i–j*, has a probability P_{ij}. The sum of the probabilities of all letter pairs is 1.

To predict the average time for typing a pinyin mark, one still needs to know the average number of letters per Chinese character. According to Section 2.2, if the frequencies of the Chinese characters are taken into account, the average number of letter per Chinese character is 2.88.

Most Chinese characters are homophonic with other characters (Liu & Räihä, 2008; Liu & Wang, 2007; Sacher, Tng, & Loudon, 2001); thus, a pinyin mark usually corresponds to multiple Chinese characters. In most pinyin input systems, the character options are listed according to their usage frequencies: characters with higher frequencies are listed before characters with lower frequencies. Accordingly, I calculated the average position of all characters with the same pronunciations. The result was 1.77. The 12-key keypad can also produce ambiguous results. A series of key presses sometimes results in multiple pinyin marks, which are listed according to their usage frequencies. I also calculated the average position for all pinyin marks. The result was 1.24.

According to Section 2.2, characters that could be entered via the prediction feature accounted for about 30.3% of all characters in the corpus. Since the predicted characters are also listed according to their usage frequency, the average position calculated for all predicted characters was 2.60.

The Movement Model

The movement model was built on the basis of a combination of Fitts' law and the language model. Inspired by information theory, Fitts' law was applied by psychologists to predict the movement time of the human motor system (Seow, 2005; Soukoreff & MacKenzie, 2004). According to Fitts' law, the time for people to move from one object to another is a logarithmic function of the distance between the two objects divided by the size of the target object. Fitts' law is usually expressed as follows:

$$MT = a + b \log_2 (A/W+1) \tag{6-6}$$

In the above equation, A represents the amplitude of motor movement and W is the size of the target object, which is usually indicated by its width (Seow, 2005; Soukoreff & MacKenzie, 2004) or for two-dimensional objects, height or width, whichever is smaller (MacKenzie & Buxton, 1992; Silfverberg, MacKenzie, & Korhonen, 2000). Constants a and b are defined by empirical experiments. The log term of Equation 6-6 is also called the index of difficulty (ID).

Fitts' law has been widely applied in prediction of expert user performance with different types of input devices including physical (Myung, 2004; Silfverberg, MacKenzie, & Korhonen, 2000) and soft keyboards (MacKenzie & Soukoreff, 2002; Zhai, Smith, & Hunter, 2002; Zhai, Sue, & Accot, 2002; Isokoski, 2004; Soukoreff & MacKenzie, 2004), pointing devices (Soukoreff & MacKenzie, 2004; Zhai, 2004) and tasks including text entry (Isokoski, 2004; Isokoski & MacKenzie, 2003; James & Reischel, 2001; MacKenzie & Soukoreff, 2002; Myung, 2004; Pavlovych & Stuerzlinger, 2004; Silfverberg, MacKenzie, & Korhonen, 2000; Soukoreff & MacKenzie, 2004) and menu selection (Cockburn, Gutwin, & Greenberg, 2007; Cockburn & Gutwin, 2009). In application of Fitts' law to predict user performance with different tasks, a basic assumption is that the users are experts in the tasks, for whom motor movements cost the majority of the effort and time while cognitive processes cost little.

I combined Fitts' law and the language model to estimate the average movement time for a pinyin mark as follows:

$$T_m = 2.88 \times \sum\sum (P_{ij} \times MT_{ij}) \tag{6-7}$$

where 2.88 is the average number of letters that a pinyin mark includes; P_{ij} is the probability of the letter pair, i–j, being entered together; and MT_{ij} is the average time needed to move the thumb from "i" to the "j" key and press it.

KLM for Navigation Time in the Disambiguation Process

I applied KLM, the simplest GOMS model (Card, Moran, & Newell, 1980, 1983), to predict navigation time T_n in the disambiguation process. Originally, KLM defined six operators: K for key presses, P for pointing to an object on the display with a mouse, H for moving the hands to the home position on the keyboard or mouse, D to draw a line, M for mentally preparing for an action or closing associated primitive actions, and R as the system response time for which users must wait. By splitting a specific task into such operators and defining times for relevant operators, analysts can estimate the time a skilled user needs to complete the task without error.

Dunlop and Crossan applied KLM to compare user performance in the multi-press method with that in the predictive methods (Dunlop &

Crossan, 2000; Dunlop & Masters, 2008). In their KLM, they chose three operators and defined fixed times for them: K for a button press (280 ms), H to move the hand to the home button (400 ms), and M for mental preparation time for executing physical actions (1350 ms).

When I predicted the navigation time in the disambiguation process, I simply chose operator K. Moreover, instead of using a fixed time for operator K, I applied Fitts' law to estimate times for different key presses.

If we assume P_{mn} is the overall probability of Chinese characters in the corpus whose target pinyin marks and character symbols are located at positions m and n, respectively, T_n can be accurately calculated as follows:

$$T_n = \sum\sum (P_{mn} \times T_{n\_mn}) \qquad (6\text{-}8)$$

Here, T_{n_mn} is the specific navigation time based on KLM for cases where the target pinyin marks and Chinese characters are located at positions m and n, respectively However, such a calculation of T_n is rather complex and hard to apply. First, according to the corpus, there are, in total, 263 combinations of m and n. Second, for each combination of m and n, the calculation of T_{n_mn} is different from the others. For example, when $m = 1$ and $n = 1$, T_{n_mn} should be calculated as:

$$T_{n\_mn} = K_{io} + K_{rr} \qquad (6\text{-}9)$$

And when $m = 2$ and $n = 2$, T_{n_mn} should be calculated as:

$$T_{n\_mn} = K_{ir} + 2K_{ro} + K_{or} \qquad (6\text{-}10)$$

In the above equations, K_{io} and K_{ir} are, respectively, the average time to move the thumb from the last letter of the pinyin marks to the "OK" or the right navigation key and press. Similarly, K_{rr} is the time required for repeated key presses (the value of K_{rr} equals a in Fitts' law when ID equals 0), K_{or} is the time to move the thumb from the "OK" key to the right navigation key and press it, and K_{ro} is the time to move the thumb from the right navigation key to the "OK" key and press it. Moreover, K_{io} and K_{ir} can be calculated as shown in Equation 6-11, where $P_{io/r}$ is the overall probability of the "i" key being the last letter of the pinyin marks and $MT_{io/r}$ is the time for moving the thumb from the "i" key to the "OK" or right navigation key.

$$K_{io/r} = \sum\sum (P_{io/r} \times MT_{io/r}) \qquad (6\text{-}11)$$

To make the calculation of T_n simpler and easier to apply, I approximated T_n with $T_{n\_1,2}$ (i.e., by setting m and n to 1 and 2, respectively, because, according to the corpus analysis, the average positions for pinyin marks and characters are 1.24 and 1.77, respectively – close to 1 and 2. Figure 6.7 shows the navigation process with a target pinyin mark in the first position and a Chinese character in the second position. With this simplification, T_n can be expressed as:

$$T_n = K_{io} + K_{or} + K_{ro} \qquad (6\text{-}12)$$

As I mentioned earlier, some characters can be entered by choosing them from lists of predictions. For those characters, T_{n_pre} can be similarly expressed as:

$$T_{n\_pre} = K_{or} + (2.6\text{-}2) \times K_{rr} + K_{ro} \qquad (6\text{-}13)$$

Since the average position for the predicted characters is 2.6, the time for pressing the right navigation key repeatedly is calculated as $(2.6 - 2) \times K_{rr}$.

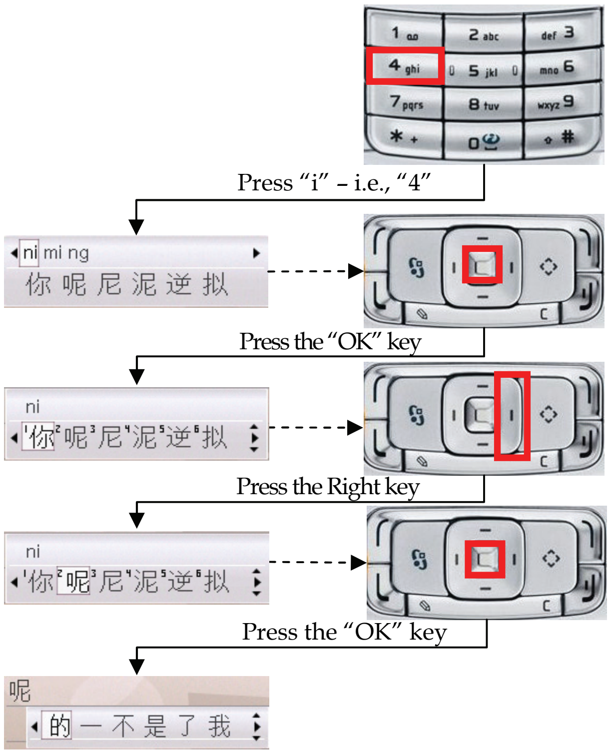


Figure 6.7. The disambiguation process when m = 1 and n = 2.

The Visual Search Model

Visual search is the process of finding a target item from among distractor items (Wolfe, 1998). Selections of a target pinyin mark and a target Chinese character in the pinyin input process include such visual search processes.

Psychologists regard visual search as a basic process of human cognition, and many studies have been done to understand its mechanisms (Wolfe, 1998). A basic experiment paradigm applied in such studies is to change the set size – i.e., the total number of stimuli – and collect reaction times when participants are instructed to identify a target item and respond. In half of the trials, the target item is present and in the other half, the target is not included. Psychologists divide visual searches into parallel searches and serial searches according to the slope of reaction times plotted against set size. If the slope is close to 0 ms per item, the case is usually regarded as one of parallel search. However, if the slope is steep, it is considered to involve serial search.

In the field of human–computer interaction, the Hick-Hyman law has been applied to estimate visual search time (Cockburn, Gutwin, & Greenberg, 2007; Cockburn & Gutwin, 2009; Soukoreff & MacKenzie, 1995). Similarly to Fitts' law, the Hick-Hyman law was inspired by information theory. It has been applied to estimate the reaction time for making a choice from a number of possibilities. However, in comparison to Fitts' law, it has not been as widely applied in HCI (Seow, 2005). It is usually expressed as follows:

$$RT = c + d \log_2 (n + 1) \tag{6-14}$$

Here, RT is the decision-making time for choosing a response from a number of possibilities in accordance with a presented stimulus; it is a logarithmic function of the number of possibilities. In Equation 11, n is the number of possibilities, and c and d are constants defined by empirical studies.

Soukoreff and MacKenzie (1995) applied the Hick-Hyman law to estimate visual search time with soft keyboards. However, Sears et al. (2001) argued that the Hick-Hyman law was not suitable for predicting time for visual search that is "scan-and-match." They also argued that more factors, such as familiarity, should be taken into account for prediction of user performance of visual search. Cockburn, Gutwin, and Greenberg (2007) argued that when people can anticipate the location of items, the Hick-Hyman law is appropriate for predicting the time for acquiring the target but a linear model should be applied when anticipation is not possible.

I applied the linear model instead of the Hick-Hyman law to estimate visual search times in the pinyin input process. There were two reasons behind this decision. First, people are not able to anticipate locations for both pinyin marks and Chinese characters, so the linear model may be more appropriate. Second, both pinyin marks and Chinese characters are complex units of information, and visual search of them may be characteristic of a serial search. A linear model is appropriate for a serial search. Thus T_v can be expressed as follows:

$$T_v = \sum\{P_m \times [e_{py} + f_{py} \times (m-1)]\} + \sum\{P_n \times [g_{cc} + h_{cc} \times (n-1)]\} \tag{6-15}$$

In the above equation, e_{py}, f_{py}, g_{cc}, and h_{cc} are constants defined by Experiment 2; m and n are, respectively, the positions for the target pinyin marks and Chinese characters; P_m is the overall probability of the target pinyin mark being at position m; and P_n is the overall probability of the target Chinese character being at position n. Equation 12 can be simplified as follows:

$$T_v = e_{py}+f_{py}\times[\sum (P_m\times m)-1)]+g_{cc}+h_{cc}\times[\sum(P_n\times n) -1)] \qquad (6\text{-}16)$$

In the above equation, the terms $\sum(P_m \times m)$ and $\sum(P_n \times n)$ provide our formulae for calculating the average positions for pinyin marks and Chinese characters, which equal 1.24 and 1.77, respectively.

For Chinese characters that can be entered by choosing them from predicted options, visual search time included only the part for Chinese characters in Equation 13.

6.3.3 Experiment 1: Setting of the Parameters for Fitts' Law

The first experiment was done to define parameters for Fitts' law.

Participants

Twelve volunteers (seven male, five female) took part in the experiment. Their ages ranged from 22 to 34 years, with an average of 27.3 years (SD = 3.78). All participants were either student interns or researchers at Nokia Research Center in Beijing. All were right-handed and held the phone in the dominant hand in the experiment. All were regular phone-users.

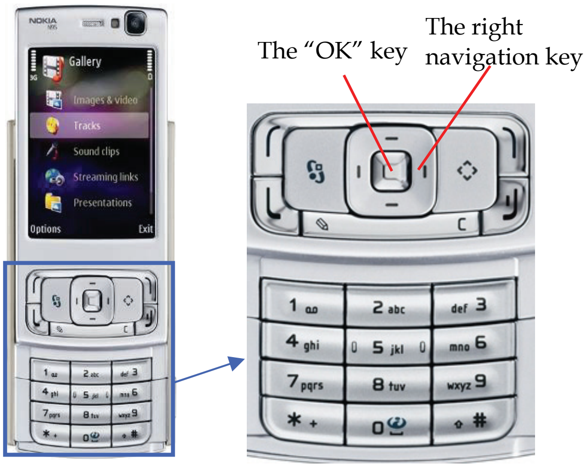


Figure 6.8. The Nokia N95 used in the experiment.

Apparatus

A Nokia N95 (see Figure 6.8) was the device used in the experiment. For the experiment, I built a program that could automatically log the time for each key press. The look and feel of the program was the same as in the device's short message application. I chose the Nokia N95 because it has a high-speed processor and can ensure the accuracy of the time logs.

Test Tasks

There were two types of tasks that users needed to complete with the thumb of their dominant hand:

i) Repeated key presses: participants were to press a key twice, as quickly as they could. In the experiment, participants were asked to press the right navigation key repeatedly and complete the tasks four times, so I collected 48 data points.

ii) Paired key presses: the participants were instructed to press two specified keys consecutively, as quickly as they could. The task included 12 pairs of keys, with distances ranging from 5.76 mm ("6"−"9") to 28.98 mm ("9"−"1"). The heights of the keys that serve as *W* in Fitts' law for calculation of the IDs range from 5.45 mm to 6.28 mm. Every participant needed to press a pair of keys four times, so, in total, I collected 48 data points for each pair of keys.

For both types of tasks, participants were to press the "OK" key before and after each pair of key presses, to indicate the start and end of a trial. Thus for each trial, four times were logged and the time for motor movement and the key press equaled the difference between the third logged time and the second logged time.

Procedure

The experiment was conducted in a lab with a coordinator and a participant present. Before data collection, the coordinator explained the objectives and tasks to the participant. Then the participant practiced freely until he or she was ready to start the data collection phase. Half of the participants started with the repeated key press tasks, and the other half started with the paired key press tasks. For the paired key press tasks, testing orders were counterbalanced among the 12 participants with the Latin-square technique.

In the data collection phase, participants were instructed to complete all tasks. The time for each key press was automatically logged. After the experiment, participants were presented with a small gift.

Results

Table 6.5 and Figure 6.9 show the results of Experiment 1. In Figure 6.9, the diamond marks indicate the average reaction times for different IDs and the line is the linear regression that I performed for the average reaction time by ID. The average reaction times increased when the ID rose. An ANOVA test indicated a significant effect of ID on reaction times ($F_{12,155} = 33.63$, $p < .001$).

Table 6.5: Parameters defined ("a" and "b") for Fitts' law

Intercept, a (ms)	Slope, b (ms/bit)	Correlation
195	101	0.992

Table 6.5 shows the constants for Fitts' law; the correlation of the linear regression was very high, showing that Fitts' law is a very good predictor of motor movement time.

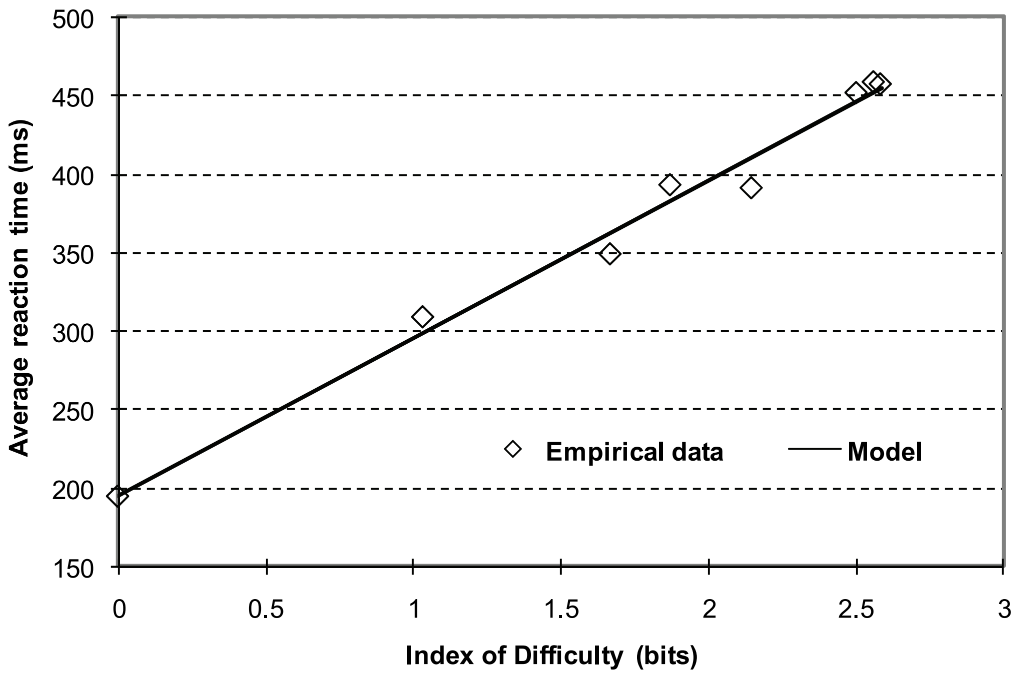


Figure 6.9. Result of Experiment 1: regression of reaction time by ID (R^2 = 0.985).

From the results, I calculated T_m according to Equation 6-7, K_{io} from Equation 6-11, and K_{or} and K_{ro}, which, in that order, equaled 1059 ms, 456 ms, 321 ms, and 285 ms. On the basis of Equation 6-12, I calculated the average navigation time for entering a predicted character, which equaled 723 ms.

6.3.5 Experiment 2: Setting of Parameters for the Linear Model of Visual Search

Experiment 2 was designed to determine parameters for the linear visual search models (e_{py}, f_{py}, g_{cc}, and h_{cc} in Equation 6-15) for pinyin marks and Chinese characters.

Participants

Twenty-four volunteers (16 male and eight female) took part in the experiment. Their ages ranged from 22 to 32 years, with an average of 26.2 years (SD = 3.63). All were either student interns or Nokia Research Center staff members in Beijing. Two were left-handed, and the others were right-handed. In the experiment, the participants held the phone with both hands, with the left thumb on the "1" key and the right thumb on the "3" key.

Apparatus

A Nokia N95 was the device used in the study. A program was designed for logging key press times and to lead the participants through the experiment (see Figure 6.10).

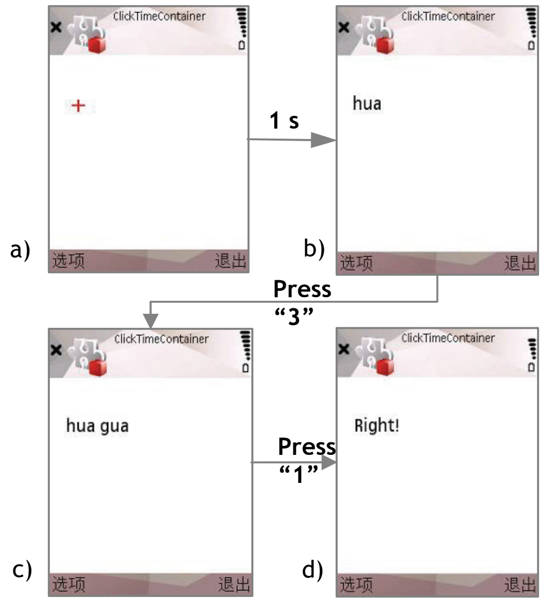


Figure 6.10. Screenshots for Experiment 2: a) at the beginning of a trial, a red cross was displayed for a second to gain the user's attention; b) a target item was displayed in the same location, and participants were to remember it and press "3" to proceed to the next step; c) a list of options appeared for users' indication of whether the target had appeared or not (target present: press "1," target absent: press "3"); and d) after participants responded, the system gave feedback.

Tasks and Materials

I applied the experiment paradigm used by psychologists to study the visual search process (Wolfe, 1998). At the beginning of a trial, a red cross was displayed for a second to gain the attention of the user. Then a target item (either a pinyin mark or a Chinese character) appeared at the same location. Once participants had recognized and remembered the target, they were to press the "3" key. After the key press, a list of options (2–4 for pinyin marks and 6 for Chinese characters) was presented either horizontally or vertically and participants were required to decide whether the target was among them or not and respond by pressing a corresponding key as accurately and quickly as they could. To avoid cheating, half of the trials included the target and half did not. If participants decided that a target was in the list, they needed to press the "1" key; otherwise, they needed to press the "3" key. User reaction times were automatically logged by the software.

There were three types of trials in the experiment, with pinyin marks listed either vertically or horizontally and Chinese characters listed horizontally. In the experiment, pinyin marks were listed both horizontally and vertically because both solutions exist in pinyin input solutions and there is no existing study exploring the differences in user performance. By contrast, Chinese characters were listed just horizontally since studies have already proven that people are more efficient at searching for Chinese characters in horizontal lists than in vertical lists (Heloisa & Ravindra, 1998) and, moreover, Chinese characters are listed horizontally in many products' pinyin input solutions.

The pinyin marks and Chinese characters used in the experiment were selected carefully. The 133 groups of pinyin marks in which between two and four pinyin marks shared the same series of key presses were all covered in both horizontal pinyin trials and vertical pinyin trials. Thus there were 266 trials for pinyin marks in the experiment, half with targets presented and half without. I chose 96 Chinese characters from among the 500 most frequently used ones as the target items. For each target item, I specified five (for target-present trials) or six (for non-present target trials) other Chinese characters with the same pronunciation as options. There were, in total, 96 trials for Chinese characters, so the experiment included 362 trials. For all trials, all variables, including whether a target was absent or presented, the target item itself, its location in the option list if presented, and the order of trials, were randomly arranged without replacements.

Procedure

The procedure in this experiment was the same as that for Experiment 1 except that the training session of Experiment 2 included 30 trials.

Results

Participants made few errors in the experiment, with an average error rate of 1.4% (*SD* = 0.92%). In my analysis of the data, all error trials were excluded. Table 6.6 and Figure 6.11 show the results of Experiment 2, which indicate that linear models are appropriate for modeling the visual search task in Experiment 2 (Soukoreff & MacKenzie, 2004). Moreover, the visual search in pinyin marks and Chinese characters is characteristic of both serial and self-terminating searches.

An ANOVA analysis was conducted to explore the effects of trial type (pinyin marks listed both horizontally and vertically and Chinese characters) and target location (first four positions) on the average visual search time of the 24 participants. The results indicated significant effects of target position ($F_{3,287} = 71.8$, $p < .001$) and type of trial ($F_{2,287} = 7.9$, $p < .001$) on response times, as well as their interaction ($F_{6,287} = 3.9$, $p < .001$). Further t-tests indicated that searching among pinyin marks, no matter whether they were listed horizontally ($t = 2.5$, $p < .05$) or vertically ($t = 2.8$, $p < .05$), required longer response times than searching among

Chinese characters. There was no significant difference between the two listing types for pinyin marks on response times (t = 0.446, ns).

Table 6.6: The linear models from Experiment 2

	Intercept (ms)	Slope (ms/item)	Correlation
Horizontal pinyin	644	153	0.999
Vertical pinyin	645	160	0.976
Chinese characters	704	62	0.968

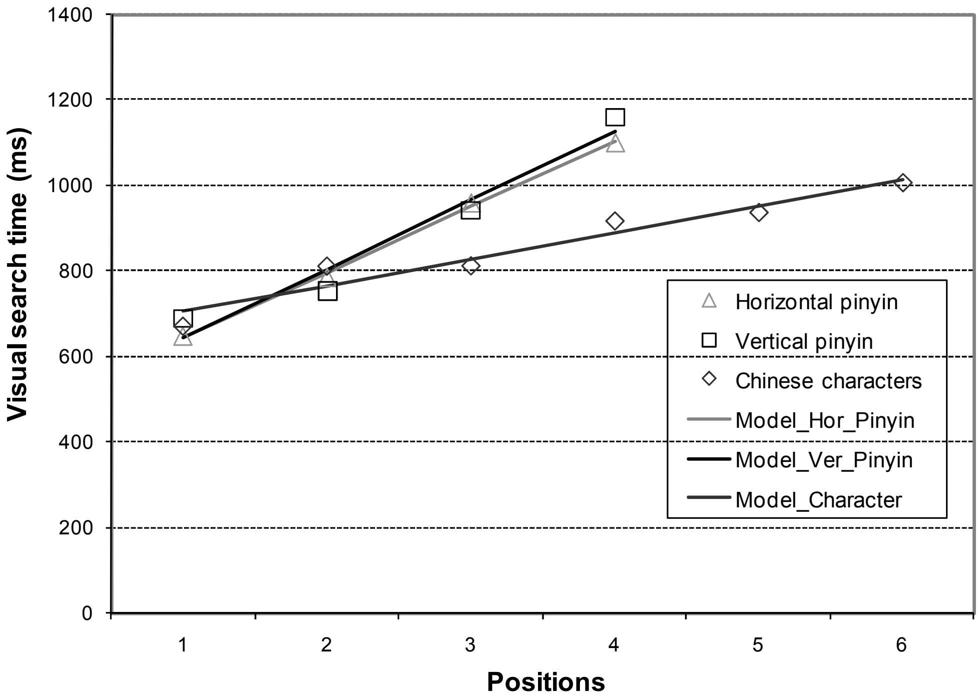


Figure 6.11. The empirical results and the linear models for visual search.

When I estimated visual search times, I did not directly apply the results of Experiment 2. This was because the task of the experiment included a sub-process of choice reaction from two reaction options (or decision-making) besides visual search. Therefore, I decided to subtract a fixed time for the choice reaction from the linear models. According to Sears et al. (2001), the choice reaction time should be calculated on the basis of the number of possible reactions instead of the number of stimuli. I applied the results of Hick's experiment concerning choice reaction and subtracted a fixed time of 247 ms (n = 2) from the linear models (Hick, 1952; Seow, 2005). I applied Hick's results because the task in his experiment was closer. Finally, the constants e_{py}, f_{py}, g_{cc}, and h_{cc} were calculated. These are presented in Table 6.7.

Table 6.7: The parameters for visual search

e_{py} (ms)	f_{py} (ms/item)	g_{cc} (ms)	h_{cc} (ms/item)
397	153	457	62

Table 6.8: The average time required to enter a Chinese character

	T_m	T_n	T_v	T
Average time per character, predictive feature off (ms)	1059	1062	938	3059
Average time per character entered with predictive feature only (ms)		723	556	1279

Table 6.8 summarizes the average time for entering a Chinese character with the predictive feature off (3059 ms) and with the predictive feature only (1279 ms). These are represented, respectively, as T and T_{pre} in equations 6-17 and 6-18. Equations 6-17 (S_{pre_off}) and 6-18 (S_{pre_on}) show how I calculate the predicted speeds when the predictive feature is off and on, respectively. P_{pre} in Equation 6-18 is the overall percentage of characters entered via prediction only. Table 6.9 presents the predicted speeds.

$$S_{pre\_off} = 60/\ (T/1000) \tag{6-17}$$

$$S_{pre\_on} = 60/\ \{[T\times(1- P_{pre}) + T_{pre}\times P_{pre}]\ /1000\} \tag{6-18}$$

6.2.6 Experiment 3: Evaluation of the Model with Empirical User Speeds

Experiment 3 was carried out to collect empirical data on users' error-free speeds, for evaluation of the predicted text entry speeds.

Participants

Twelve volunteers (eight male, four female) took part in the experiment. Their ages ranged from 24 to 32, with an average of 27.9 (SD = 3.06). All were either student interns or staff members at the Nokia Research Center in Beijing. One was left-handed, and the others were right-handed. All were users of pinyin text entry solutions on mobile phones, and they had used mobile phones for, on average, 5.3 years (SD = 1.21).

Apparatus

The Nokia N95 was the device used in the study. I used the program designed for Experiment 1 to collect data.

Test Tasks

Participants were instructed to enter two short messages twice, once character by character and the other time applying the phrase-based predictive input feature. Task orders were balanced among the 12 participants. There were, in total, 31 characters to be entered by participants and about seven characters (about 23%) that could be entered with the prediction feature. The single-letter correlation (MacKenzie & Soukoreff, 2003) with the corpus for the two text messages was 0.932. The average number of letters for the 31 characters was 2.94, and the average positions for pinyin marks and Chinese characters, respectively, were 1.10 and 1.84.

Procedure

The procedure in this experiment was the same with that in experiments 1 and 2 except that in the training session for this experiment, the participants were instructed to enter three short messages, consisting of 39 characters each.

Results

The time of all extra key presses – for example, for making and clearing errors – was fully removed from the task completion time, to make sure the text entry rates calculated are comparable with the predicted ones.

Table 6.9: Average text entry speeds in CCPM

<table>
<tr><th colspan="2"></th><th>Average user speeds, with SD</th><th>Predicted speeds</th></tr>
<tr><td colspan="2">Predictive feature off</td><td>19.1 (2.32)</td><td>19.6</td></tr>
<tr><td rowspan="2">Predictive feature on, with the percentages of characters entered with it</td><td>23.0%</td><td>21.9 (2.28)</td><td>22.6</td></tr>
<tr><td>30.3%</td><td>----</td><td>23.8</td></tr>
</table>

Table 6.9 and Figure 6.12 show the results of Experiment 3 and the predicted speeds. According to Table 6.9, the average user speed was 19.1 CCPM ($SD = 2.32$) when the predictive feature was off. When the predictive feature was on, the average user speed was 21.9 CCPM ($SD = 2.28$). The empirical data match the predictions well. The percentage differences in the predicted speeds were both within 4% of the empirical data, at 2.6% and 3.2%, while the predictive feature was, respectively, off and on. A paired t-test indicated that user speeds were significantly higher when the predictive feature was on ($t = 2.95$, $p < .05$). The percentage increases of speeds were correspondingly 14.7% from empirical data and 15.3% from the predictive model.

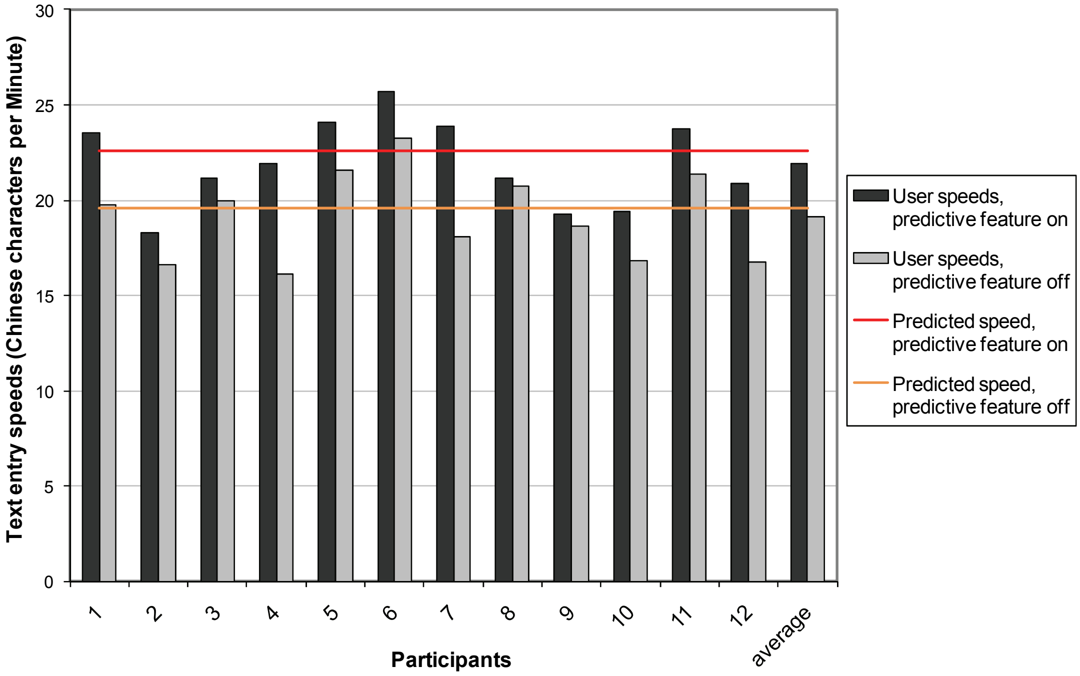


Figure 6.12. Text entry speeds without errors.

6.2.7 Experiment 4: Evaluation of the Model with Empirical Movement and Key Press Time

In the study described above, the predictive model was evaluated via comparison of its predictions with the empirical results for users' error-free text entry rates. However, further validation of the model may still be needed at sub-process or even stroke level. Experiment 4, along with the next experiment described, was conducted to validate the model again on the stroke level.

The objective of this experiment was to check the validation of the movement model estimated via Fitts' law. I designed another approach to collect the times required for some single key clicks in the disambiguation process and compared these with the results predicted by Fitts' law.

Participants

Twelve volunteers (half male, half female) took part in the experiment. They ranged from 24 to 32 years of age, with an average age of 26.7 years (*SD* = 2.89). All were either student interns or staff members at the Nokia Research Center in Beijing. All participants were right-handed. They were instructed to hold the phone in their dominant hand and operate it with the thumb.

Apparatus

The device used in this experiment was still a Nokia N95. A program was designed specifically for this experiment, to lead the participants to complete the tasks. The program also logged the key press times.

Test Tasks

The task involved scrolling and target selection by pressing relevant keys. Figure 6.13 shows the process and associated screenshots. With the software, participants can press any key to activate a frame where seven numbers (from 0 to 6) are shown with a target item in red. First, participants needed to press the "OK" key to highlight the first item, as shown by Figure 6.13 (c). Second, participants moved the highlighting to the target item by pressing the right navigation key. Third, participants were to select the target item by pressing the "OK" key. Each participant needed to complete 105 trials in the experiment, with the target locations randomly arranged. For each location, 180 data points in total were collected. If a participant made an error, the trial would present the item again, later in the experiment, until the participant was correct in all trials.

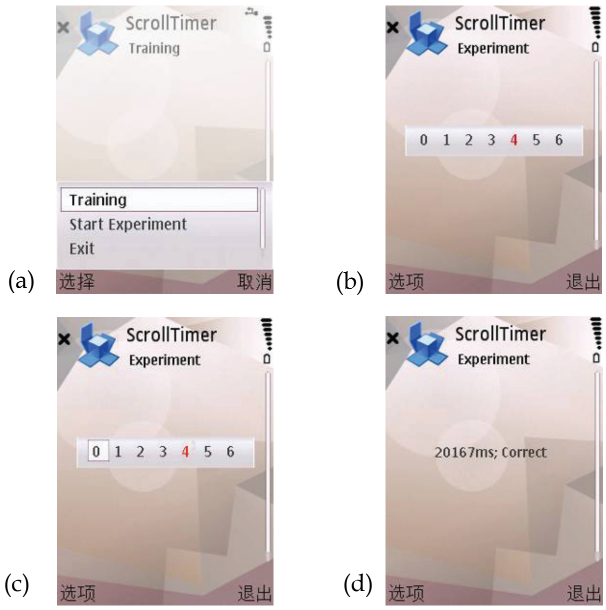


Figure 6.13. The software used in this experiment: a) an option menu is used to choose either training or data collection sessions; b) the frame shows the numbers from "0" to "6," with the target number in red; c) clicking the "OK" key will highlight "0" in the list, and users need to move the cursor to the "4" in red and select it by pressing the "OK" key; and d) the software will show results, including the response time and whether the response is correct.

The task is designed to mimic the disambiguation process, but it is much simpler, since numbers replaced Chinese characters or pinyin marks and the target item was shown in a different color. Moreover, the target item actually indicated the number of times to press the right navigation key for highlighting the target. For example, if the target item was 4, participants needed to press the right navigation key four times to acquire the target. Thus the task required little use of cognitive resources for either visual search or navigation coordination and is appropriate for use to collect key press times in the disambiguation process.

Procedure

The procedure in this experiment was the same as those of experiments 1, 2, and 3 except that the training session in this experiment included 50 trials.

Results

The average error rate of the 12 participants was 1.3%. Table 6.10 shows the average movement times obtained from this experiment and predicted by Fitts' law. The predicted results were all within 7% of the empirical results, which again validates the prediction by Fitts' law.

Table 6.10: Results of this experiment and the results predicted by Fitts' law

	Time required for repeated key presses	Time for movement from "OK" key to right navigation key and for pressing it	Time for movement from the right navigation key to the "OK" key and for pressing it
Average time, with *SD*	199 (20.8)	333 (54.0)	305 (43.3)
Results predicted by Fitts' law	195	321	285

6.2.8 Experiment 5: Evaluation of the Model with Empirical Disambiguation Time

Experiment 5 was carried out to evaluate the predictive model by comparing the empirical and predicted time for the disambiguation process. The empirical results for time required by the disambiguation process also show what the selection process looks like.

Participants

Sixteen volunteers from the Nokia Research Center Beijing office took part in the experiment. Their ages ranged from 22 years to 34 years, with an average of 26.7 years (*SD* = 3.38). All of them were users of SMS and daily users of the pinyin input method.

Apparatus

The device used in this experiment too was the Nokia N95. The software used in Experiment 1 was applied to collect data concerning key presses.

Test Tasks

The tasks for participants to complete in the experiments were to copy the Chinese text presented, with a Nokia N95. The text given was carefully designed such that characters in the text appeared in all positions, from the first to the sixth, after their pinyin marks were selected, so that enough data points for T_d could be collected for all positions in the disambiguation process.

Participants needed to enter, in total, 347 Chinese characters with the Nokia N95. The pieces of text presented were divided across four sessions, and participants were required to take a 15-minute break between any two sessions.

Procedure

The procedure was the same as in experiments 1, 2, 3, and 4, except that in the training session for this experiment, the participants were instructed to enter eight short messages, each consisting of 40 characters.

Results

Table 6.11 and Figure 6.14 show the results of this experiment and their comparison with the predictions as to the time required by the disambiguation process with pinyin character input methods.

Table 6.11: Results from this experiment as regards empirical time required by the disambiguation process and its comparison with the results predicted by the model

Time required for disambiguation process (ms)	Target positions					
	1st	2nd	3rd	4th	5th	6th
Empirical results, with *SD*	465 (291)	979 (511)	1391 (763)	1631 (704)	1784 (725)	1942 (714)
Predicted results	652	1125	1382	1639	1896	2152

I also conducted an analysis of correlation between the predicted and empirical results. The correlation coefficient between the two data sets is 0.986. The high correlation supports the validity of the predictive model.

On the other hand, although the trends between the predicted and empirical results are consistent, the differences between the two sets of data for first, second, fifth, and sixth positions were still between 100 ms and 210 ms. The results can be explained by the serial position effect, in which recall accuracy varies as a function of an item's position in a list. Both the first few items (primacy effect) and the last few (recentness effect) were recalled more accurately than items listed in the middle (Deese & Kaufman, 1957). In the disambiguation in the pinyin input process, the characters at the start or end of the option list were subject to more cognitive processing in the visual search process, so selection of them saved some time.

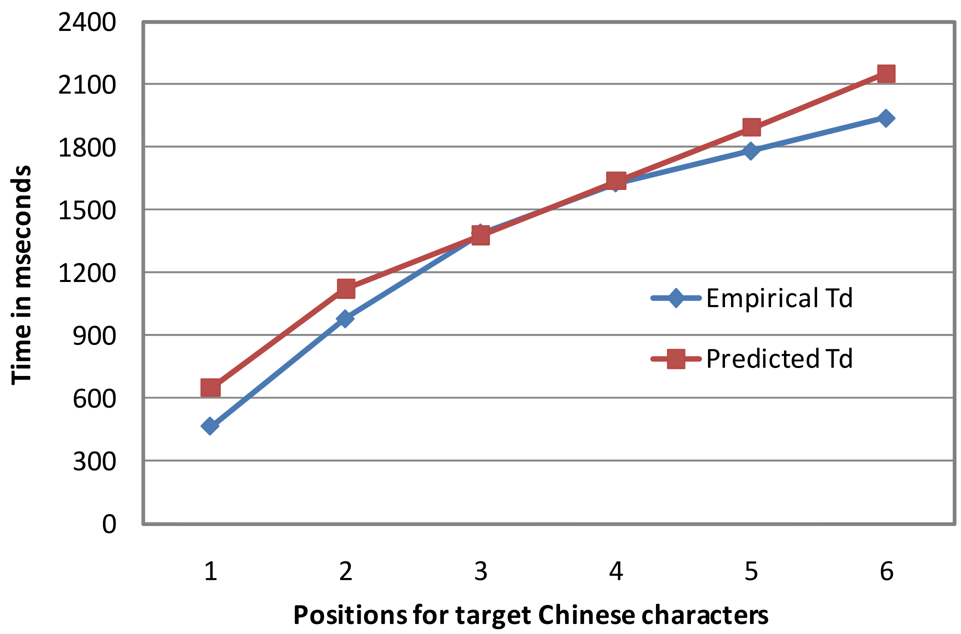


Figure 6.14. The results from this experiment on the empirical time required for the disambiguation process and its comparison with the results predicted by the model.

6.2.9 Discussion

The predictive model proved to be valid, as shown by the close match between the average user speeds and our predictions. On the other hand, the empirical results also indicated large individual differences among the participants in terms of text entry speeds. It is worth noting that these two facts are not in contradiction with each other. Since the constants in our models were calculated on the basis of average user performance with sub-tasks, the predicted speeds were already averages of user performance. Further evaluations of the model on the stroke level also indicated good matches of the predicted results and empirical results.

It was expected that the model, taking into consideration the characteristics of the corpus, would be applied in evaluations of Chinese text entry solutions for mobile phones; now I have completed the process with only one phone, and other phones would yield different values for the parameters. It is also possible to apply the corpus characteristics and the parameters separately.

According to our model, the disambiguation process required 69.2% (predictive feature on) or 65.4% (predictive feature off) of the total time with the pinyin input methods on mobile phones, indicating a greater percentage of the time than the disambiguation process of a pinyin input method based on a QWERTY keyboard takes (about 52%) (Wang, Zhai, & Su, 2001). Innovative solutions are needed to optimize the disambiguation process in Chinese text entry solutions for mobile devices.

When I calculated the predicted speed while the predictive feature is on as Equation 6-18 shows, I directly used for T the average time (3059 ms) calculated on the basis of all characters in the corpus. However, the characters that were entered via the predictive feature should have been removed from this number. I assume that this was one reason I found a slightly higher predicted speed for cases with the predictive feature. This is also one of the points for improvement that I need to address in future work.

In comparison with state-of-the-art predictive models for text entry and menu selection tasks (Cockburn, Gutwin, & Greenberg, 2007; Isokoski & MacKenzie, 2003; Pavlovych & Stuerzlinger, 2004), our model does not cover some variables, such as the learning process, or new features such as phrasal input. However, I view the present work as a good start and anticipate that such issues will be explored in the future.

6.4 SUMMARY

Chinese pinyin text entry solutions based on the 12-key keypad of mobile phones are critical since they are the primary methods adopted by local users. I presented two studies, focusing on pinyin phrasal input methods and pinyin character input methods, respectively.

I compared five Chinese pinyin phrasal input methods for mobile phones (Cstar, Guobi, Nokia, T9, and Zi), representing two types of interaction designs: methods requiring and not requiring selection of pinyin marks. The participants achieved better performance with the Cstar and Nokia methods. However, they also preferred Guobi subjectively, although it was not good by performance metrics. It is believed that the results indicate that omitting the selection process for pinyin marks is a trend for phrasal pinyin input methods that are based on the 12-key keypad. Some further design guidelines were proposed for the interaction design of pinyin phrasal input methods.

I presented a model that integrates a language model with Fitts' law for key presses, KLM for navigation, and a linear model for visual search to predict user performance with two Chinese pinyin input methods. I evaluated the model by comparing its predictions with the empirical user speeds. The predictions were proven effective: when the prediction feature was on and off, the predicted speeds were, respectively, 3.2% and 2.6% higher than the empirical user speeds. I view the model as a useful start. First, there is a lack of predictive models or insights to explore the patterns of user behavior in relation to the Chinese text entry systems of mobile phones. I hope that, as the first published model, our work can draw forth further work on this subject. Second, practitioners and researchers in this field can benefit from the model by applying it in evaluation of Chinese pinyin text entry solutions on mobile phones and identification of design opportunities.

7 Summary and Conclusions

Chinese text entry is a critical and urgent research field, in view of the widespread and frequent use of mobile devices by Chinese users. However, there have been few studies either to explore the design of new Chinese text entry methods for mobile devices or to understand user behavior in the interaction processes. This thesis has addressed both topics by presenting two design cases and several experimental studies focusing on user behavior in the interaction process.

Chinese is nearly unique in some of its characteristics: it is logosyllabic in nature; its characters are complex in structure and homophonic, with multiple tones. These characteristics significantly affect how Chinese text is entered in computing systems. First, Chinese characters cannot be directly entered with keyboards designed for alphabetic languages. Therefore, two coding systems were created to map the tens of thousands of Chinese characters to the 12-key keypad: the pinyin coding system, based on the Mandarin pronunciation of Chinese characters in the form of Latin letters, and the stroke coding system, based on the standard sequences for writing Chinese characters. Second, since the complex structure of Chinese characters carries rich information to differentiate them from each other, the Chinese handwriting recognition technique has attained a satisfactory accuracy level and been incorporated by many mobile products.

A few key goals motivate creation of new text entry solutions for mobile devices: optimal user performance (low error rates with high speeds), universal solutions covering more user groups and work contexts, ease of learning, and positive user experience. Many approaches have been explored and applied to reach these goals, although the resulting designs often end up with a tradeoff among the goals. Optimizing user performance with mobile text entry solutions primarily involves three types of approaches: the disambiguation approach (mainly for keypad-based text entry solutions), decreasing the motor movement distances (mainly for soft keyboards), and providing a cognitively familiar UI layout or key mappings.

User studies are an essential part of the design process for new mobile text entry solutions. To ensure the internal and external validity of user studies in the thesis, a few relevant guidelines and measures were reviewed. First, comparative user studies are preferred. Second, text copying tasks with short and easily remembered sentences had best be adopted in user studies of text entry solutions. The primary measurements that need to be covered are text entry rate in WPM or CCPM, error rates, and keystrokes per character for showing the ambiguity level.

Chapter 3's analysis of existing Chinese text entry solutions implied that an affordable Chinese HWR solution is required if mobile devices are to accommodate more potential users. First, Chinese text entry solutions based on the 12-key keypad, including the pinyin and stroke methods, can hardly accommodate all potential users. People's native dialect, when it is not Mandarin, often negatively affects the use of the pinyin coding system and causes systematic errors in the production of pinyin marks. Also, people develop different writing orders for Chinese characters, while the stroke coding system requires a standard stroke order. These disadvantages of the pinyin and stroke coding systems bring either extra learning or obstacles for people. Second, Chinese HWR has potential to offer good immediate usability. However, it is supported by only those mobile products with touchscreens, which makes it expensive and does not match user expectations: Chinese HWR is expected even when touch interactions for all functions are not. Therefore, I proposed for mobile phones a Chinese HWR solution based on a touchpad: a cheaper solution that also enables combination of a touchpad with the 12-key keypad.

Chapter 3 also presented a customized user-centered design approach to develop the concept into interaction designs. Three modules of empirical studies to involve users in the design process were planned, for different purposes: to build user experience benchmarks for our concept via competitor analysis; to compare interaction options for the concept, including feedback and UI styles; and to evaluate the interaction design as well as the concept. In total, eight studies were conducted. I chose and presented three of them because these were representative for each module. The benchmark study of UI layout for Chinese HWR indicated that users preferred the box designs, supporting continuous input and virtual buttons. A quantitative study to explore the effects of user group (novice vs. expert) and UI design (full-screen vs. three-box design) on user performance indicated that only the main effect of user group is significant. Though the three-box design supports continuous input, users are not faster with it, since the errors they notice interrupt continuous writing. Finally, the evaluation based on a working prototype indicated that the concept and UI designs were well accepted by end users.

Chapter 4 explored the design issues of Chinese pinyin input with a rotator. A rotator, a ring or round device that can be rotated either clockwise or counterclockwise, is very good for navigation and also enables selections. It is not a direct input device as the keyboard is; it requires users to scroll in a list of items and select the target items to complete the text entry process. Since target selection from a list of options is an essential sub-process in Chinese pinyin solutions, it is worth exploring the potential of a selection-based Chinese text entry solution based on the pinyin coding system and a rotator.

I proposed two new designs based on the idea of “consonant plus vowel” and compared them to the traditional date stamp design. Although there were no significant differences among the three design solutions in user performance, the efficiency perceived by users was greater for one of the new designs, and it was their technique of choice for their own phone. Moreover, it is worth noting that the results were obtained in a setup that favored the initial design since no predictive input was implemented in the new designs. The interaction efficiency with the rotator was primarily affected by two factors: number of items in the scrolling list and number of selection actions. Thus, the prediction feature is quite important for text entry solutions based on a rotator, because it can significantly decrease the number of items in the scrolling list. I argued that it was because prediction was not implemented in the new designs that they did not outperform the traditional design.

In Chapter 5, I presented two studies of Mandarin dictation solutions for mobile phones. In the first section, I presented two rounds of evaluations of isolated Mandarin message dictation solutions with working prototypes. In both evaluations, both objective and subjective data were collected for understanding user behaviors and attitudes toward the isolated message dictation applications. It was found that the recognition rate increased after an optimization of the recognition engine. However, enrollment (the adaptation of the system to the user’s voice) does not help to increase recognition rates. Average text entry rates with the isolated dictation solutions on mobile phones were in the range 8–10 CCPM. Participants expected continuous dictation solutions, for they believed that continuous speaking is more natural and can bring better performance. The second study explored effects of speaking styles (phrasal input vs. sentence input) along with recognition rate and message length on user performance and satisfaction, with a Wizard of Oz method. The results indicated that all factors, including speaking style, recognition rate, and message length, affected user performance and satisfaction. Further task analysis indicated that error corrections took the greatest proportion of the time in the dictation process, followed by selection from among the recognition options and speaking time.

Chapter 6 presented two studies of the 12-key-keypad-based Chinese pinyin text entry solutions. The first study compared five phrasal text entry solutions, belonging to two categories: with and without the need for processes for selection of target pinyin marks. The results indicated that at least one interaction solution from each category enables satisfactory user performance. However, subjectively users preferred the category of solutions requiring no selection of target pinyin marks. Therefore, the future trend for the design of interaction for pinyin phrasal input would be that user selection of target pinyin marks is not required. Also, three additional design guidelines were proposed and discussed for the pinyin phrasal input feature.

The second study aimed to build for Chinese pinyin character input solutions a user model to estimate users' error-free input speeds. I divided the input process with a Chinese pinyin character input method into two parts: the production and typing of a pinyin mark and the disambiguation process for selecting the target pinyin marks and Chinese characters. I combined Fitts' law and a language model to estimate the average time required for production and typing of a pinyin mark. Then I divided the disambiguation process itself into two sub-processes: the visual search process and the navigation process. I utilized a keystroke-level model and a linear model to estimate, respectively, the navigation time and visual search time required in the disambiguation process. Then I conducted five experiments, the first two to set the parameters for Fitts' law and the linear law and the other three experiments to evaluate the model against empirical results. The evaluation indicated that the prediction of our model is effective, suggesting that the model can be used to evaluate user performance of Chinese pinyin text entry solutions on mobile phones.

In Table 7.1, I summarize user performance (text entry rates and error rates) with four types of Chinese text entry methods on mobile devices according to the results in this thesis. The four types of methods are Chinese handwriting recognition methods, rotator based pinyin input methods, isolated character dictation methods and pinyin input methods based on the 12-key keypad. According to the table, the keypad-based pinyin input methods supporting phrasal input enable the fastest text entry rates. Text entry rates with Chinese HWR methods are slightly lower than with keypad-based pinyin input methods. However, error rates with Chinese HWR methods are comparatively higher indicating it is still possible to improve them. With a Chinese SMS dictation application that supports isolated character speech input and the average 4-best recognition rate of around 96%, users can enter text with rates from 8 to 10 CCPM. Though users considered dictation a useful text input method, they were neutral on its usability. Rotator-based Chinese pinyin input yields much lower user speeds. However, a rotator has several good elements: it performs fine in navigation tasks, its small size enables it to be easily integrated into mobile devices, it is fun to use, and it can be used to enter written languages that have more characters. Further exploration of it is necessary, and our study provides guidelines for interaction designs employing a rotator.

Several additional research topics and questions were identified in this thesis for the future. First, there are high expectations that further work will improve the target selection process that is necessary in mobile Chinese text entry solutions. Second, studies of the interaction process with keypad-based stroke methods are necessary. Third, interaction designs for, and studies of, continuous Chinese HWR and Chinese specific soft keyboards are also clearly required.

Table 7.1: Empirical results for user performance with different Chinese text entry methods according to the studies in this thesis

<table>
<tr><th colspan="3" rowspan="2">Chinese text entry methods explored in this thesis</th><th colspan="2">Average empirical results (SD)</th></tr>
<tr><th>Error rate (deletion actions per Chinese character)</th><th>Text entry rate (CCPM)</th></tr>
<tr><td rowspan="2">HWR (with pen and mobile touchscreen device)</td><td colspan="2">Novice</td><td>0.26 (0.191)</td><td>14.5 (3.09)</td></tr>
<tr><td colspan="2">Expert</td><td>0.18 (0.139)</td><td>18.8 (3.95)</td></tr>
<tr><td rowspan="2">Rotator (novice)</td><td colspan="2">Date stamp</td><td>0.04 (0.042)</td><td>6.5 (1.05)</td></tr>
<tr><td colspan="2">Con. + vowel (Alpha)</td><td>0.05 (0.042)</td><td>6.6 (1.28)</td></tr>
<tr><td>Speech dictation</td><td colspan="2">Isolated character dictation</td><td>0.05–0.32</td><td>8–10</td></tr>
<tr><td rowspan="3">Keypad pinyin</td><td colspan="2">Character input with prediction (expert)</td><td>0.11 (0.102)</td><td>21.9 (2.28) (speeds without errors)</td></tr>
<tr><td rowspan="2">Phrasal input</td><td>Novice user</td><td rowspan="2">Nokia: 0.084 (0.026)</td><td>Nokia: 25.5 (8.0)</td></tr>
<tr><td>Trained user</td><td>Nokia: 34.2 (10.2)</td></tr>
</table>

8 References

Ai, H., Harris, T., & Rose, C. P. (2006). The effect of miscommunication rate on user response preferences. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, 448–453.

Alhonen, J., Cao, Y., Ding, G., Liu, Y., Olsen, J., Wang, X., & Yang, X. (2007). Mandarin short message dictation on Symbian Series 60 mobile phones. In *Proceedings of Mobility 2007*, ACM Press, 431–438.

Ann, T. K. (1982). Cracking the Chinese puzzles. *Stockflows*, **5**, 278–374.

Ao, X., Wang, X., Tian, F., Dai, G., & Wang, H. (2007). Crossmodal error correction of continuous handwriting recognition by speech. In *Proceedings of the 12th International Conference on Intelligent User Interfaces (IUI '07)*, ACM Press, 243–250.

Bellman, T., & MacKenzie, I. S. (1998). A probabilistic character layout strategy for mobile text entry. In *Proceedings of Graphics Interface (GI '98)*, Canadian Information Processing Society, 168–176.

Benko, H., Wilson, A. D., & Baudisch, P. (2006). Precise selection techniques for multi-touch screens. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, ACM Press, 1263–1272.

Bernsen, N. O., & Dybkjær, L. (1998). *Designing Interactive Speech Systems: From First Ideas to User Testing*. New York: Springer.

Brewster, S., Lumsden, J., Bell, M., Hall, M., & Tasker, S. (2003). Multimodal 'eyes-free' interaction techniques for wearable devices. In

Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '03), ACM Press, 473–480.

Card, S. K., Moran T. P., & Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM,* **23** (7), 396–410.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale: Erlbaum.

Carroll, J. M. (Ed.) (2003). *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*. San Francisco: Morgan Kaufmann.

Chang, E., Seide, F., Meng, H., Chen, Z., Shi, Y., & Li, Y. C. (2002). A system for spoken query information retrieval on mobile devices. *IEEE Transactions on Speech and Audio Processing*, **10** (8), 531–541.

Chang, E., Zhou, J., Di, S., Huang, C., & Lee, K. F. (2000). Large vocabulary Mandarin speech recognition with different approaches in modeling tones. In *Proceedings of 6th International Conference on Spoken Language Processing*, V2, 983–986.

Chen, C. J., Gopinath, R. A., Monkowski, M. D., Picheny, M. A., & Shen, K. (1997). New methods in continuous Mandarin speech recognition. In *EUROSPEECH '97*, 1543–1546.

Chen, J. M., & Yuen, C.-K. (1991). Effects of Pinyin and script type on verbal processing: Comparisons of China, Taiwan, and Hong Kong Experience. *International Journal of Behavioral Development,* **14** (4), 429–448.

Clarkson, E., Clawson, J., Lyons, K., & Starner, T. (2005). An empirical study of typing rates on mini-QWERTY keyboards. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, 1288–1291.

China Internet Network Information Center (CNNIC) (2010). *Statistic Report on Internet Development in China.* Retrieved from http://www.cnnic.net.cn/html/Dir/2010/01/15/5767.htm

Cockburn, A., & Gutwin, C. (2009). Predictive model of human performance with scrolling and hierarchical lists. *Human-Computer Interaction,* **24** (3), 273–314.

Cockburn, A., Gutwin, C., & Greenberg, S. (2007). A predictive model of menu performance. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*, ACM Press, 627–636.

Cockburn, A., & Siresena, A. (2003). Evaluating mobile text entry with the Fastap keypad. In *Proceedings of British Computer Society Conference on Human Computer Interaction*, 77–80.

Contemporary Chinese Dictionary (Third Version) (1996). Shangwu Publishing House.

Cox, A. L., Cairns, P. A., Walton, A., & Lee, S. (2008). Tlk or txt? Using voice input for SMS composition. *Personal and Ubiquitous Computing*, **12**, 567–588.

Cui, Y., & Lantz., V. (2005). Stroke break analysis: a practical method to study timeout value for handwriting recognition input. In *Proceedings of International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI '05)*, ACM Press, 263–266.

Curran, K., Woods, D., & Riordan, B. O. (2006). Investigating text input methods for mobile phones. *Telematics and Informatics,* **23** (1), 1–21.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies: why and how. In *Proceedings of the 1st International Conference on Intelligent User Interfaces (IUI '93)*, ACM Press, 193–200.

Dai, R., Liu, C., & Xiao, B. (2007). Chinese character recognition: history, status and prospects. *Frontiers of Computer Science in China,* **1** (2), 126–136.

Darroch, I., Goodman, J., Brewster, S., & Gray, P. (2005). The effect of age and font size on reading text on handheld computers. In *Lecture Notes in Computer Science*, 3585, Springer, 253–266.

Davis, R., Saponas, T. S., Shilman, M., & Landay, J. A. (2007). SketchWizard: Wizard of Oz prototyping of pen-based user interfaces. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)*, ACM Press, 119–128.

Deese, J. & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology,* **54** (3), 180–7.

Ding., X. (2002). Chinese character recognition: a review. *Chinese Journal of Electronics,* **30** (9), 1364–1368.

Dow, S., MacIntyre, B., Lee, J., Oezbek, C., Bolter, J. D., & Gandy, M. (2005). Wizard of Oz support throughout an iterative design process. *IEEE Pervasive Computing,* **4**, 18–26.

Dunlop, M. D., & Crossan, A. (2000). Predictive text entry methods for mobile phones. *Personal Technology,* **4** (2), 134–143.

Dunlop, M. D., & Masters, M. M. (2008). Investigating five key predictive text entry with combined distance and keystroke modeling. *Pervasive and Ubiquitous Computing,* **12**, 589–598.

Dunlop, M. D., & Taylor, F. (2009). Tactile feedback for predictive text entry. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI '09)*, ACM Press, 2257–2260.

Eisenhardt, M. K. (1989). Building theories from case study research. *The Academy of Management Review,* **14** (4), 532–550.

Entwistle, M. S. (2003). The performance of automated speech recognition systems under adverse conditions of human exertion. *International Journal of Human-Computer Interaction,* **16** (2), 127–140.

Fiedler, A., Gabsdil, M., & Horacek, H. (2004). A tool for supporting progressive refinement of Wizard-of-Oz experiments in natural language. In *Intelligent Tutoring Systems, 7th International Conference (ITS 2004)*, Lecture Notes in Computer Science 3220, Springer, 325–335.

Frankish, C., Hull, R., & Morgan, P. (1995). Recognition accuracy and user acceptance of pen interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '95)*, ACM Press, 503–510.

Fu, S. W. K., Lee, C. H., & Clubb, O. L. (1996). A survey on Chinese speech recognition. *Communications of COLIPS*, **6** (1), 1–17.

Furui, S. (2007). Speech-based interfaces. In: MacKenzie, I. S. & Tanaka-Ishii, K. (Eds.), *Text Entry Systems: Mobility, Accessibility, Universality*, Morgan Kaufmann, 159–174.

Gao, D., Zhong, Y., & Ceng, L. (1995). The influence of the frequency of use on the speed of recognition in the Chinese characters. *Chinese Journal of Psychological Science,* **18**, 225–228.

GB 2312. (1980). *Codes of Chinese character set for information interchange – Primary set*. Standardization Administration of China.

GB 13000. (1993). *Codes of Chinese character set for information interchange – Extended set*. Standardization Administration of China.

GB 18030. (2005). *Code of Chinese character set for information interchange – Extended set*. Standardization Administration of China.

Green, N., Kruger, J., Faldu, C., & Amant, S. R. (2004). A reduced QWERTY keyboard for mobile text entry. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, 1429–1432.

Griffiths, H. (2006). Hello mobile email, blogging and even instant messaging. *Revolution,* May 9.

Grimes, A. D., & Schulz, F. K. (2002). Bias and causal associations in observational research. *The Lancet (Epidemiology series)*, **359**, 248–252.

Grudin, J. (1991). Interactive systems: Bridging the gaps between developers and users. *IEEE Computer*, **24** (4), 59–69.

Guo, R., & Jin, L. (2007). Segmentation of on-line cursive handwritten Chinese word based on stroke speed feature and stroke vector feature. In *Proceedings of the IEEE International Conference on Automation and Logistics,* IEEE Press, 1576–1579.

Heloisa, M. S., & Ravindra S. G. (1998). Effectiveness of menu orientation in Chinese. *Human Factors,* **40** (4), 569–576.

Hewett, T. T., Baecker, R., Card, S., Carey, J., Gasen, J. B., Mantei, M. M., Perlman, G., Strong, G. W., & Verplank, B. (1992). ACM SIGCHI Curricula for Human-Computer Interaction. Retrieved from http://old.sigchi.org/cdg/cdgR.html

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, **4**, 11–26.

Hoffmann, A., Spelmezan, D., & Borchers, J. (2009). TypeRight: a keyboard with tactile error prevention. In *Proceedings of 27th International Conference on Human Factors in Computing Systems (CHI '09),* ACM Press, 2265–2268.

Hoggan, E., Brewster, S. A., & Johnston, J. (2008). Investigating the effectiveness of tactile feedback for mobile touchscreens. In *Proceedings of 26th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, ACM Press, 1573–1582.

Hornbaek, K. (2006). Current practice in measuring usability: challenges to usability studies and research. *International Journal of Human-Computer Studies*, **64**, 79–102.

Howell, M., Love, S., & Turner, M. (2005). The impact of interface metaphor and context of use on the usability of a speech-based mobile city guide service. *Behavior & Information Technology,* **24** (1), 67–78.

Huang, C., Chang, E., Zhou, J. L., & Lee, K. F. (2000). Accent modeling based on pronunciation dictionary adaptation for large vocabulary Mandarin speech recognition. In *Proceedings of International Conference on Spoken Language Processing*, 3, 818–821.

Hunter, M., Zhai, S., & Smith, B. A. (2000). Physics-based graphical keyboard design. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, 157–158.

Hwang, S., & Lee, G. (2005). Qwerty-like 3x4 keypad layouts for mobile phone. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems,* ACM Press, 1479–1482.

ISO. (1998). *Ergonomic Requirements for Office Work with Visual Display Terminals (VDT)s – Part 14: Menu Dialogues*. ISO/IEC 9241-14.

ISO. (1999). *Human-Centred Design Processes for Interactive Systems*. ISO/IEC 13407.

ISO. (2008). *Ergonomics of Human System Interaction – Part 210: Human-Centred Design for Interactive Systems* (formerly known as 13407). ISO DIS 9241-210.

Isokoski, P. (2004). Performance of menu-augmented soft keyboards. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*, ACM Press, 423–430.

Isokoski, P. & MacKenzie, I. S. (2003). Combined model for text entry rate development. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, 752–753.

James, C. L., & Reischel, K. M. (2001). Text input for mobile devices: Comparing model predictions to actual performance. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*, ACM Press, 365–371.

Jeng, J.-Y. (2005). The influence of speaking rate, utterance unit and position on segmental duration of Mandarin. *Journal of Tai Nan University – Social Science Edition,* **39** (1), 161–185.

Jokela, T., Iivari, N., Matero, J., & Karukka, M. (2003). The standard of User-Centered Design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11. In *Proceedings of the Latin American Conference on Human-Computer Interaction (CLIHC '03),* ACM Press, 53–60.

Karat, C.-M., Halverson, C., Karat, J., & Horn, D. (1999). Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*, ACM Press, 568–575.

Karat, J., Horn, D., Halverson, C., & Karat, C.-M. (2000). Overcoming unusability: developing efficient strategies in speech recognition systems. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, 141–142.

Karpov, E., Kiss, I., Leppänen, J., Olsen, J., Oria, D., Sivadas, S., & Tian, J. (2006). Short message dictation on Symbian Series 60 mobile phones. In *Proceedings of Speech in Mobile and Pervasive Environments Workshop (SiMPE Workshop '06)* in conjunction with Mobile HCI 2006.

Kernchen, R., Mossner, K., & Tafazolli, R. (2005). Adaptivity for multimodal user interfaces in mobile situations. In *Proceedings of Autonomous Decentralized Systems*, 469–472.

Klemmer, S. R., Sinha, A. K., Chen, J., Landay, J. A., Aboobaker, N., & Wang, A. (2000). Suede: a Wizard of Oz prototyping tool for speech user interfaces. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST '00)*, ACM Press, 1–10.

Kristensson, P.-O., & Denby., L. C. (2009). Text entry performance of state of the art unconstrained handwriting recognition: A longitudinal user study. In *Proceedings of 27th International Conference on Human Factors in Computing Systems (CHI '09),* ACM Press, 567–570.

Kristensson, P.-O., & Zhai, S. (2004). SHARK2: a large vocabulary shorthand system for pen-based computers. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*, ACM Press, 43–52.

Kujala, S. (2003). User involvement: a review of the benefits and challenges. *Behaviour & Information Technology*, **22** (1), 1–16.

Lagoá, P., Santana, P., Guerreiro, T., Gonçalves, D., & Jorge., J. (2007). BloNo: a new mobile text entry interface for the visually impaired. In *Proceedings of HCI International*, 4555, 908–917.

LaLomia, M. J. (1994). User acceptance of handwritten recognition accuracy. In *Companion Proceedings of the CHI '94 Conference on Human Factors in Computing Systems*, ACM Press, 107–108.

Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P. O. S., & Kort, J. (2009). Understanding, scoping and defining user experience: A survey approach. In *Proceedings of 27th International Conference on Human Factors in Computing Systems (CHI '09)*, ACM Press, 719–728.

Leventhal, L. M., McKeeby, J. W., & Mynatt, B. T. (1991). Screen keyboards: An empirical study of the effects of shape and character layout. In *Proceedings of INTERACT '91,* 108–112.

Levy, D. (2002). The Fastap keypad and pervasive computing. In *Proceedings of International Conference on Pervasive Computing (Pervasive 2002),* Lecture Notes in Computer Science 2414, Springer, 58–68.

Li, J. (2005). *Research on Wu Dialectal Chinese Speech Recognition*. Doctor Thesis, Tsinghua University, 2005.

Lin, M., & Sears, A. (2005). Chinese character entry for mobile phones: a longitudinal investigation. *Interacting with Computers,* **17**, 121–146.

Lin, M., & Sears, A. (2007). Constructing Chinese characters: keypad design for mobile phones. *Behaviour & Information Technology,* **26** (2), 165–178.

Ling, R. (2005). The socio-linguistics of SMS: an analysis of SMS use by a random sample of Norwegians. In: Ling, R. & Pederson, E. P. (Eds.), *Mobile Communications: Re-negotiation of the Social Sphere,* Springer, 335–349.

Liu, C.-L., Jaeger, S., & Nakagawa, M. (2004). Online recognition of Chinese characters: The state-of-the-art. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **26** (2), 198–213.

Liu, H., & Liu, Y. (2008). Evaluation of Chinese finger writing recognition on touch-sensitive keypad of mobile phones. In *Workshop of Cultural Usability and Human Work Interaction Design – Techniques that Connects* in conjunction with the Fifth Nordic Conference on Human-Computer Interaction (NordiCHI 2008).

Liu, Y., Ding., K., & Liu, N. (2009). Immediate user performances with touch Chinese text entry solutions on handheld devices. In *Proceedings of International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI '09)*, ACM Press, 56–57.

Liu, Y., Jiang, L., & Yang, X. (2009). Will input style affect Mandarin short messages in mobile device?: a Wizard of Oz study. In *Proceedings of Speech in Mobile and Pervasive Environments Workshop (SiMPE Workshop '09)* in conjunction with Mobile HCI 2009, 41–44.

Liu, Y., & Räihä, K.-J. (2008). RotaTxt: Chinese pinyin input with a rotator. In *Proceedings of International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI '08)*, ACM Press, 225–233.

Liu, Y., & Räihä, K.-J. (2010). Predicting Chinese text entry speeds on mobile phones. In *Proceedings of 28th International Conference on Human Factors on Computing Systems (CHI '10)*, ACM Press, 2183–2192.

Liu, Y., & Wang, Q. (2007). Chinese pinyin phrasal input on mobile phone: usability and developing trends. In *Proceedings of Mobility 2007*, ACM Press, 548–554.

Lumsden, J., & Brewster, S. (2003). A paradigm shift: alternative interaction techniques for use with mobile & wearable devices. In *Proceedings of the 2003 Conference of the Centre For Advanced Studies on Collaborative Research*, IBM Press, 197–210.

Lyons, K., Plaisted, D., & Starner, T. (2004). Expert chording text entry on the Twiddler one-handed keyboard. In *Proceedings of the Eighth International Symposium on Wearable Computers* (ISWC '04), IEEE CS Press, 94–101.

Lyons, K., Skeels C., & Starner, T. (2005). Providing support for mobile calendaring conversations: a Wizard of Oz evaluation of dual-purpose speech. In *Proceedings of International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI '05)*, 243–246.

Lyons, K., Starner, T., Plaisted, D., Fusia, J., Lyons, A., Drew, A., & Looney, E. W. (2004). Twiddler typing: one-handed chording text entry for mobile phones. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*, ACM Press, 671–678.

Ma, D., Ichikawa, F., Liu, Y., & Jiang, L. (2007). Use of Chinese short messages. In *Proceedings of HCI International 2007, Vol. 9,* Lecture Notes in Computer Science 4558, Springer, 582–591.

MacKenzie, I. S. (2002a). KSPC (keystrokes per character) as a characteristic of text entry techniques. In *Proceedings of 4th International Symposium on Human-Computer Interaction with Mobile Devices (Mobile HCI '02)*, 195–210.

MacKenzie, I. S. (2002b). Mobile text entry using three keys. In *Proceedings of the Second Nordic Conference on Human-Computer Interaction (NordiCHI 2002)*, ACM Press, 27–34.

MacKenzie, I. S. (2003). Motor behavior models for human-computer interaction. In J. M. Carroll (Ed.), *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*, Morgan Kaufmann, 27–53.

MacKenzie, I. S. (2007). Evaluation of text entry techniques. In: MacKenzie, I. S. & Tanaka-Ishii, K. (Eds.), *Text Entry Systems: Mobility, Accessibility, Universality*, Morgan Kaufmann, 75–102.

MacKenzie, I. S. (2009). The one-key challenge: Searching for a fast one-key text entry method. In *Proceedings of the ACM Conference on Computers and Accessibility (ASSETS '09)*, ACM Press, 91–98.

MacKenzie, I. S., & Buxton, W. (1992). Extending Fitts' law to two-dimensional tasks. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*, ACM Press, 219–226.

MacKenzie, I. S., Kober, H., Smith, D., Jones, T., & Skepner, E. (2001). LetterWise: prefix-based disambiguation for mobile text input. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST '01)*, ACM Press, 111–120.

MacKenzie, I. S., & Soukoreff, R. W. (2002). Text entry for mobile computing: models and methods, theory and practice. *Human-Computer Interaction*, **17**, 147–198.

MacKenzie, I. S., & Soukoreff, R. W. (2003). Phrase sets for evaluating text entry techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, 754–755.

MacKenzie, I. S., & Tanaka-Ishii, K. (Eds.) (2007). *Text Entry Systems: Mobility, Accessibility, Universality*. San Francisco: Morgan Kaufmann.

MacKenzie, I. S., Zhang, S. X., & Soukoreff, R. W. (1999). Text entry using soft keyboards. *Behaviour and Information Technology,* **18**, 235–244.

Majaranta, P., & Räihä, K.-J. (2007). Text entry by gaze: utilizing eye tracking. In: MacKenzie, I. S. & Tanaka-Ishii, K. (Eds.), *Text Entry Systems: Mobility, Accessibility, Universality*, Morgan Kaufmann, 175–188.

Marcus, A. (2003). A great leap forward. *Interactions* (January and February), 21–25.

Ministry of Industry and Information Technology of China (MIIT). (2009). Statistics (in Chinese). Retrieved from http://www.miit.gov.cn/n11293472/n11293832/n11294132/n11302706/12426281.html

Molin, L. (2004). Wizard-of-Oz prototyping for co-operative interaction design of graphical user interfaces. In *Proceedings of the Third Nordic Conference on Human-Computer Interaction (NordiCHI '04)*, 425–428.

Munteanu, C., Penn, G., Baecker, R., & Zhang, Y. (2006). Automatic speech recognition for webcasts: how good is good enough and what to do when it isn't. In *Proceedings of the 8th International Conference on Multimodal Interfaces (ICMI '06)*, ACM Press, 39–42.

Myung, R. (2004). Keystroke-level analysis of Korean text entry methods on mobile phones. *International Journal of Human-Computer Studies,* **60** (5-6), 545–563.

Nesbat, S. B. (2003). A fast, full-text entry system for small electronic devices. In *Proceedings of the 5th International Conference on Multimodal Interfaces (ICMI '03),* ACM Press, 4–11.

Nielsen, J. (1993). *Usability Engineering*. San Diego: Academic Press.

Patel, N., Clawson, J., & Starner, T. (2009). A model of two-thumb chording on a phone keypad. In *Proceedings of International Conference on Human-Computer Interaction in Mobile Devices and Services (Mobile HCI '09),* ACM Press, 69–72.

Pavlovych, A., & Stuerzlinger, W. (2004). Model for non-expert text entry speed on 12-button phone keypads. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*, ACM Press, 351–358.

Pellegrino, F., Farinas, J., & Rouas, J.-L. (2004). Automatic estimation of speaking rate in multilingual spontaneous speech. In *Proceedings of International Conference on Speech Prosody*, 517–520.

Poupyrev, I., & Maruyama, S. (2003). Tactile interfaces for small touch screens. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST '03)*, ACM Press, 217–220.

Price, K. J., & Sears, A. (2005). Speech-based text entry for mobile handheld devices: an analysis of efficacy and error correction techniques for server-based solutions. *International Journal of Human-Computer Interaction,* **19** (3), 279–304.

Proschowsky, M., Schultz, N., & Jacobsen, E. N. (2006). An intuitive text input method for touch wheels. In *Proceedings of* SIGCHI *Conference on Human Factors in Computing Systems (CHI '06),* ACM Press, 467–470.

Qiao, J., Qiao, Y., & Qiao, S. (1990). Six-digit coding method. *Communications of the ACM*, **33**, 491–494.

Ranjan, A., Balakrishnan, R., & Chignell, M. (2006). Searching in audio: the utility of transcripts, dichotic presentation, and time-compression. In *Proceedings of SIGCHI conference on Human Factors in Computing Systems (CHI '06)*, ACM Press, 721–730.

Ronkainen, S., Koskinen, E., Liu, Y., & Korhonen, P. (2010). Environment analysis as a basis for designing multimodal and multi-device user interfaces. *Human-Computer Interaction,* **25** (2), 148–193.

Ryu, H., & Cruz, K. (2005). LetterEase: Improving text entry on a handheld device via letter reassignment. In *Proceedings of the 17th Australia Conference on Computer-Human Interaction (OZCHI '05)*, ACM Press, 1–10.

Sacher, H. (1998). Interactions in Chinese: designing interfaces for Asian languages. *Interactions,* **5** (5), 28–38.

Sacher, H., Tng, T.-H., & Loudon, G. (2001). Beyond translation: approaches to interactive products for Chinese consumers. *International Journal of Human-Computer Interaction,* **13** (1), 41–51.

Sears, A., Jacko, J. A., Chu, J., & Moro, F. (2001). The role of visual search in the design of effective soft keyboards. *Behaviour & Information Technology,* **20** (3), 159–166.

Seow, S. C. (2005). Information Theoretic models of HCI: A comparison of the Hick-Hyman Law and Fitts' Law. *Human-Computer Interaction,* **20** (3), 315–352.

Serrano, M., Nigay, L., Demumieux, R., Descos, J., & Losquin, P. (2006). Multimodal interaction on mobile phones: development and evaluation using ACICARE. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI '06)*, ACM Press, 129–136.

Sheng, J. (1985). A pinyin keyboard for inputting Chinese characters. *Computer*, **18** (1), 60–63.

Shin, H., Lee, W., Lee, G., & Cho, I. (2009). Multi-Point touch input method for Korean text entry. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, 3871–3876.

Shneiderman, B. (2000). The limits of speech recognition. *Communications of the ACM,* **43** (9), 63–65.

Shneiderman, B. (2003). *Leonado's Laptop: Human Needs and the New Computing Technologies*. Cambridge: MIT Press.

Shuang Pin. (2008). Shuang Pin Method. Retrieved from http://www.hongen.com/pc/pcketang/ime/imejssp1.htm

Shuang Pin CE. (2008). Shuang Pin Method in Windows CE system. Retrieved from http://msdn2.microsoft.com/en-us/library/ms921792.aspx

Silfverberg, M. (2007). Historical overview of consumer text entry technologies. In: MacKenzie, I. S. & Tanaka-Ishii, K. (Eds.), *Text Entry Systems: Mobility, Accessibility, Universality*, Morgan Kaufmann, 3–26.

Silfverberg, M., MacKenzie, I. S., & Korhonen, P. (2000). Predicting text entry speed on mobile phones. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*, ACM Press, 9–16.

Sinha, A. K., Shilman, M., & Shah, N. (2001). MultiPoint: a case study of multimodal performance for building presentations. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, 431–432.

Siok, W. T., Perfetti, C. A., Jin, Z., & Tan, L. H. (2004). Biological abnormality of impaired reading is constrained by culture. *Nature,* **431**, 71–76.

Soukoreff, R. W. (2002). Text entry for mobile systems: models, measures, and analyses for text entry research. *Master's thesis*, York University.

Soukoreff, R. W., & MacKenzie, I. S. (1995). Theoretical upper and lower bounds on typing speed using a stylus and soft keyboard. *Behaviour & Information Technology,* **14**, 370–379.

Soukoreff, R. W., & MacKenzie, I. S. (2003). Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '03),* ACM Press, 113–120.

Soukoreff, R. W., & MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation: Perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies,* **61**, 751–789.

Starner, T. (2004). Keyboards Redux: fast mobile text entry. *Pervasive Computing,* **3** (3), 97–101.

Stocky, T., Faaborg, A., & Lieberman, H. (2004). A commonsense approach to predictive text entry. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, 1163–1166.

Takahashi, H., Ogasawara, A., Ogasawara, M., & Ren, X. (2005). The effects of PDA pen-length on the performance of older adults. In *Proceedings of the 2005 International Conference on Active Media Technology*, 283.

Tan, L. H., Spinks, A. J., Eden, F. G., Perfetti, A. C., & Siok, W. T. (2005). Reading depends on writing, in Chinese. *PNAS,* **102** (24), 8781–8785.

Tanaka-Ishii, K., Zhou, M., & Kim, J.-D. (2007). Text entry in east Asian languages. In: MacKenzie, I. S. & Tanaka-Ishii, K. (Eds.), *Text Entry Systems: Mobility, Accessibility, Universality*, Morgan Kaufmann, 203–226.

Tang, C., & van Heuven, V. J. (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua*, **119** (5), 709–732.

Tarasewich, P. (2003). Evaluation of thumbwheel text entry methods. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, 756–757.

Tian, F., Lv, F., Wang, J., Wang, H., Luo, W., Kam, M., Setlur, V., Dai, G., & Canny, J. (2010). Let's play Chinese characters – mobile learning approaches via culturally inspired group games. In *Proceedings of 28th International Conference on Human Factors on Computing Systems (CHI '10)*, ACM Press, 1603–1612.

Viikki, O., Kiss, I., & Tian, J. (2001). Speaker and language independent speech recognition in mobile communication systems. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*. 4 pp.

Vredenburg, K., Isensee, S., & Righi, C. (2002). *User-Centred Design. An Integrated Approach*. Upper Saddle River: Prentice Hall.

Wang, H.-M., Ho, T.- H., Yang, R.-C., Shen, J.- L., Bai, B.-R., Hong, J.-C., Chen, W.-P., Yu, T.-L., & Lee, L.-S. (1997). Complete recognition of continuous Mandarin speech for Chinese language with very large vocabulary using limited training data. *IEEE Transaction on Speech and Audio Processing*, **5** (2), 195–200.

Wang., J. (2003). Human-computer interaction research and practice in China. *Interactions,* **10** (2), 88–96.

Wang, J., Zhai, S., & Su, H. (2001). Chinese input with keyboard and eye-tracking: an anatomical study. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*, ACM Press, 349–356.

Wang, R. H. (1999). National performance assessment of speech recognition systems for Chinese. In *Proceedings of Oriented COCOSDA Workshop*, 1999.

Wang, X., Li, J., Ao, X., Wang, G., & Dai, G. (2006). Multimodal error correction for continuous handwriting recognition in pen-based user interfaces. In *Proceedings of the 11th International Conference on Intelligent User Interfaces (IUI '06)*, ACM Press, 324–326.

Ward, D., Blackwell, A., & MacKay, D. (2000). Dasher—a data entry interface using continuous gesture and language models. In *Proceedings of the 13th Annual ACM Symposium on User Interference Software and Technology (UIST '00)*, ACM Press, 129–136.

Ward, D. J., Blackwell, A. F., & MacKay, D. J. C. (2002). Dasher: A gesture-driven data entry interface for mobile computing. *Human–Computer Interaction,* **17**, 199–228.

Wigdor, D., & Balakrishnan, R. (2003). TiltText: Using tilt for text input to mobile phones. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST '03),* ACM Press, 81–90.

Wigdor, D., & Balakrishnan, R. (2004). A comparison of consecutive and concurrent input text entry techniques for mobile phones. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*, ACM Press, 81–88.

Wobbrock, J. O. (2007). Measures of text entry performance. In: MacKenzie, I. S. & Tanaka-Ishii, K. (Eds.), *Text Entry Systems: Mobility, Accessibility, Universality*, Morgan Kaufmann, 47–74.

Wobbrock, J. O., & Myers, B. A. (2006). Trackball text entry for people with motor impairments. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '06),* ACM Press, 479–488.

Wobbrock, J. O., Myers, B. A., & Aung, H. H. (2004). Writing with a joystick: A comparison of date stamp, selection keyboard, and EdgeWrite. In *Proceedings of Graphics Interface (GI '04)*, Canadian Human-Computer Communications Society, 1–8.

Wolf, C. G., Glasser, A. R., & Fujisaki, T. (1991). An evaluation of recognition accuracy for discrete and run-on writing. In *Proceedings of the Human Factors Society 35th Annual Meeting*, Human Factors and Ergonomics Society.

Wolfe, J. M. (1998). Visual Search. In: H. Pashler (Ed.), *Attention*, University College London Press, 13-56.

Wu, C., Zhang, K., & Hu, Y. (2003). Human performance modeling in temporary segmentation Chinese character handwriting recognizers. *International Journal of Human-Computer Studies,* **58**, 483–508.

Zhai, S. (2004). Characterizing computer input with Fitts' law parameters — The information and noninformation aspects of pointing. *International Journal of Human-Computer Studies,* **61** (6), 791–809.

Zhai, S., Hunter, M., & Smith, B. A. (2000). The Metropolis Keyboard—an exploration of quantitative techniques for virtual keyboard design. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST '00)*, ACM Press, 119–128.

Zhai, S., & Kristensson, P.-O. (2003). Shorthand writing on stylus keyboard. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*, ACM Press, 97–104.

Zhai, S., & Kristensson, P.-O. (2007). Introduction to Shape Writing. In: MacKenzie, I. S. & Tanaka-Ishii, K. (Eds.), *Text Entry Systems: Mobility, Accessibility, Universality*, Morgan Kaufmann, 139–158.

Zhai, S., Kristensson, P.-O., & Smith, B. A. (2005). In search of effective text input interfaces for off the desktop computing. *Interacting with Computers,* **17**, 229–250.

Zhai, S., Smith, B. A., & Hunter, M. (2002). Performance optimization of virtual keyboards. *Human–Computer Interaction,* **17** (2,3), 89–129.

Zhai, S., Sue, A., & Accot, J. (2002). Movement model, hits distribution and learning in virtual keyboarding. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*, ACM Press, 17–24.

Zhang, Y., et al. (2008). *Kangxi Dictionary*. Social Sciences Academic Press of China.

Zhou, L., Shi, Y., Zhang, D., & Sears, A. (2006). Discovering cues to error detection in speech recognition output: A user-centered approach. *Journal of Management Information Systems*, **22** (4), 237–270.

Zhou, Y. (1953). *Research on Chinese Pinyin*. Shanghai East Bookstore.

Zhou, Y. (2007). *Han yu pinyin*. San Lian Bookstore.

Ziegler, J. C., & Goswami, U. C. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin,* **131** (1), 3–29.

Zou, Y., Yu, K., & Wang, K. (2008). Continuous Chinese handwriting recognition with language model. In *Proceedings of the 11th International Conference on Frontiers in Handwriting Recognition.*

1. **Timo Partala:** Affective Information in Human–Computer Interaction
2. **Mika Käki:** Enhancing Web Search Result Access with Automatic Categorization
3. **Anne Aula:** Studying User Strategies and Characteristics for Developing Web Search Interfaces
4. **Aulikki Hyrskykari:** Eyes in Attentive Interfaces: Experiences from Creating iDict, a Gaze-Aware Reading Aid
5. **Johanna Höysniemi:** Design and Evaluation of Physically Interactive Games
6. **Jaakko Hakulinen:** Software Tutoring in Speech User Interfaces
7. **Harri Siirtola:** Interactive Visualization of Multidimensional Data
8. **Erno Mäkinen:** Face Analysis Techniques for Human–Computer Interaction
9. **Oleg Špakov:** iComponent – Device-Independent Platform for Analyzing Eye Movement Data and Developing Eye-based Applications
10. **Yulia Gizatdinova:** Automatic Detection of Face and Facial Features from Images of Neutral and Expressive Faces
11. **Päivi Majaranta:** Text Entry by Eye Gaze
12. **Ying Liu:** Chinese Text Entry with Mobile Devices