

Chinese Input with Keyboard and Eye-Tracking - An Anatomical Study

Jingtao Wang⁺Shumin Zhai⁺⁺Hui Su⁺

⁺ IBM China Research Lab
No. 7, 5th Street, Shangdi
Beijing, 100085, P. R. China
+86 10 6298 6677 {ext.309, 330}
{jingtaow, suhui}@cn.ibm.com

⁺⁺ IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120, USA
+1 408 927 1112
zhai@almaden.ibm.com

ABSTRACT

Chinese input presents unique challenges to the field of human computer interaction. This study provides an anatomical analysis of today's standard Chinese input process, which is based on *pinyin*, a phonetic spelling system in Roman characters. Through a combination of human performance modeling and experimentation, our study decomposed the Chinese input process into sub-tasks and found that choice reaction time and numeric keying, two components resulted from the large number of homophones in Chinese, were the major usability bottlenecks. Choice reaction alone took 36% of the total input time in our experiment. Numeric keying for multiple candidates selection tends to take the user's attention away from the computer visual screen. We designed and implemented the EASE (Eye Assisted Selection and Entry) system to help maintaining complete touch-typing experience without diverting visual attention to the numeric keys. The EASE approach used a common selection key (spacebar) and implicit eye-tracking to replace the numeric keystrokes. Our experiment showed that such a system could indeed work, even with today's imperfect eye-tracking technology.

Keywords

Chinese text input, *pinyin* input, gaze, eye-tracking, gaze-tracking, multi-modal interface, performance modeling.

INTRODUCTION

Text entry is one of the most frequent human computer interaction tasks. Although great strides have been made towards speech and handwriting recognition, typewriting remains and will likely be the main text entry method in the future. A recent study [5] showed that text entry by today's continuous speech recognition is still far slower than typing

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGCHI'01, March 31-April 4, 2001, Seattle, WA, USA.
Copyright 2001 ACM 1-58113-327-8/01/0003...\$5.00.

(13.6 vs. 32.5 cwpm for transcription and 7.8 vs. 19.0 cwpm for composition). Furthermore, the study also revealed many human-factors issues that were not previously realized. For example, many users found it "harder to talk and think than type and think" and considered the keyboard to be more "natural" than speech for text entry.

Once learned, touch-typing offers two critical characteristics: rapid speed and full attention focus on the screen. A skilled typist can type 60 words per minute, far beyond what handwriting could achieve [13]. The other perhaps even more important property is that touch-typing frees the user's visual attention so she can focus on her task on the computer display.

Numerous difficulties rise when the user's language is Chinese (or many other non-alphabetic languages). Currently the most popular method used for Chinese text input in China mainland is *pinyin* input. *Pinyin* is the official Chinese phonetic alphabet based on Roman characters. For example, in *pinyin* the Chinese character 中 (center, middle) is "zhong" and the *pinyin* for the word "北京", consisting of two Chinese characters, is "Beijing".

About 97% of computer users in China use *pinyin* or some variations of it for daily input [3]. There have been other non-*pinyin*-based Chinese text input methods that encode the logographic Chinese characters, but the amount of learning and memorization has prevented them from becoming popular.

The complication to *pinyin* input, however, is that most Chinese characters are homophonic with many others. In Mandarin Chinese, there are only about 410 distinct syllables¹ [6]. In contrast, there are 6,763 Chinese characters in the national standard database GB2312. This means that on average, each syllable corresponds to 16.8 characters, notwithstanding the relatively small number of

¹ Unlike in English, a syllable in Chinese only takes two forms: a vowel or a consonant-vowel combination. Note also that each syllable has four tones.

characters with multiple pronunciations. As a result, when a user types the *pinyin* of a character, e.g. “zhong”, the computer software displays many candidate characters with the same pronunciation, together with identifying numbers, typically in a “page” (usually a one-line graphical window). For example, the first eight candidate characters for “zhong” could be

1中 2种 3重 4众 5种 6终 7忠 8肿 ,

which correspond to the following meanings: “1.center, 2.type, 3.heavy, 4.mass, 5.kind, 6.finale, 7.loyal, 8.swollen”. The user must then select a choice from the multiple candidates by typing the identifier number, e.g. “1”. When the candidates take more than one line, the user has to scroll to another line of candidates by hitting a page key. When the intended character is the first in a line, the user can select it by typing either number 1 or the spacebar.

HUMAN PERFORMANCE DIFFICULTIES IN PINYIN INPUT

The multiple choice selection process, in our view, is what makes today’s *pinyin* input much less efficient than typing in English. Our view is based on two human performance factors.

Choice reaction time

The first factor is the information processing involved in multiple-choice reaction. It is possible to apply the Hick’s law to *estimate*, in a first order approximation, the time T_r in such a process:

$$T_r = I_c H \quad (1)$$

where I_c is a constant between 150 to 200 ms [2] [12] and H is the information entropy (or uncertainty) in the n number of stimuli measured in bits:

$$H = \sum_{i=1}^n p_i \log_2(1/p_i + 1) \quad (2)$$

p_i is the probability for item i to be the target. In the current context, it is the probability for choice i to be the intended character. In the worst case, when all choices are equally probable,

$$H = \log_2(n + 1) \quad (3)$$

If we assume the target is in the first 7 choices appearing in the first line of display, then $H = 3$ bits; $T_r = 450 \sim 600$ ms.

On average, each Chinese character’s *pinyin* has 4.2 Roman characters. For a skilled typist, an average keystroke takes 200 ms (corresponding to 60 wpm in English) [2, 13]. Her *pinyin* typing time for each Chinese character is therefore around $200 \times 4.2 = 840$ ms. This means that in this case the ratio between the time spent on actual *pinyin* typing (840 ms) to the time to identify the correct choice (450 ~ 600 ms) is a startling 1: 0.53 ~ 1: 0.71.

If the intended character is not on the first page, T_r could be even longer. On average, however, the H value in equation (2) could be lower than 3 bits. First, as the user gains experience, she may be able to anticipate where the correct choice is likely to occur, hence changing the p_i distribution and reducing the entropy value H . Second, the choices in *pinyin* software are usually listed by their frequency rank, giving a skewed distribution of p_i values, which reduces the H value in equation (2).

There are limitations to the foregoing analysis of the portion of choice reaction time in *pinyin*-based input. First, the input experience with each Chinese character varies significantly both between users and within a user over time. Second, the number, the sequence and the frequency distribution of the homophonic characters for each *pinyin* syllable also change significantly. Third, the validity to Hick’s law in the context of Chinese character recognition needs to be questioned. In short, while the Hick’s law approach gives us some baseline points, it is still necessary to empirically measure the actual average choice selection time in Chinese input.

Many methods have been invented to reduce the frequency and number of choices in *pinyin*-based input. These methods include the following five categories.

1. Using additional keystrokes to represent shape or structure information of a Chinese character.
2. Enable the user to input both characters and words. A Chinese word is usually composed of one, two or three Chinese characters. If the user types the *pinyin* of “北京” (Beijing) separately, i.e. ‘bei’ and ‘jing’, the two syllables have 29 and 40 candidates respectively. If she types them as one unit, then only two candidates “1北京 2背景” (1. Beijing, 2. Background) are the likely choices in daily language.
3. Using a Chinese language model to reduce the uncertainty at the phrase or sentence level [14, 15]. The language model can be either rule-based or bigram or trigram-based, which are commonly used in speech recognition.
4. Adding the Chinese intonation information after the *pinyin* of a character. The four tones in Chinese pronunciation are often encoded as 1-4 for additional entry.
5. Continuous completion. The system continuously composes possible choices based on the *pinyin* characters typed. As the *pinyin* stream grows longer, the number of possible choices is reduced, until the user decides to select a choice.

Some of these measures help to reduce the choice reaction time, but still cannot completely eliminate it.

Numeric key typing

The other problem in multiple homophonic character selection lies in the difficulty of touch-typing the numeric keys for selecting the target character. On a standard QWERTY keyboard, typing the numeric keys is much more difficult than typing the alphabetic keys. This is in part due to the distance between the numeric keys and "home row" - the ASDFGHJKL keys. We observed this distance degrades the human open loop motor control accuracy, making it difficult to type correctly without visual guidance. A survey, to be detailed shortly, shows that only a little over 30 percent of computer users claim to be able to touch-type on the numeric keys, albeit with reduced accuracy.

For the majority of users who cannot touch-type the numeric keys, the two important advantages of typewriting – speed and the low demand of visual attention – both suffer when Chinese input is needed.

The QWERTY typewriter was invented by Christopher Sholes, Carlos Glidden and Samuel Soule in 1867. Its initial purpose was not in speed, but the superior legibility over handwriting [13]. The rapid speed of typewriting is largely a result of the method of touch-typing without looking at the keyboard, discovered independently by L.V. Longley and F. E. McFurrin in the 1880s. Since the 1920s, touch-typing has been accepted without controversy as the superior typing method, and its wide adoption rapidly increased typing speed [13].

Other than speed, the low visual attention demand in touch-typing is even more relevant to computer input. Without touch-typing, the user has to divide her visual attention between the screen and the keyboard, making the interaction process less seamless.

To conclude, in addition to multiple choice processing, the frequent use of numeric keying is another human computer interaction bottleneck for Chinese users. Avoiding numeric keying and hence maintaining the user's touch-typing ability is a compelling goal in Chinese text input. An apparent solution to this issue is to use alphabetic characters rather than numeric keys to label the homophonic candidates. However, it would be not possible to distinguish between a character that is part of the *pinyin* and a character this is a label for a candidate, unless a mode switching scheme is involved. For example, the string *jing* can be the *pinyin* for a character or the *pinyin* plus label *g* for a different character.

We have investigated a novel approach to replace the numeric keying without mode switching - by means of eye tracking.

“WHAT YOU SEE IS WHAT YOU GET” - EYE ASSISTED SELECTION AND ENTRY (EASE)

Using the eyes as a channel of input in advanced user interfaces has long been a topic of interest to the HCI field [1, 4, 7, 11]. As we previously argued elsewhere [16], there

are two shortcomings to use the eye gaze as a *direct control* channel, regardless of the maturity of eye-tracking technology. First, given the one-degree size of the fovea and the subconscious jittery motions that the eyes constantly produce, eye gaze cannot be very precise. Second, and perhaps more importantly, the eye, as one of our primary perceptual devices, is not naturally suited for deliberate control functions. Sometimes its movements are voluntarily controlled while at other times it is driven by external events. Based on these observations, Zhai and colleagues [16] proposed the MAGIC pointing method in which eye gaze was used to define a cursor's starting position and the hand remained to be the fine control organ. We may generalize the idea of MAGIC pointing to the following principles.

1. The hand and the eye work in combination.
2. The eye gaze defines the framework, context, or constraint. Its input to control should be implicit.
3. The hand acts in the context of the eye-gaze. It should remain the explicit control organ.

Applying these principles to Chinese text input, we rejected the possibility of using the eye-gaze as a deliberate selection method, which is a common method in “eye typing” [10]. For the same reason, we also rejected the idea of blinking or dwell time threshold as a way of making a selection.

Instead we propose to let the user type a common key, such as the spacebar which can be easily touch-typed, for the purpose of selection. However, what is selected is the Chinese character the user sees when she presses on the common key. In other words, “what you see is what you get”. Such a method took advantage of the following observations:

The user has to visually search and locate the intended character as a natural part of homophonic choice selection in Chinese input.

These candidates can be displayed in separation greater than one visual degree, hence above the threshold of eye-tracking.

The candidates are usually displayed in one row at the bottom of the screen. This means that the eye-tracker only has to be one dimensional, which may simplify the eye-tracking and make it more reliable.

The common key can be touch-typed easily, in comparison to looking and typing on the numeric keys on the top roll or on the side of a keyboard and then switch the gaze back to the screen.

Similar to MAGIC pointing [16], the current invention does not require the user to consciously stare and select an object by the eye. Instead, the eye gaze is implicitly used when the key is pressed.

We have implemented these ideas in a prototypical system dubbed EASE (Eye Assisted Selection and Entry). EASE consists of two parts. One is an eye-tracking system and the other is the Chinese *pinyin* software with gaze input.

Our eye tracker was developed at the IBM Almaden Research Center. It uses two near infrared (IR) time multiplexed light sources, composed of two sets of IR LED's, which are synchronized with the camera frame rate. One light source is placed very close to the camera's optical axis; the other light source is relatively far away from the axis. The two light sources produce different effects on the pupil area of eye as shown in Figure 1. The pupil area in the image is detected robustly by comparing the two images. The user's point of gaze can be determined by incorporating both the location information of the pupil area and that of the corneal reflection point. The detailed descriptions of our eye tracker can be found in [8]. The raw data from an eye tracker cannot be directly used for gaze-based interaction, due to noise from image processing, eye movement jitters, and samples taken during saccades (ballistic eye movements). We used a simple filtering method described in [16] to find likely fixations.

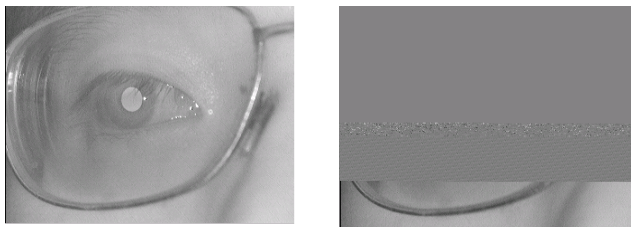


Figure 1. Bright and dark pupil image result from the time multiplexed IR illumination.

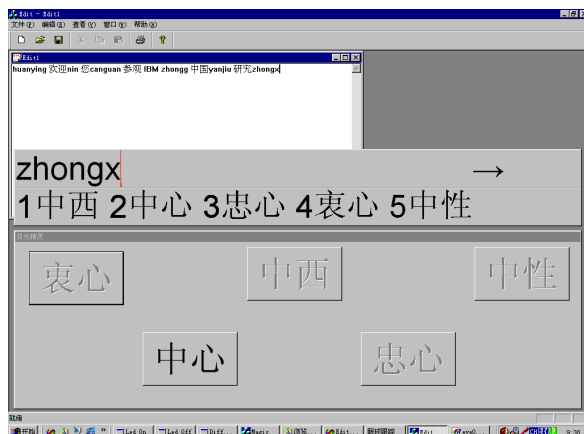


Figure 2. The EASE Chinese input system interface

Although the IBM Almaden eye-tracker is more reliable than the commercially available eye-trackers due to the dual IR illumination scheme, it still has several limitations. One is the 30 fps sampling frequency, which causes up to 100 ms delay in detecting fixations. The other is the requirement of the user staying steady in head position during use.

The EASE software was written in Visual C++ 6.0 and ran on the Simplified Chinese version of MS Windows NT 4.0. Our *pinyin* input implementation works at both character and word level. It has 6,763 Chinese *characters* in the National Standard GB-2312 1980 as well as 9425 most frequently used Chinese *words*. The multiple character or word homonyms for the same *pinyin* are arranged in the order of usage frequency, collected from a large Chinese corpus.

Instead of the linear candidate display order in traditional *pinyin* input method, a 'w' shape order is designed to display candidates to make the candidate selection procedure more robust, in order to minimize the limitations of today's technology. The user interface of the EASE prototype is shown in Figure 2.

USER STUDIES

Purpose

It is evident in the foregoing analysis that typing a character or a word in a *pinyin*-based system is an interaction task composed of a series of sub-tasks, including:

- t_1 - cognitively generating the *pinyin* of the Chinese character to be typed.
- t_2 - typing the *pinyin* character by character to the computer.
- t_3 - visually searching the target character from a list of candidates. If the expected character is not found, do page down/up operation and scan the next list of candidates, until the target character is found.
- t_4 - typing the numeric key corresponding to the target character.

The total time to input one Chinese character or word is:

$$T = t_1 + t_2 + t_3 + t_4 \quad (4)$$

As we previously pointed out, t_3 and t_4 could be the performance bottleneck in Chinese input. We have analyzed t_3 in light of Hick's law and estimated that it could be on par with t_2 -- the actual *pinyin* typing time. However, such estimation was taken with many highly simplified assumptions. The average t_3 as a portion of the total T can only be obtained with high confidence through experimentation.

The problem with t_4 is the visual attention distraction due to the difficulty of touch-typing numeric choice keys. To that end, we have proposed and designed the EASE method to maintain the user's visual attention on the screen. Can the EASE method indeed work? Is there any hand eye timing mismatch that prevents successful use of the EASE system? Should t_4 be shorter with the EASE method since the user does not have to look at and type on the numeric keys but touch-type on a common key instead? To answer these questions, we conducted a survey and an experiment.

Survey

The purpose of the survey was to verify our analysis of users' numeric key touch-typing ability. The survey was conducted in the department of computer science at the IBM Almaden Research Center (ARC) and the IBM China Research Lab (CRL). 127 employees, 109 from ARC and 18 from CRL, responded to an email survey. The majority of the respondents were computer science researchers and engineers; others were management and support staff. The average typing experience was 19 years. The results are summarized in the following Figure 3.

	Alphabetic keys	Top row numeric keys
ARC	79.8%	30.2%
CRL	88.9%	37.5%
Total	81.1%	31.2%

Figure 3. Percentage of staff who can touch-type the alphabet and numeric keys in ARC & CRL

From this survey we can see that less than half of the alphabet touch typists can touch-type the numeric keys, supporting our analysis on the difficulty of touch-typing on numeric keys due to the distance to the home row. Although the number of respondents in CRL was too small for a sound comparison, a higher percentage of respondents in CRL could touch-type the numeric keys. This may suggest that practice could overcome the difficulty of numeric key touch-typing. However, even if there is a practice effect, the effect is small. Note that numeric keys are not always the least frequently used in English (Figure 4). The number 1 appears more frequently than letter j, x and z, but people are still better at typing j, x, z than the number 1, suggesting practice cannot overcome the location difference.

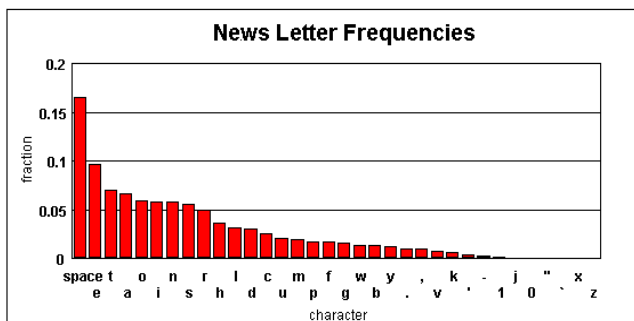


Figure 4. Letter frequency in English, from [17]

Experiment

Experimental Task

We conducted a Chinese typing experiment, in which subjects typed a paragraph that translated to “I’d like to invent a Chinese text input method. In this method a character key is used to replace the numeric keys in selecting the correct choice of characters.” The paragraph has a total of 39 Chinese characters. When typed in a

standard Chinese *pinyin* input method, this paragraph needed 127 alphabet plus 39 numeric keystrokes. The numeric keystrokes were 23.5% of the total keystrokes needed. This is somewhat higher than the 19.2% average, due to the shorter average *pinyin* length in the paragraph (3.26 vs. 4.2 Roman character). The number of page up/down operations needed in typing the paragraph depended on the number of candidates displayed per page. Ten page-down operations were needed for 5 candidates per page; seven or eight for 7 candidates per page.

The paragraph was read by an experimenter at each subject’s typing pace. A special software application was developed and used to record each subject’s input and performance, including the starting and ending time of each paragraph, the keys typed, the eye movement data, page up/page down actions etc. The input was conducted in character mode only.

Experimental Conditions

Four experimental conditions were designed to tease out the relative portions of t_1 to t_4 in the total Chinese text entry time and to evaluate the feasibility of the eye-gaze assisted choice selection method. The four conditions were:

1. *Pure pinyin*. In this condition the subjects translated and typed the Chinese characters they heard to the computer. The purpose of this condition is to establish a baseline with $t_1 + t_2$ only.
2. *Traditional*. This was the standard method of typing *pinyin* then selecting by numeric keying. When the intended character appeared as the first on a page of candidates, regardless which page, the user had the choice of using the spacebar or the numeric key 1 to select it. In this condition, all phases from t_1 to t_4 were involved with the total time being $T = t_1 + t_2 + t_3 + t_4$.
3. *Ideal Gaze Assisted Selection*. In this condition subjects typed *pinyin* and then selected the targeted character by the EASE method. However, no actual eye-tracking was employed. The subjects were asked to press the spacebar when their eyes had located the targeted candidate. This condition simulated the ideal eye-tracking system without the limitation of the current technology – delay and the lack of free head movement. The time spent in this condition is $t_1 + t_2 + t_3 + t_4$. But the value of t_4 is expected to be lower than the t_4 in Condition 2, because it is easier and faster to type the spacebar than a numeric key.
4. *Current Gaze Assisted Selection*. This condition is same as condition 3, except the current EASE system was actually used. The time consumed in this section can be represented as $t_1 + t_2 + t_3 + t_4 + d$. $t_1 + t_2 + t_3 + t_4$ should be similar to condition 3. d represents the time taken to recover the tracking if the subject moved the head out of the tracking area randomly plus the time delay caused by the image processing computation and sampling.

It is possible for the subjects to cheat in condition 3, once they realized there was no real eye-tracking. They could press on the spacebar without visually locating the target. To prevent such cases, three checkpoints were set in the section. At each checkpoint, the correct candidate could be in the second or the third page of the candidate lists. Thus the subjects must page down/up a certain number of times to get the right page. If the experiment program in any of the three checkpoints detected an incorrect number of page down/up operations, the subject was considered cheating and asked to repeat the same section. In the experiment, no cheating was ever detected.

Conditions 2 and 3 were repeated over another independent variable, the length of the candidate list in each screen. Two levels, 5 and 7 candidates in each page, were tested.

Experimental Subjects

Twelve subjects participated in this experiment; ten of them were staff members in the IBM China Research Lab. The other two were undergraduate students in Beijing. The average typing experience of the subjects was 7.75 years, ranging from 2 to 10. The average age was 29. Eleven subjects had no prior experience with eye-tracking devices. All the 12 subjects were familiar with *pinyin* input method. A within subject design was used. Each subject was presented with all four conditions. The order of the conditions presented was counterbalanced in a Latin Square pattern across the 12 subjects

Experimental Results

Figure 5 shows the means of all four experimental conditions with 95% confidence error bars. Data in Condition 2 and 3 were from the sets of data collected with 5 candidates page length. Analysis of variance showed a significant condition effect: $F_{3,33} = 155, p < .0001$. Pair-wise mean comparison (t-tests) showed that time performances between all pairs of conditions were significantly different from each other ($p < .05$) except Condition 2 (traditional method) and Condition 4 (with current gaze tracking).

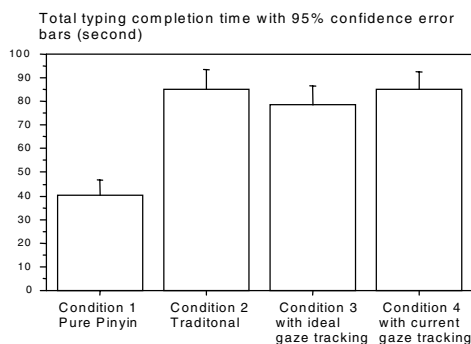


Figure 5. Completion time of the four sections by each subject

In *Condition 1*, the pure *pinyin* condition, the mean completion time was 40.3 seconds, with the fastest at 28 sec and slowest at 65 sec. The average typing speed was hence 317 ms per stroke (ranging 220 ~ 511 ms), which was equivalent to 37.8 (ranging 54 ~ 29) words per minute in English (assuming 5 strokes per English word). These numbers were similar to experiments conducted in English speaking countries (e.g. 32 corrected wpm in [4]). Note that this condition only reflected $t_1 + t_2$ — the *pinyin* generation and keying. The result suggests that time t_1 , the time to cognitively generate the *pinyin* sequence, cannot be a very significant portion of the total time, if we assume the typing skills of our experimental subjects were in the normal range.

The mean completion time in *Condition 2*, which involved the complete traditional Chinese input process ($t_1 + t_2 + t_3 + t_4$), was 85.1 (71 ~ 122) seconds - more than twice the time in Condition 1! This means that $t_3 + t_4$ on average was greater than $t_1 + t_2$, the components equivalent to English typing. As we hypothesized, the multiple-choice selection involved in *pinyin*-based typing is indeed a very serious performance bottleneck.

Condition 3, typing with a simulated (ideal) gaze tracking, had a mean time 78.8 (ranging 64 ~ 106) seconds.

The difference between Condition 2 and 3 was 6.25 seconds. This “saving” in the ideal gaze condition compared to Condition 2 (7.3%) was statistically significant.

Since the only difference between Condition 2 and 3 lay in t_4 , we have

$$39(t_{42} - t_{43}) = 6.25 \text{ sec} \quad (5)$$

where t_{42} and t_{43} are t_4 in Condition 2 and 3 respectively. t_{43} is the time needed to type the spacebar. Given the ease of accessing the spacebar, we can assume t_{43} is in the range of 200 ms, which means

$$t_{42} = 6.25/39 + 0.2 = 0.36 \text{ (second)} \quad (6)$$

360 ms is greater than the average alphabet key stroking time (317 ms). Remember that in Condition 1 subjects were allowed to use either the spacebar or a numeric key when the intended target is the first in the list. In fact, an average of 18 selections out of 39 were made with the spacebar in Condition 1. Due to the nature of the paragraph tested, the characters tended to be the most frequent and hence were the first choice in the list. Considering this, the average time to type on the numeric key could be:

$$(39 \times 360 - 18 \times 200)/(39-18) = 497 \text{ ms.} \quad (7)$$

The estimated value of t_{42} allowed us to calculate a complete estimated decomposition of Condition 2 (traditional Chinese typing), shown in Figure 6. The chart illustrates that generating and typing *pinyin* took only about half the total time. Multiple choice processing took another 36%, which corresponds to about 780 ms for each choice selection. And this is the major slowdown of *pinyin* input.

The remaining 16% was spent on typing the choice selection key.

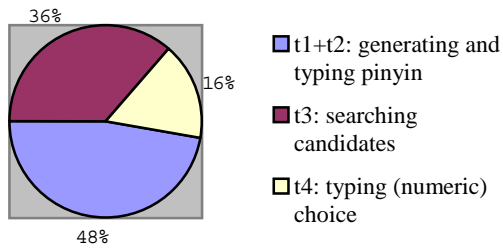


Figure 6. Time composition estimates of pinyin Chinese character input

Condition 4, implemented with the current eye-tracking technology, took about the same amount of time as in Condition 2 (no statistical difference). It appeared that the potential time saving, reflected in Condition 3, was offset by the limitations of the current technology. In other words, d , which corresponds to the time loss due to eye-tracking lag and recovery when the user's head moves out and in the valid tracking area, was relatively large. However, even with these limitations that can be solved in the near future, the system still worked just as well as the pure hand selection method, without the need to type on the numeric keys and hence help the user focus her attention on the computer screen.

We also conducted a 2 x 2 variance analysis of the completion time with regard to Condition (2 vs. 3) and the number of candidates per page (5 vs. 7). Both condition ($F_{1,11} = 10.7, p < .01$) and the number of candidates per page ($F_{1,11} = 28.8, p < .001$) had significant effect on completion time, but the interaction of the two was not significant ($F_{1,11} = .12, p = .73$). On average, the completion time was reduced by 4.8 seconds from 5 to 7 candidates per page. This reduction could be caused by two factors. One is the reduction of the number of paging key strokes (from 10 to 7 or 8 paging strokes, depending on the checkpoints), which should amount to about $3 \times 0.497 = 1.49$ seconds. The rest of the difference had to do with the Hick's law effect when a greater number of smaller choices are replaced smaller number of larger choices. For example, $\log_2(1+7) < \log_2(1+3) + \log_2(1+4)$.

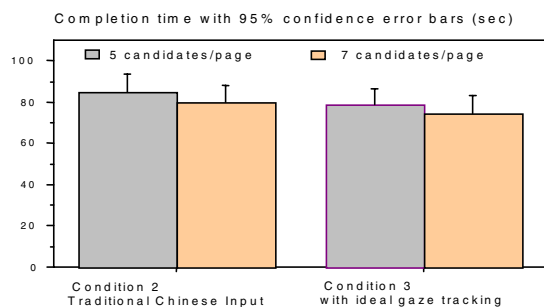


Figure 7. Completion time was shorter with 7 candidates on each page

After the experiment, participants' subjective reactions were collected. Two points stood out. Eleven of the twelve subjects said they liked the elimination of typing on numeric keys in the EASE condition. Over half of the subjects, on the other hand, complained about the inconvenience of head movement restriction imposed on them by the current eye-tracking system.

DISCUSSIONS AND CONCLUSIONS

This study has made two related contributions to Chinese text input. The first is an anatomical analysis of the *pinyin*-based Chinese text input. The second is the exploration of using eye-tracking technology to solve the difficulty of touch-typing numeric selection keys and hence freeing user's visual attention.

The methodology in this study is a combination of experimentation and a fast, first order approximation based, engineering-like modeling approach, most frequently advocated by Card, Moran and Newell [2, 9]. We found that the modeling approach, although not as accurate, gave us a better *understanding of the mechanisms* in the human computer interaction process than detailed direct empirical measurement. Through the combination of two methods, we anatomized Chinese text input into four serial acts. *Pinyin* generation (t_1), *pinyin* typing (t_2), multiple choice reaction (t_3), and selector typing (t_4). In one *baseline* case, we estimated the portions of the total completion time to be 48% ($t_1 + t_2$), 36% (t_3) and 16% (t_4). Such a decomposition clearly points out the performance bottleneck in *pinyin*-based input: more than half the time was spent on looking and selecting the right choice – the components that do not even exist in English typing. We emphasize that this is a *baseline* estimate. In particular, the portion of t_3 may decrease if words or phrases, in addition to characters, were used as a unit to reduce the number of choices. Note also that the paragraph used in the experiment had a shorter *pinyin* length, which means that the *pinyin* typing portion (t_2) is smaller than the average.

We found that the primary problem with t_4 is the numeric keys used for choice selection. These keys make it more difficult for the user to be a complete touch typist and hence diverts the user's attention from the information on the computer display. Our proposed EASE solution, which used a combination of eye-tracking and a common selection key for multiple choice selection, could solve the attention diversion problem without time performance cost. Our experiment showed if the time delay and the head movement constraint in today's eye-tracking technology could be removed, the EASE solution could even save part of the t_4 .

Other than motivating the design of the EASE system, our anatomical analysis of Chinese input also points to the directions of future research and design efforts in Chinese input. For example, the cognitive load of "translating" Chinese into *pinyin*, a spelling system not naturally used

outside of primary school teaching and computer input, is not a major bottleneck to performance, but reducing the frequency and number of multiple choices in the input process is critical.

The implication of our study also goes beyond Chinese input. First, the lessons learned in here may benefit text input system designs in many other non-alphabetic languages. Second, the issue of frequent multiple-choice selection is not limited to text input. The idea of avoiding the visual attention distraction of the numeric keystrokes by means of an EASE-like approach may be worth considering when frequent multiple-choice selection is needed.

In summary, the study 1) provided an anatomical analysis of today's standard Chinese input process; 2) identified the most important problems in today's system; 3) addressed the attention diverting problem in Chinese typing by designing and implementing the EASE system, which generalized the idea of implicit use of eye-tracking in MAGIC pointing; 4) found the weaknesses of today's eye-tracking technology.

Much more work remains to be done in the future. The modeling work needs to be more generalized and more accurate. Richer experimental data should be collected and analyzed in finer granularity. Better eye-tracking systems need to be developed, which are feasible in the near future given the rapid increase in computing power. Furthermore, it is possible to build an intentional model to interpret the gaze data in the context of Chinese character selection to compensate the time delay in the system. The use of the EASE approach can also be more general. Specifically, the paging key in scrolling the candidate characters can be implemented in the same approach by displaying a "next" and a "previous" icon. When the user's eye does not find the right choice through the current list of candidates and goes to the "next" icon, pressing on the same common selecting key (e.g. spacebar) should bring up another page of candidates.

ACKNOWLEDGMENTS

We would like to thank Alison Sue, Qian Wu, Qianying Wang, Teenie Matlock, and Dan Russell for their advice and assistance.

REFERENCES

1. Bolt, R.A. Eyes at the interface. In *Proc. of Human Factors in Computer Systems*. 1982. Gaithersburg, Maryland: ACM p. 360-362.
2. Card, S.K., T.P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*. 1983, Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.
3. Chen, Y., *Chinese Language Processing*, 1997, Shanghai Education publishing company.
4. Jacob, R.J.K. What You Look At is What You Get: Eye Movement-Based Interaction Techniques. in *Proc. of CHI'90: ACM Conference on Human Factors in Computing Systems*. 1990: Addison-Wesley/ACM Press p. 11-18.
5. Karat, C.M., C. Halverson, D. Horn, and J. Karat. Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *Proc. of CHI'99: ACM Conference on Human Factors in Computing Systems*. 1999. Pittsburgh, PA, USA p. 568-574.
6. Lexicon of Contemporary Chinese, Third version, ISBN 7-100-01777-7/H.519, Shangwu Publishing House, 1996
7. Levine, J.L., *An Eye-Controlled Computer*, 1981, IBM TJ Watson Research Center: Yorktown Heights, New York.
8. Morimoto, C, D. Koons, A. Amir and M. Flickner, Pupil detection and tracking using multiple light sources, *Image And Vision Computing*, Volume 18, Issue 4.
9. Newell, A., S. Card, The prospects for psychological science in human-computer interaction. *Human-Computer Interaction*, 1985. 1: p. 209-242.
10. Salvucci, D., Inferring Intent in Eye-Based Interfaces: Tracing Eye Movements with Process Models. . In CHI '99: ACM Conference on Human Factors in Computing Systems. 1999, ACM Press
11. Ware C. and H.H. Mikaelian. An evaluation of an eye tracker as a device for computer input. in *Proc. of CHI+GI: ACM Conference on Human Factors in Computing Systems and Graphics Interface*. 1987. Toronto p. 183-188.
12. Welford, A.T., *Fundamentals of Skill*. 1968, London.
13. Yamada, H., A historical study of typewriters and typing methods: from the position of planing Japanese parallels. *Journal of Information Processing*, 1980. 2(4): p. 175-202.
14. Zhang, S., *Intelligent Approach: From Pinyin to Chinese Characters*. Language Engineering, Tsinghua University Press 1997.
15. Zhang, X., *Intelligent Chinese Pinyin-Character Conversion Based on Phrase Analysis and Dynamic Semantic Collocation*. Language Engineering, Tsinghua University Press 1997
16. Zhai, S., C. Morimoto, and S. Ihde. Manual And Gaze Input Cascaded (MAGIC) Pointing. In *Proc. of CHI'99: ACM Conference on Human Factors in Computing Systems*. 1999: ACM Press p. 246-253
17. Zhai, S., Hunter, M., Smith, B.A., The Metropolis Keyboard -- An Exploration of Quantitative Techniques for Virtual Keyboard Design, In *Proc. of UIST 2000*, p. 119-12