Predicting Chinese Text Entry Speeds on Mobile Phones

Ying Liu
Nokia Research Center
No.5 Donghuan Zhonglu, BDA Area
Beijing 100176, China
ying.y.liu@nokia.com

Kari-Jouko Räihä
Unit for Computer-Human Interaction (TAUCHI)
Department of Computer Sciences
FIN-33014 University of Tampere, Finland
kari-jouko.raiha@cs.uta.fi

ABSTRACT
Chinese text entry on mobile phones is critical considering the large number of Chinese speakers worldwide and as a key task in many core applications. But there is still a lack of both empirical data and predictive models that explore the pattern of user behavior in the process. We propose a model to predict user performance with two types of Chinese pinyin input methods on mobile phones. The model integrates a language model (digraph probability) with Fitts' law for key presses, a keystroke-level model for navigation, and a linear model for visual search in pinyin marks and Chinese characters. We tested the model by comparing its predictions with the empirical measures. The predictions are satisfactory and the percentage differences are all within 4% of the empirical results, suggesting that the model can be used to evaluate user performance of Chinese pinyin text entry solutions on mobile phones.

Author Keywords
Chinese, text entry, mobile, performance, model, speed.

ACM Classification Keywords
H5.2. User Interfaces: Evaluation/methodology, theory and method.

General Terms
Performance, Theory

INTRODUCTION
Chinese text entry solutions on mobile phones are important. China is the world's largest single market for mobile phones. By May of 2009, there were 687 million mobile phone users in mainland China [27], many of whom also use short message services (SMS). SMS use has seen a tremendous increase in China since its release in 2000 [15]. In May of 2009, Chinese users exchanged more than 64 billion short messages, with over 3 messages per user per day [27]. Moreover, forthcoming mobile applications including email, instant messaging and office applications also involve the task of entering Chinese characters on mobile phones.

Although there are many novel devices designed to improve user performance on text entry tasks, the 12-key keypad is still the dominant input device because of its familiarity, and also because its compact size is suitable to be held and used with one hand. Chinese is an ideographic language and its characters cannot be entered directly with Roman keyboards. Thus two types of coding systems have been invented to map the unlimited Chinese characters to the 12-key keyboard: the pinyin coding system based on the Mandarin pronunciations and the stroke coding system based on a standard stroke order [12, 13, 14]. The pinyin coding system is the most widely used in both personal computers and mobile phones.

In this paper, we present a predictive model that estimates the average text entry speed of users' error-free Chinese pinyin input on mobile phones. Two facts motivated us to build the model. First, considering the large number of Chinese speakers around the world, published studies on user performance of mobile Chinese pinyin input are relatively few in number, and results of the limited number of studies often contradict each other. Lin and Sears reported a text entry speed of 4.04 words per minute (WPM[1]) when the participants naturally balanced between input speed and error rate [12]. Liu and Wang reported user speeds of 15 to 34 WPM for Chinese pinyin input supporting phrases [14]. Second, there is no predictive model built to estimate user performance or understand mechanisms involved in the Chinese text entry tasks on mobile phones. A predictive model can assist researchers and practitioners to compare different text entry solutions without conducting tedious empirical studies and identify the improvement areas and design opportunities [1, 3, 4, 6, 9, 11, 17, 18, 19, 24, 25].

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

[1] WPM refers to Chinese characters per minute in this paper.

CHINESE PINYIN TEXT ENTRY ON MOBILE PHONES

Pinyin is the standard coding system of Mandarin pronunciation in the form of Roman letters [12, 13, 14, 20, 28]. A pinyin mark, whose length varies between one and six letters, usually consists of a consonant and a vowel, with the exception of a few marks that consist of vowels alone (See Table 1). Pinyin is the primary coding system that can be applied to nearly all types of keyboards including physical and soft keyboards, the QWERTY keyboard for computers and the 12-key keypad for mobile phones. Moreover, text entry solutions based on the pinyin coding system are the primary methods that users in mainland China are using on both personal computers and mobile phones [13, 14].

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

Table 1. The 23 consonants and the 33 vowels of pinyin.

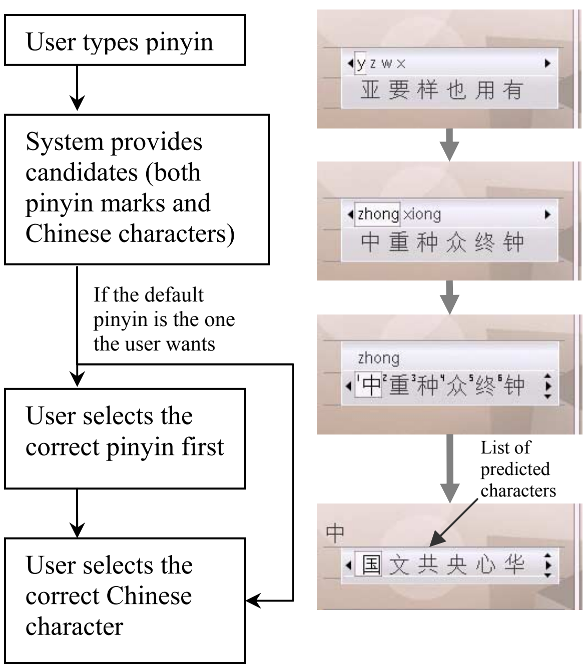

Figure 1. The input process of Chinese T9 Pinyin input with the 12-key keypad.

Entering Chinese characters with the T9 pinyin input on mobile phones requires three steps (see Figure 1). First, users press keys to enter a pinyin mark. Second, users need to select the target pinyin mark from a few options, because 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 highlighted pinyin mark happens to be the target one, users can click the "ok" key to choose it. Otherwise, users need to move the highlight to the target pinyin mark and choose it. Third, since most Chinese characters are homophonic with several others, users need to select the target character from a list of options sharing the same pinyin mark.

Most Chinese pinyin input systems also support predictive input based on Chinese phrases (referred to as "the predictive feature" in the rest of the paper). A phrase consists of one or more Chinese characters and is the smallest meaningful unit in Chinese language [14]. Once a Chinese character is entered, the next character can be predicted based on relevant phrases. For example, after the character "中" is entered, characters that can be combined with it to form phrases are predicted and listed by the input system (see the last image in Figure 1). Users can choose the target character from the list without typing its pinyin mark.

MODEL FOR SUBTASKS OF CHINESE TEXT ENTRY

According to the task analysis in the previous section, there are two subtasks in the Chinese pinyin input process on mobile phones: the retrieval and typing of a pinyin mark and the disambiguation task for selecting the target pinyin mark and Chinese character. Thus we built the predictive model as expressed below:

$$T = T_m + T_d \qquad (1)$$

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 because we 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 highlight 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 from the first items since they are more likely to be the intended ones than the last ones. Often it is not necessary to search the whole list before the desired mark or character is found. Thus we split the disambiguation process into two subprocesses: visual search and navigation. Consequently, T_d includes times for visual search (T_v) and navigation (T_n):

$$T_d = T_v + T_n \qquad (2)$$

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

Based on this model, we applied several theories to predict the times for the subtasks. These theories are presented in the next three subsections. First, we present the corpus used for building the language model. Second, we present the movement model by combing Fitts' law and the language model (T_m). Third, we present the keystroke-level model (KLM) to predict the average navigation time (T_n). Finally, we present the linear model to estimate the average visual search time (T_v).

The Corpus and Language Model

Since the language model is for text entry on mobile phones, we collected a corpus[2] of 630,000 text messages that contained a total of 9,200,000 Chinese characters (these statistics exclude punctuation marks). After analysis, we found that the corpus contained 4,912 different Chinese characters, which corresponds to 404 syllables or pinyin marks. By comparison, the GB2312, a standard Chinese character set for simplified Chinese supported by most Chinese text entry systems, includes 6,763 single Chinese characters with 404 different syllables that cover about 99.75% of all Chinese characters [7]. Since our corpus is a collection of text messages entered with Chinese text entry solutions, the characters covered by it are a subset of the GB2312. However, our corpus is more representative of the current mobile text entry context than the GB2312 which was defined in 1980 for more general purposes.

Based on the corpus, we built the language model. First, we transcribed the Chinese characters to corresponding pinyin marks and calculated the frequencies of all pinyin marks. Based on the frequencies of pinyin marks, we calculated the digraph probabilities for each legal pair of letters in pinyin. The linguistic model resulted in a 26×26 matrix of letter pair frequencies. 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 one.

To predict the average time for typing a pinyin mark, we still need to know the average number of letters per Chinese character. In our corpus the result is 3.24, but if we take into account the frequencies of the Chinese characters, the figure drops to 2.88. Both numbers are smaller than 4.2, which was believed to be the average number of Roman characters per Chinese character in [29].

Most Chinese characters are homophonic with other characters [13, 14, 20], thus a pinyin mark usually corresponds to multiple Chinese characters. In the corpus, it was found that a pinyin mark corresponded to a minimum of 1 character and a maximum of 74 Chinese characters. In most pinyin input systems, the character options are listed according to their usage frequencies: characters with higher frequencies listed before characters with lower frequencies. Based on this, we calculated the average position of all characters in characters with the same pronunciations. The result was 1.77.

[2] The corpus includes two parts: one part was licensed from a third party and the other part was collected by the authors with user agreements.

The 12-key keypad can also produce ambiguous results. A series of key presses sometimes results in multiple pinyin marks that are listed according to their using frequencies. We also calculated the average position for all pinyin marks. The result was 1.24.

We also examined the phrases in the corpus. Figure 2 shows the proportions of phrases that consist of different numbers of Chinese characters taking into account their frequencies in our corpus. Single characters make up the largest proportion of the corpus, followed by the 2-character and 3-character phrases.

Once the first character of a phrase is entered, the other characters can be entered by directly selecting them one by one from the prediction lists. We calculated the proportions of characters that could be entered by the predictive feature and the result was 30.3% for the corpus. Since the predicted characters are also listed according to their usage frequencies, the calculated average position for all predicted characters was 2.60.

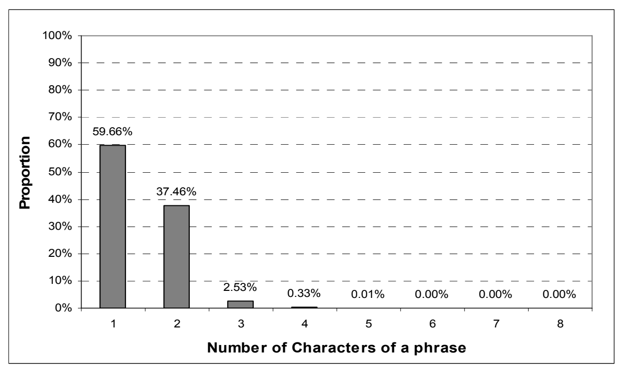

Figure 2. Proportions of phrases with different numbers of characters.

Movement Model

The movement model was built based on a combination of Fitts' law and the language model. Fitts' law was inspired by information theory and applied by psychologists to predict movement time of the human motor system [22, 26]. 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:

$$MT = a + b \log_2 (A/W+1) \qquad (3)$$

In Equation 3, A represents the amplitude of motor movement and W is the size of the target object, which is usually indicated by its width [22, 26] or for two-dimensional objects, height or width, whichever is smaller [16, 24]. Constants a and b are defined by empirical experiments. The log term of Equation 3 is also called the "index of difficulty" (ID).

Fitts' law has been widely applied in predicting expert user performance with different types of input devices including physical [18, 24] and soft keyboards [9, 17, 25], pointing devices [26, 31] and tasks including text entry [9, 10, 11, 17, 18, 19, 24, 25] and menu selection [3, 4]. When applying Fitts' law to predict user performance with different tasks, a basic assumption is that the users are experts of the tasks, for whom motor movements cost the majority of their effort and time while cognitive processes cost little.

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

$$T_m = 2.88 \times \sum (P_{ij} \times MT_{ij}) \quad (4)$$

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,* to be entered together, and MT_{ij} is the average time needed to move to the thumb from "i" to the "j" key and press it.

KLM for Navigation Time in the Disambiguation Process

We applied KLM, the simplest GOMS model [1, 2], 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 display with a mouse; H for moving hands to the home position on keyboard or mouse; D to draw a line; M for mentally preparing to do an action or closing associated primitive actions; and R as the system response time that users had to wait for. By splitting a specific task to such operators and defining times for relevant operators, analysts can estimate the time required by a skilled user to complete the task without error.

Dunlop and Crossan applied KLM to compare user performance of the multi-press method with that of the predictive methods [5, 6]. In their KLM, they chose three operators and defined fixed times for them: K for 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 we predicted the navigation time in the disambiguation process, we simply chose operator K. Moreover, instead of using a fixed time for operator K, we applied Fitts' law to estimate times for different key presses.

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

$$T_n = \sum\sum (P_{mn} \times T_{n\_mn}) \quad (5)$$

Here T_{n_mn} is the specific navigation time based on KLM for cases where the target pinyin marks and Chinese characters are located respectively at positions m and n. However, such a calculation of T_n is rather complex and hard to apply. First, according to the corpus, there are a total of 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:

$$T_{n\_mn} = K_{io} + K_{rr} \quad (6)$$

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

$$T_{n\_mn} = K_{ir} + 2K_{ro} + K_{or} \quad (7)$$

In the above equations, K_{io} and K_{ir} are respectively the average time to move the thumb from the last letter of pinyin marks to the "ok" or the right navigation keys and press them. 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 8, where $P_{io/r}$ is the overall probability of the "i" key being the last letter of pinyin marks and $MT_{io/r}$ is the movement time for moving the thumb from the "i" key to the "ok" or the right navigation key.

$$K_{io/r} = \sum (P_{io/r} \times MT_{io/r}) \quad (8)$$

To make the calculation of T_n simpler and easier to apply, we approximated T_n by $T_{n\_1,2}$, i.e., by setting m and n to 1 and 2, respectively, because according to the corpus

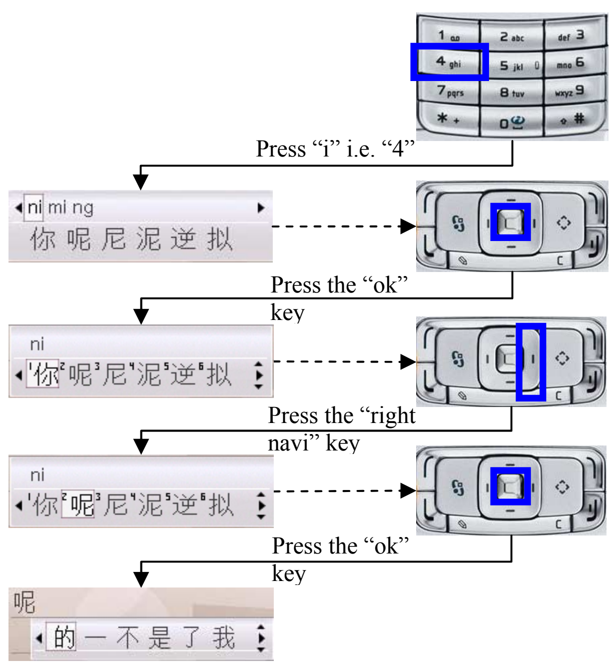

Figure 3. The disambiguation process when *m* equals 1 and *n* equals 2.

analysis the average positions for pinyin marks and characters are 1.24 and 1.77, respectively, and thus close to 1 and 2. Figure 3 shows the navigation process with a target pinyin mark at the 1st position and a Chinese character at the 2nd position. With this simplification, T_n can be expressed as:

$$T_n = K_{io} + K_{or} + K_{ro} \qquad (9)$$

As we 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:

$$T_{n\_pre} = K_{or} + (2.6 - 2) \times K_{rr} + K_{ro} \qquad (10)$$

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

The Visual Search Model

Visual search is the process of finding a target item among distractor items [30]. Thus 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 mechanism [30]. 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 based on the slope of reaction times by set size. If the slope is close to zero ms per item, it is usually regarded as a parallel search. However, if the slope is steep, it is regarded as a serial search.

In the human-computer interaction field, the Hick-Hyman law had been applied to estimate visual search time [3, 4, 25]. Similar to Fitts' law, the Hick-Hyman law was also inspired by information theory. It has been applied to estimate the reaction time of making a choice from a number of possibilities. However, compared to Fitts' law, it has not been as widely applied in HCI [22]. It is usually expressed as follows:

$$RT = c + d \log_2 (n + 1) \qquad (11)$$

Here RT is the decision making time to choose a response from a number of possibilities in accord 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 applied the Hick-Hyman law to estimate visual search time with soft keyboards [25]. However, Sears et al. [21] 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 like familiarity should be taken into account when predicting user performance of visual search. Cockburn et al. [3, 4] argued that when people could anticipate the location of items, the Hick-Hyman law was proper to predict the time of acquiring the target; but when the anticipation was not possible, a linear model should be applied.

We 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: thus the linear model might 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:

$$T_v = \sum\{P_m \times [e_{py} + f_{py} \times (m-1)]\} + \sum\{P_n \times [g_{cc} + h_{cc} \times (n-1)]\} \quad (12)$$

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

$$T_v = e_{py} + f_{py} \times [\sum(P_m \times m) - 1)] + g_{cc} + h_{cc} \times [\sum(P_n \times n) - 1)] \qquad (13)$$

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

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.

EXPERIMENTS

We conducted two experiments to define parameters for the model and one experiment to evaluate the model. Experiments I and II were to define the parameters for Fitts' law and the linear models for visual search in pinyin marks and Chinese characters. Experiment III was to collect and compare user speeds with the predicted speed.

Experiment I: Parameters for Fitts' Law

Participants

Twelve volunteers (7 male, 5 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 in Nokia Research Center in Beijing. All were right-handed and held the phone in their dominant hands in the experiment. All were regular phone users.

Apparatus
A Nokia N95 was used in the experiment. We built for the experiment 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 short message application in the device. We chose a Nokia N95 because it has a high-speed processor and can ensure the accuracy of the time logs.

Figure 4. The Nokia N95 used in the experiment.

Test Tasks

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

i) Repeated key presses: participants were to press a key continuously and 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 we 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"). Heights of the keys that serve as W in Fitts' law to calculate 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, we 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 key press equaled the difference of the third logged time and the second logged time.

Procedure
The experiment was conducted in a quiet 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 they were 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 counter-balanced among the twelve 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.

Experiment II: Visual Search Time of the Disambiguation Processes
The experiment sought to define parameters for the linear visual search models (e_{py}, f_{py}, g_{cc}, and h_{cc} in Equation 13) for pinyin marks and Chinese characters.

Participants
Twenty four volunteers (16 male, 8 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 staff in Nokia Research Center 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
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 (Figure 5).

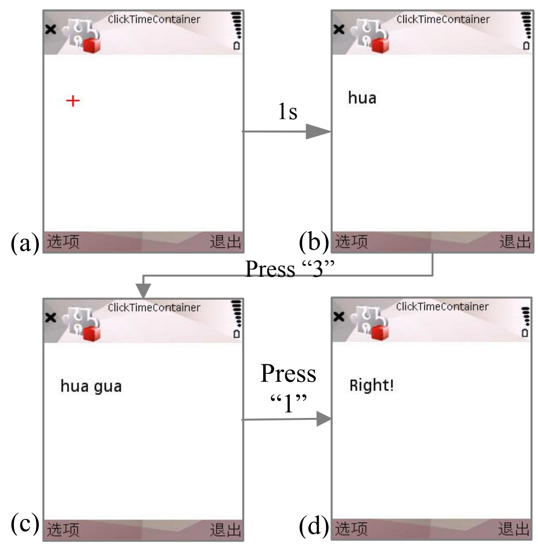

Figure 5. Screen shots for experiment III: a) at the beginning of a trial, a red cross was displayed for 1 s to attain the user's attention; b) a target item was displayed at the same location and participants were to remember it and press "3" to go to the next step; c) a list of options appeared for users to decide whether the target appeared or not (target present: press "1"; target absent: press "3"); d) after participants' responses, the system gave feedback.

Tasks and Materials
We applied the experiment paradigm used by psychologists to study the visual search process [30]. At the beginning of a trial, a red cross was displayed for 1 second to attain the attention of the user. Then a target item (either a pinyin

mark or a Chinese character) appeared at the same location. Once participants recognized and remembered the target, they were to press the "3" key. After the key press, a list of options (2 to 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 using the software.

There were three types of trials in the experiment: 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 systems and there is no existing study exploring the differences in user performance. However, Chinese characters were just listed horizontally since past studies had already proven that people are more efficient searching for Chinese characters in horizontal lists than in vertical lists [8] and moreover, Chinese characters are listed horizontally in many existing pinyin input solutions in products.

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

Procedure
The procedure in this experiment was the same with that of Experiment I except that the training session of Experiment II included 30 trials.

Experiment III: Empirical Text Entry Speeds
The experiment was carried out to collect empirical data to compare with the predicted text entry speeds.

Participants
Twelve volunteers (8 male, 4 female) took part in the experiment. Their ages ranged from 24 to 32 years, with an average of 27.9 years (*SD* = 3.06). All were either student interns or staff in 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 had used mobile phones for 5.3 years on average (*SD* = 1.21).

Apparatus
Nokia N95 was the device used in the study. We used the program designed for experiment I 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 a total of 31 characters in the two text messages and about 7 characters (about 23%) could be entered with the predictive feature. The single-letter correlation of the two text messages with the corpus was 0.932. The average number of letters for the 31 characters was 2.94. The average positions for pinyin marks and Chinese characters were 1.10 and 1.84.

Procedure
The procedure in this experiment was the same with those of Experiment I and II except that in the training session of this experiment, the participants were instructed to enter three short messages consisting of 39 characters.

RESULTS

Experiment I
Table 2 and Figure 6 show the results of experiment I. In Figure 6, the diamond marks indicate the average reaction times for different IDs and the line is the linear regression that we conducted of the average reaction time by ID.

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

Table 2. Parameters *a* and *b* for Fitts' law.

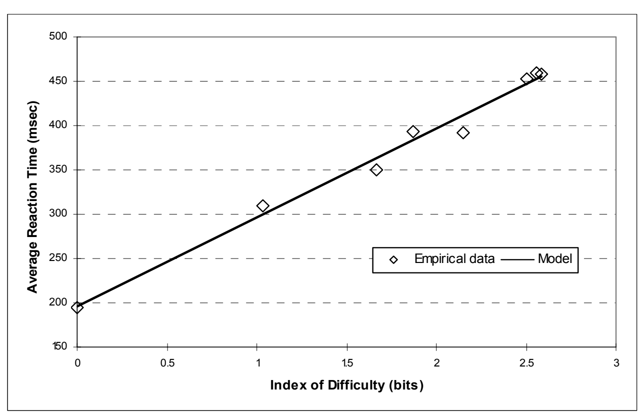

Figure 6. Regression of Reaction Time by ID.

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 2 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.

Based on the results, we calculated T_m according to Equation 4, K_{io} based on Equation 8, and K_{or} and K_{ro}, which respectively equaled 1059 ms, 456 ms, 321 ms and 285 ms. Based on Equation 9, we calculated the average navigation time to enter a predicted character, which equaled 723 ms.

Experiment II

Participants made few errors in the experiment, with an average error rate of 1.4% (*SD* = 0.92%). When we analyzed the data, all error trials were excluded. Figure 7 and Table 3 show the results of experiment II, indicating that linear models are proper to model the visual search task [26]. Moreover, the visual search in pinyin marks and Chinese characters is characteristic of both serial and self-terminating searches: searches stop once the target is identified.

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

Table 3. The linear models from Experiment II.

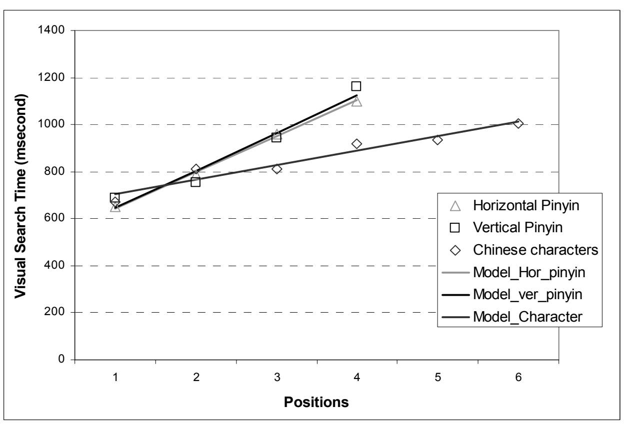

Figure 7. The empirical results and the linear models for visual search.

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 the target position ($F_{3,\ 287}$ = 71.8, $p < .001$) and the 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).

When we estimated visual search times, we did not directly apply the results of experiment II. This was because the task of the experiment included a subprocess of choice reaction from two reaction options (or decision making) besides visual search. Thus we decided to remove a fixed time for the choice reaction from the linear models. According to Sears et al. [21], the choice reaction time should be calculated based on the number of possible reactions instead of the number of stimuli. We applied the results of Hick's experiment on choice reaction and subtracted a fixed time of 247 ms ($n = 2$) from the linear models [8, 22]. We 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 and are presented in Table 4.

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

Table 4. The parameters for visual search.

Based on Table 4 and Equation 13, the average visual search time T_v was 938 ms. For characters that were entered with the predictive feature, the average visual search time was 556 ms.

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

Table 5. Average time required to enter a Chinese character.

Table 5 summarizes the average time of entering a Chinese character with the predictive feature off (3059 ms) and with the predictive feature on (1279 ms) and they are respectively represented as T and T_{pre} in Equations 14 and 15. Equations 14 (S_{pre_off}) and 15 (S_{pre_on}) show how we calculate the predicted speeds when the predictive feature is off and on. In Equation 15, P_{pre} is the overall percentage of characters entered by the predictive feature only. The predicted speeds are presented in Table 6.

$$S_{pre\_off} = 60 / (T/1000) \quad (14)$$

$$S_{pre\_on} = 60 / \{[T\times(1 - P_{pre}) + T_{pre}\times P_{pre}] / 1000\} \quad (15)$$

Experiment III

To be comparable with text entry rates of other languages, we used WPM to present results on text entry rate. Time of all extra key presses, for example, for making and clearing errors were all removed from the task completion time to make sure the calculated text entry rates are comparable with the predicted ones.

		Average user speeds with *SD*	Pre-dicted speeds
Predictive feature off		19.1 (2.32)	19.6
Predictive feature on and different percentages of characters entered with it	23%	21.9 (2.28)	22.6
	30.3%	----	23.8

Table 6. Average user speeds and predicted speeds (WPM).

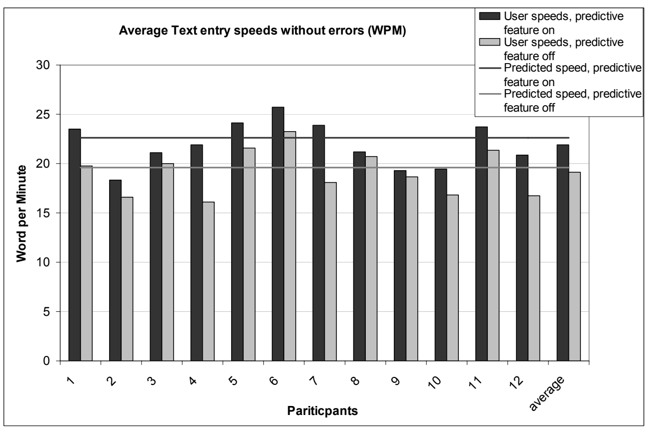

Figure 8. Text entry speeds without errors.

Table 6 and Figure 8 show the results of the average user speeds and the predicted speeds. The average user speed was 19.1 WPM (*SD* = 2.32), when the predictive feature was off. When the predictive feature was on and 23% of characters were entered with it, the average user speed was 21.9 WPM (*SD* = 2.28). The empirical data and the predictions match each other well. The percentage differences of the predicted speeds were both within 4% of the empirical data: respectively 2.6% and 3.2% while the predictive feature was 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 14.7% based on empirical data and 15.3% based on the predictive model.

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 big individual differences among the participants on text entry speeds. It is worth noting that the two facts are not in contradiction with each other. Since the constants in our models were calculated based on average user performance with subtasks, the predicted speeds were also averages of user performance.

It was expected that the model, with the characteristics of the corpus, be applied together to evaluations of Chinese text entry solutions on mobile phones; now we have only carried through the process with 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 respectively 69.2% and 65.4% of the total time with the pinyin input methods on mobile phones when the predictive feature was on and off, indicating a higher percentage of time than the disambiguation process of a pinyin input method based on a QWERTY keyboard takes (about 52%) [29]. Innovative solutions are needed to optimize the disambiguation process in Chinese text entry solutions on mobile devices.

When we calculated the predicted speed while the predictive feature is on as Equation 15 shows, we directly used for T the average time (3059 ms) calculated based on all characters of the corpus. However, the characters that were entered by the predictive feature should have been removed from this number. We assume that this was one reason why we found a slightly faster predicted speed for cases with the predictive feature. This is also one of the improvement points that we need to address in our future work.

Compared with the state of the art predictive models on text entry and menu selection tasks [3, 10, 19], our model does not cover some variables like the learning process or new features like phrase input. But we view the present work as a good start and anticipate that such issues will be explored in the future.

CONCLUSIONS

Chinese text entry on mobile phones is critical considering the large amount of users and as a key task in many core applications. We 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. We evaluated the model by comparing its predictions with the empirical user speeds. The predictions were satisfactory: when the predictive feature was on and off, the predicted speeds were respectively 3.2% and 2.6% higher than the empirical user speeds.

We view the model as a useful start. First, there is a lack of predictive models or insights to explore the pattern of user behaviors in the Chinese text entry systems of mobile phones. Being the first published model, we hope our work can draw further work on this subject. Second, the practitioners and researchers in this area can benefit from the model by applying it in evaluations of Chinese pinyin text entry solutions on mobile phones and identification of design opportunities.

ACKNOWLEDGMENTS

We thank Ning Liu, Yuan Feng, Xia Wang and Zhen Liu for their support. We also thank Poika Isokoski and Scott MacKenzie for sharing their knowledge. The work of the second author was supported by grant 1130044 from the Academy of Finland to visit the University of Canterbury in Christchurch, New Zealand.

REFERENCES

1. Card, S.K., Moran, T.P., and Newell, A. The keystroke-level model for user performance time with interactive systems. *Communications of the ACM 23*, 7 (1980), 396-410.
2. Card, S.K., Moran, T.P., and Newell, A. *The Psychology of Human-Computer Interaction*. Erlbaum, 1983.
3. Cockburn, A., Gutwin, C., and Greenberg, S. A predictive model of menu performance. In *Proc. CHI 2007*, ACM Press (2007), 627-636.
4. Cockburn, A., and Gutwin, C. A predictive model of human performance with scrolling and hierarchical lists. *Human-Computer Interaction 24*, 3 (2009), 273-314.
5. Dunlop, M.D., and Crossan, A. Predictive text entry methods for mobile phones. *Personal Technologies 4*, 2 (2000), 134-143.
6. Dunlop, M.D., and Masters, M.M. Investigating five key predictive text entry with combined distance and keystroke modeling. *Personal and Ubiquitous Computing 12*, 8 (2008), 589-598.
7. GB 2312-1980. Code of Chinese graphic character set for information interchange--Primary set (1980).
8. Hick, W.E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology 4*, 1 (1952), 11-26.
9. Isokoski, P. Performance of menu-augmented soft keyboards. *Proc. CHI 2004*, ACM Press (2004), 423-430.
10. Isokoski, P., and MacKenzie, I.S. Combined model for text entry rate development. *Ext. Abstracts CHI 2003*, ACM Press (2003), 752-753.
11. James, C.L., and Reischel, K.M. Text input for mobile devices: Comparing model predictions to actual performance. *Proc. CHI 2001*, ACM Press (2001), 365-371.
12. Lin, M., and Sears, A. Constructing Chinese characters: keypad design for mobile phones. *Behaviour & Information Technology 26*, 2 (2007), 165-178.
13. Liu, Y., and Räihä, K.-J. RotaTxt: Chinese pinyin input with a rotator. *Proc. Mobile HCI 2008*, ACM Press (2008), 225-233.
14. Liu, Y., and Wang, Q. Chinese pinyin phrasal input on mobile phone: Usability and developing trends. *Proc. Mobility 2007*, ACM Press (2007), 540-546.
15. Ma, D., Ichikawa, F., Liu, Y., and Jiang, L. Use of Chinese short messages. *Proc. HCI International 2007*, Springer (2007), LNCS 4558, 582-591.
16. MacKenzie, I.S., and Buxton, W. Extending Fitts' law to two-dimensional tasks. *Proc CHI 1992*, ACM Press (1992), 219-226.
17. MacKenzie, I.S., and Soukoreff, R.W. Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction 17*, 2&3 (2002), 147-198.
18. Myung, R. Keystroke-level analysis of Korean text entry methods on mobile phones. *International Journal of Human-Computer Studies 60*, 5-6 (2004), 545-563.
19. Pavlovych, A., and Stuerzlinger, W. Model for non-expert text entry speed on 12-button phone keypads. *Proc. CHI 2004*, ACM Press (2004), 351-358.
20. Sacher, H., Tng, T.-H., and Loudon, G. Beyond translation: approaches to interactive products for Chinese consumers. *International Journal of Human-Computer Interaction 13*, 1 (2001), 41-51.
21. Sears, A., Jacko, J.A., Chu, J., and Moro, F. The role of visual search in the design of effective soft keyboards. *Behaviour & Information Technology 20*, 3 (2001), 159-166.
22. Seow, S.C. Information theoretic models of HCI: A comparison of the Hick-Hyman law and Fitts' law. *Human-Computer Interaction 20*, 3 (2005), 315-352.
23. Shih, H.M., and Ravindra, S.G. Effectiveness of menu orientation in Chinese. *Human Factors 40*, 4 (1998), 569-576.
24. Silfverberg, M., MacKenzie, I.S., and Korhonen, P. Predicting text entry speed on mobile phones. *Proc. CHI 2000,* ACM Press (2000), 9-16.
25. Soukoreff, R.W., and MacKenzie, I.S. Theoretical upper and lower bounds on typing speed using a stylus and a soft keyboard. *Behaviour & Information Technology 14*, 6 (1995), 370-379.
26. Soukoreff, R.W., and MacKenzie, I.S. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, 61, 6 (2004), 751-789.
27. Statistics from Ministry of Industry and Information Technology of China. http://www.miit.gov.cn/n11293472/n11293832/n11294132/n11302706/12426281.html (in Chinese)
28. Tanaka-Ishii, K., Zhou, M., and Kim, J.-D. Text entry in East Asian languages. In I.S. MacKenzie and K. Tanaka-Ishii (eds.), *Text Entry Systems: Mobility, Accessibility, Universality*, pp. 203-225. Morgan Kaufmann, 2007.
29. Wang, J., Zhai, S., and Su, H. Chinese input with keyboard and eye-tracking: an anatomical study. *Proc. CHI 2001*, ACM Press (2001), 349-356.
30. Wolfe, J.M. Visual search. In H. Pashler (ed.), *Attention,* pp. 13-72. Psychology Press, 1998.
31. Zhai, S. Characterizing computer input with Fitts' law parameters: the information and non-information aspects of pointing. *International Journal of Human-Computer Studies 61*, 6 (2004), 791-809.