Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Int. J. Human-Computer Studies 58 (2003) 483–508

International Journal of
Human-Computer
Studies

www.elsevier.com/locate/ijhcs

Human performance modeling in temporary segmentation Chinese character handwriting recognizers

Changxu Wu[a,*], Kan Zhang[a], Yongge Hu[b]

[a] *Engineering Psychology and Human Factors Laboratory, Institute of Psychology, Chinese Academy of Sciences, Beijing, PR China*
[b] *Intel Research China Center, Beijing, PR China*

Received 5 January 2003; accepted 18 January 2003
Paper accepted for publication by Editor, B. Gaines

Abstract

Human performance in Chinese character handwriting recognizers is critical to the satisfaction and acceptance of their users. Based on Teal's [CHI'92 (1992) p. 295] interactive model, a static model describing the independent factors in determining the task completion time was set up with a simple mathematical inference; in addition, a dynamic model describing these factors' direct and indirect causal relationship was established by the path analytic method. Results in Experiment 1 indicated that both the static model and the dynamic model could fit observed task completion time satisfactorily with minor modifications. In addition, with users' average writing time around 1500 ms for each frequently used character, it was found that the user's performance was impaired significantly when segmentation time was longer than 1040 ms. An integrated model was devised after combining the static and dynamic models. Experiment 2 testified the integrated model in another handwriting recognizer and found that it could still fit human performance data with users in three different training conditions. Implications of the integrated model are that: (1) when recognition accuracy and number of inputting characters are constant, the weights of average writing time for each character, segmentation time, recognition time in determining task completion time are equal but bigger than the weight of the repairing time; (2) when the repairing time, average writing time for each character, segmentation time and recognition time are constant, there is an inverse model between task completion time and recognition accuracy; when recognition accuracy is from 50% to 93%, every 1% increase of recognition accuracy will reduce task completion time from 1989 to 1915 ms; when recognition accuracy increases from 94% to

*Corresponding author. Industrial and Operations Engineering Department, University of Michigan, 1205 Beal Avenue, Ann Arbor, MI 48109-2117, USA.
E-mail address: changxuw@umich.edu (C. Wu).

doi:10.1016/S1071-5819(03)00026-0

100%, every 1% increase of recognition accuracy will reduce task completion time from 1895 to 1392 ms. Guidelines in designing these recognizers were given based on these implications.

Keywords: Performance modeling; Handwriting; Recognition accuracy; Segmentation time

Nomenclature

ST	segmentation time, ms
R	recognition time, ms
UD	user delay, ms
WT	average writing time per character, ms/character
RA	recognition accuracy, %
D	task completion time/duration, ms
N	number of input character, nondimensionless
T	repairing time, ms

1. Introduction

Nowadays pen computing and pen-based interface have wide application in personal digital assistance (PDA) and other mobile computing systems (Mackenzie and Chang, 1999; Davis et al., 1998; Oomika and Nakagawa, 1997; Frankish et al., 1995).

On the one hand, to improve the human performance in these systems is crucial, because the poorly implemented system can lead to an unacceptable system (Kalaswsky, 1999) and there is a strong causal relationship between the performance and user's satisfaction (Frankish et al., 1995). In addition, small reductions in average transaction time can be worth millions of dollars every year to service providers (Hone and Baber, 1999).

Human performance modeling is the base to understand the detailed process of man–machine interaction and to improve the performance. First, with this model, it is possible to predict the user's performance of these systems accurately with a limited number of parameters rather than run the tedious empirical usability tests. Compared with the method of factorial experiment, the method of modeling allows a large range of different parameters to be tested, helping the researcher to pin-point these parameters at which a given dialogue design would be most effective (Hone and Baber, 1999). Second, it can be used to understand the detailed process of interaction between the user and these interactive handwriting recognizers, even speech recognizers. The factors studied in this model, including recognition accuracy, average repair time, and execution time, still play the key roles in determining the human performance in the other graphic characters and speech recognition systems

(Karl, 1993). Third, theoretical performance will have wider applications in estimating human performance than the empirical linear model. Theoretical performance modeling plus the minor modifications for the empirical experiment may hold higher degree of fitness to another set of experimental data than the empirical linear model because the latter is only based on the goodness of fit in one case rather than generalize the parameters of the other cases (Spine et al., 1984; Casali et al., 1990). Fourth, it can facilitate the development of the intelligent algorithm to maximize the user's performance in these systems. For example, if the system can adapt to different speed of writing of the user by automatically changing the segmentation time in the model to maximize user's performance. However, there has been little research to model human performance in putting and repairing the graphic character systems.

On the other hand, to the 1.3 billion Chinese people, inputting Chinese characters into computer remains a major impediment in human–computer interaction (Sarcher et al., 2001; Cheng, 1996). The development of the handwriting system has shed lights on the solution to this bottleneck (Sarcher et al., 2001). However, most of the previous Chinese human–computer interaction studies are concerned with keyboard inputting with different methods to map the graphic characters onto the QWERTY keyboard (Cheng, 1996; Zhang, 1992). There is little human–computer interaction research, including human performance modeling research in Chinese character handwriting recognizers, to improve the human performance in these systems.

To separate continuously inputted characters one by one, there are two kinds of methods used in separation: first, temporal segmentation method segments characters based on temporary interval between each character inputted (Lu et al., 1995); second, spatial segmentation method separates characters by their spatial allocation on the screen; the latter method will take up more space in the user interface and this problem will become more severe when user plans to input legible characters onto the $2 \times 2\,\mathrm{cm}^2$ small screen user interface of the mobile phones (see Fig. 1). The former method will impair the human performance with the existence of segmentation time in the system (see Section 1.1 segmentation time), which makes it necessary for the related human performance research to find out the performance models and optimal length of optimal segmentation time in these recognizers.

In addition, theoretically, the human performance may include task completion time and error rate. Casali et al. (1990) study in speech recognition system treated task completion time and error rate to be repaired as the indexes of the human performance. In this study, since the subjects were required to repair all of the recognition error during input process, the higher the error rate, the longer the time will be used by the subject to complete the task, i.e. the error rate was included and reflected by the task completion time in this study. Therefore, in this study, task completion time was treated as the main index of the human performance in this study.

The detailed modeling includes: decomposing the detailed process of handwriting, finding out the factors that may affect the performance, and setting up mathematical and statistical models to describe their quantitative relationship with the performance.

Fig. 1. A mobile phone with a Chinese character handwriting recognizer. Adapted from Wu (2001).

Table 1
Factors that may affect the human performance in handwriting recognition system

Sources of the factors	Factors
Temporary segmentation handwriting recognizer	Segmentation time
User	Handwriting speed (average handwriting time per character) (Frankish et al., 1995); user strategy (Teal and Rudnicky, 1992)
Interaction between the user and recognizer	Recognition accuracy (Wolf et al., 1991); repairing time (Wolfgang and Yang, 1998)
Others	Task type (Frankish et al., 1995); the possible direct and indirect causal relationship among these factors (Frankish et al., 1995)

1.1. Factors in temporary segmentation handwriting recognizer that may influence human performance

Previous human factors research had studied some factors that influence the human performance in handwriting recognition system (Table 1).

Segmentation time: Segmentation time is the default length of the interval of different strokes set in the system. In the process of on-line handwriting, for example (see Fig. 2), if the length of interval between strokes ① and ② is shorter than that default length (segmentation time), the system will treat these two strokes as part of a whole character and wait for the input of the third stroke. If the interval is longer than the segmentation time, the recognizer will treat all of the previous inputted strokes as a completed character.

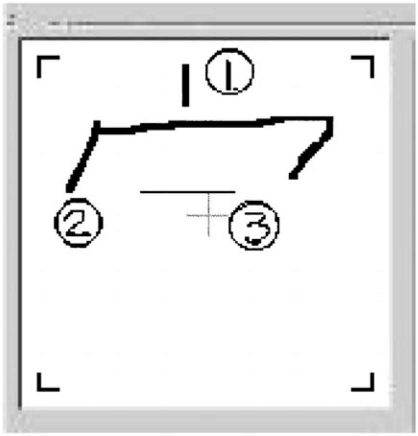

Fig. 2. Handwriting window of a temporary segmentation Chinese character handwriting recognizer.

However, the length of segmentation time triggers some difficulties for the user in handwriting: if the segmentation time is set too short, the user cannot keep up with the system and some strokes which are only a part of a character are processed by the system as a whole character; if the length of segmentation time is set too long, user will wait for an unnecessary time after writing all the strokes of a character, and the performance of the user may be impaired. Even though there are some studies in segmentation algorithm in Chinese and Japanese characters (Chien, 1995; Krishnan and Moriya, 1993; Cho et al., 1993), it is still not clear about the effects of different segmentation time on user's task completion time are still not clear.

Recognition accuracy: It was found that recognition accuracy was affected by many factors. These factors included style of writing, amount of training, interval of disuse, and alphabet (Wolf et al., 1991). Constraint of inputting (lower case and upper case) also significantly influenced the recognition accuracy found by an empirical study in handwriting recognizer (Mackenzie and Chang, 1999). Some researchers did successfully manipulate recognition accuracy by the method of "Wizard of Oz" (LaLomia, 1994; Buskirk and LaLomia, 1995). However, it will become difficult to "trick" the subjects when there are inputting errors coming from the subjects and the repairing process handled by the subjects themselves. For example, when there is no inputting error from subjects, subjects will treat the incorrect character as the wrongly recognized one. However, if subjects wrongly input a character but the character that appears on the screen is absolutely correct, it will lead the subjects to find the truth of "Wizard of Oz". Consequently, it is difficult to manipulate this variable when there are inputting errors coming from the subjects and repairing process handled by the subjects.

Repairing process and repairing time: Even though a recognition rate of 97% or higher was acceptable (LaLomia, 1994), the commercial product might have some difficulty in meeting such a standard of acceptance. Empirical study in handwriting recognizer indicated that the accuracy observed (87–93%) was below the walk-up accuracy claimed by the developers of the recognizer (Mackenzie and Chang, 1999). The recognition accuracy will be much lower in the first stage of handwriting when the templates in the recognizer have less data of the user's writing style for automatic learning. Moreover, the repair process is more common for the speech recognizer. Consequently, it is necessary to study the whole process of handwriting with repairing process (see Section 2.2). Furthermore, it is also difficult to manipulate this variable in the handwriting process because it is determined by the writing style and other complex characteristics of the subject, e.g. motor control ability and visual reaction speed, etc.

User's execution time and writing time: User's execution time is an important part of normal human–computer interaction (Teal and Rudinicky, 1992). First, there is an obvious individual difference in user's execution time and writing time, which makes systematic manipulation difficult; in Experiment 1 of the present research, writing time in the process of handwriting was recorded by the software automatically; Second, in interacting with handwriting recognizer, for most individuals, handwriting characteristics are well established and stable (Frankish et al., 1995). Consequently, in Experiment 2 of the present study, to facilitate the application of the performance model in realistic world, writing time in the process of handwriting was replaced by the subjects' initial writing speed recorded at the beginning of the experiment.

Other factors are also essential to be taken into consideration in the study of human–computer interaction in regard to handwriting recognizer: first, user's strategy of selection: existence of segmentation time caused some delay when the recognizer responded to the user's inputting action. Such a delay, like the system response time, will affect a user's selection of task strategies (Teal and Rudnicky, 1992; Paddy and Draper, 1996); second, task context: study in recognition accuracy of handwriting recognizer found that the influence of recognition accuracy on user satisfaction depends on the task context (Frankish et al., 1995).

Moreover, Frankish's research (1995) indicated that there are direct and indirect dynamic causal relationships among these factors.

Based on the review above, in present study, segmentation time, recognition accuracy, repairing time, and handwriting time were the major four parts of the performance models. Moreover, first, a static performance model which treats the factors independent of each other will be established to provide the basic composition and estimation of the task completion time; second, since there are possible direct and indirect dynamic causal relationships among these factors, a dynamic performance model which takes these relationships into consideration will be set up to reduce the variables in the static performance model. In addition, user's strategy selection was controlled by experiment instruction, which was verified by a pilot study. Only one task context, the copying task, was applied in the present study, so that the irritant variable of the task context was also controlled in this study.

2. The static performance model

2.1. Condition without the error repair and its performance model

The detailed information on the inputting process in computer is involved in the analysis of an interaction cycle in terms of system and user events (Shneiderman, 1984). There is a model set up to analyse the detailed process of the English words typed under the condition without the error repair (Teal and Rudnicky, 1992).

System events in the model contain the system response time, display time and lockout time (see Fig. 3), and in the process of the temporary characters handwriting, system events contain the segmentation time (ST) and recognition time (R).

User events in the model are composed of the three components: (1) user planning and acquisition time is the time that elapses between when user initiates the current system activity and the user begins entering the next task into the computer; (2) user delay (UD) is defined as the time that elapses between when the system is ready for the next system input and when the user actually begins to enter the next activity into the system, e.g. the duration between when the recognized character appears on the screen and when the user presses the "back space" key on the keyboard; (3) the execution time is the time required by the user to enter the task into the system. In the process of the Chinese characters handwritings, it refers to the time taken by the user to write one Chinese character (WT).

Therefore, if the user input N characters correctly and continually, the performance model, which is represented by the task completion time/duration under the condition without the error repair, can be inferred:

$$\mathrm{Duration}_N = \sum_{i=1}^{N}(\mathrm{ST} + R + \mathrm{UD} + \mathrm{WT}) + \mathrm{Error}. \tag{1}$$

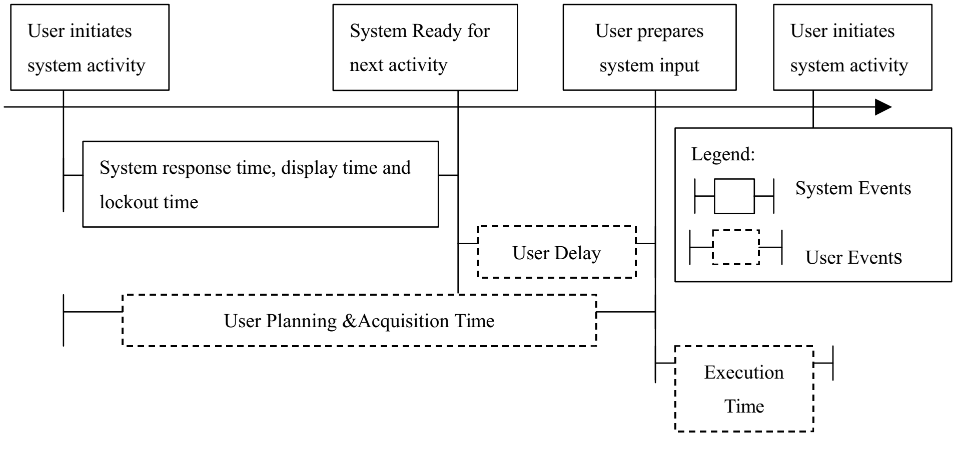

Fig. 3. Model of user and system events occurring during a normal human–computer interaction (Teal and Rudnicky, 1992, briefed).

To the frequently used Chinese characters, the distribution of the simple visual reaction time is $M = 439$ ms, S.D. $= 48$ ms (Gao et al., 1995). To facilitate the fit to performance data for the model, only the mean of the simple visual reaction time to the Chinese characters was used in this study.

Consequently, if average (WT) represents the average handwriting time for each Chinese character, performance model will be transformed into (2):

$$\begin{aligned} \text{Duration}_N &= N(\text{Average(WT)} + \text{ST} + R + \text{UD}) + \text{Error} \\ &= N(\text{Average(WT)} + \text{ST} + R + 439) + \text{Error}. \end{aligned} \tag{2}$$

2.2. *Condition with the immediate error repair and its performance model*

Generally, there are two kinds of repair methods in temporary segmentation handwriting recognizer: immediate repair and delayed repair. The former is to repair the wrong character immediately after it appears on the screen. Since it can facilitate the immediate transmit from the track of the handwriting into the analysable codes in the central processor (e.g. handwriting name and telephone number can be processed immediately by PDA), which is more typical for common usage, this repair method was adopted and studied in this study.

Adding the repair process to the model without the error repair, the user events in the repair process will contain the five components (see Fig. 4):

Box ① and ② are the processes of repairing of the wrong character by the user. Detailed process includes press the "Back Space" key (Box ①) and write that character again (Box ②).

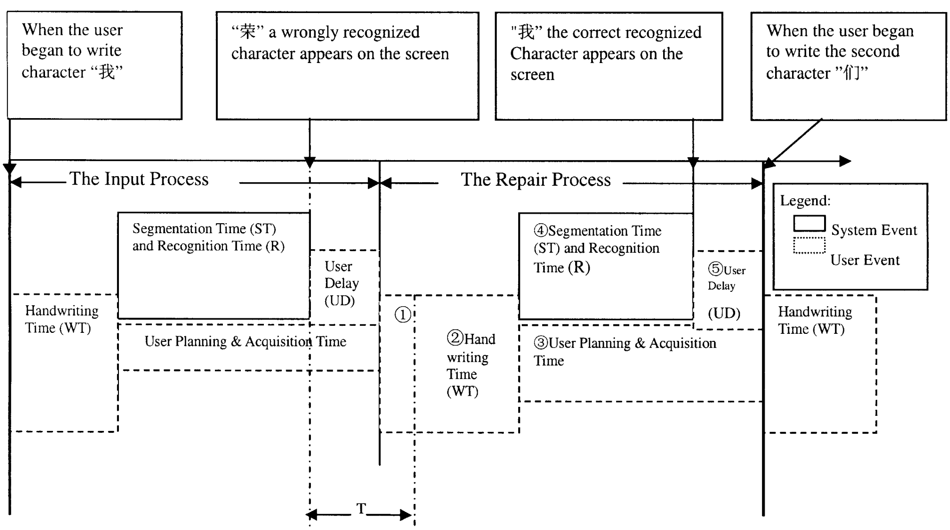

Fig. 4. Model of user and system events occurring during the handwriting and immediate repair process.

To facilitate the verification of the mathematical model, assume the repairing time (T) (see Fig. 4): the interval between when the wrongly recognized character appears on the screen and when user begins to write the correct recognized character.

Box ③, ④ and ⑤ were the same as the corresponding boxes in the input process.

The total handwriting time was composed of two parts:

(1) The time taken to input N characters correctly:

$$N(\mathrm{WT} + \mathrm{ST} + R + \mathrm{UD}). \tag{3}$$

(2) The extra time taken to repair the wrongly recognized characters:

Recognition accuracy of the system was assumed to be the ratio between the number of correctly recognized characters and the total number of characters inputted. If user input N characters with immediate repair, it is assumed that x times of the repair process will occur.

Since $\mathrm{RA} = N/(N + x)$, we have

$$x = N[(1/\mathrm{RA}) - 1] \tag{4}$$

and total time taken in the repair process in each trial:

$$(T - \mathrm{UD}) + \mathrm{WT} + \mathrm{ST} + R + \mathrm{UD}.$$

Consequently, the time taken by the whole repair process in each trial:

$$N[(1/\mathrm{RA}) - 1][(T - \mathrm{UD}) + \mathrm{WT} + \mathrm{ST} + R + \mathrm{UD}]. \tag{5}$$

Theoretically, the performance model of handwriting and immediate repair process will be (3) + (5) see (6):

$$\begin{aligned}\text{Duration} = {} & N(\mathrm{WT} + \mathrm{ST} + R + \mathrm{UD}) + N((1/\mathrm{RA}) - 1) \\ & \times [(T - \mathrm{UD}) + \mathrm{WT} + \mathrm{ST} + R + \mathrm{UD}] + \text{Error},\end{aligned} \tag{6}$$

$$\begin{aligned} = {} & N(\mathrm{WT} + \mathrm{ST} + R + 439) + N((1/\mathrm{RA}) - 1) \\ & \times [(T - 439) + \mathrm{WT} + \mathrm{ST} + R + 439] + \text{Error},\end{aligned} \tag{7}$$

$$= N[(1/\mathrm{RA})(T + \mathrm{WT} + \mathrm{ST} + R) + 439 - T] + \text{Error}. \tag{8}$$

3. The dynamic performance model

After establishing the basic components of the task completion time, it is necessary to build up the dynamic direct and indirect causal relationship among these factors to reduce the variables in the static model.

First, user will adapt their work style to the response time (Shneiderman, 1992), which suggests that with the reduction of segmentation time (ST), user might write more rapidly and their average writing time per character (WT) will decrease (path A, Fig. 5).

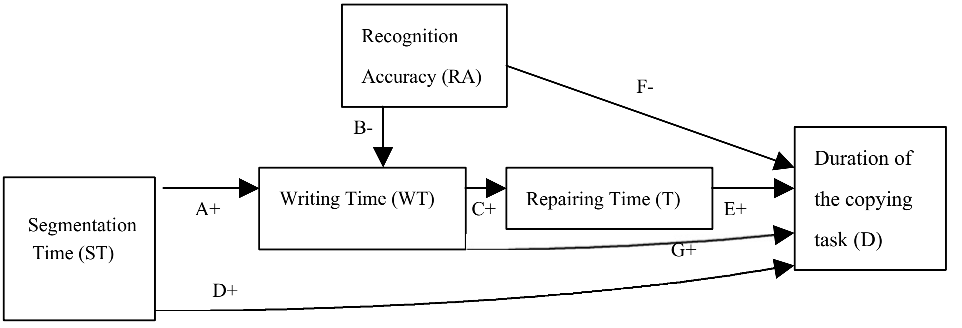

Fig. 5. The predicted dynamic model of causal relationships between variables (+ indicates a positive linear relationship between the two variables, – indicates a negative one).

Second, feedback from the recognizer could in principle allow user to develop an internal model of the recognition process, and to adapt their writing styles accordingly. If the subjects find the recognition accuracy is relatively low, they can slow down their handwriting time to adapt the recognizer (path B).

Third, according to the theory in motor abilities, there is a possible superability underlying various kinds of movement skills, e.g. reaction time, speed of movement and finger dexterity, etc. (Schmidt, 1991). It suggests that with the decrease of writing time, including the increase of the movement speed and finger dexterity, the visual reaction time and movement time will be decreased by the effect of the possible superability. Since the visual reaction time (user delay) and movement time (pressing the "backspace" key) are the major two parts of the repairing time, the decrease of writing time will lead to decrease of repairing time (T) (path C).

Fourth, based on Teal's model (Teal and Rudnicky, 1992) and the detailed process of handwriting (Fig. 3), the performance is determined directly by segmentation time (path D), repairing time (path E), recognition accuracy (the frequency of repairing) (path F), and writing time (path G). Empirical study in handwriting recognizer (Mackenzie and Chang, 1999) also showed that text-entry speed depends on the user's printing and writing speed (path G).Therefore, the dynamic model of causal relationships between variables is summarized in Fig. 5.

4. The verification of the static and the dynamic performance model

In Experiment 1, an empirical experiment was conducted to verify the validity of static and dynamic model separately with some minor modifications in the models themselves; based on the successful verification of these models in Experiment 1, two models are combined together to reduce the number of variables in the static model.

Then, Experiment 2 was conducted to calculate the degree of fitness of the integrated model to the experimental data in a real commercial product.

4.1. Experiment 1

The main purpose of Experiment 1 was to find out whether both the static and dynamic model could fit the observed user performance data well or not. The effect of segmentation time on task completion time was also investigated in Experiment 1.

4.1.1. Method

Participants: Twelve university students (nine male, three female), 19–21 years old, who never used such kind of handwriting system participated in our major experiment. They were not major in statistics and psychology. To exclude the effects of handedness, subjects copied the sentence using their dominant hands (all were right handed). Their experience in using computer and word processor were recorded. Subjects were thanked and rewarded after the experiment.

Material: Six sentences elected by a pilot study (Wu and Zhang, 2000), each contains 24 Chinese characters (mean: totally 179 strokes per sentence, S.D.: 11 strokes per sentence), were used in the major tests. In the pilot study, 10 subjects in the same range of age and education background spent no significant duration of time to copy these six sentences, excluding the individual difference treated as the covariate variable. Moreover, the content of the sentences were the academic terms in statistics and psychology, which all the subjects were unfamiliar with.

Apparatus: A PC with 523,080 KB RAM, and a handwriting recognition System "MengQue Pen" Model: ET-0405-U were used in the experiment. Software was developed which could manipulate different lengths of segmentation time in recognition and record task completion time as well as the other three factors in the model.

Experimental design: In this experiment, segmentation time was manipulated systematically and the other three factors were recorded because only the segmentation time could be well manipulated in the study of on-line recognition with repairing process (see Section 1.1 the review of these factors). In the verification of the static model, the theoretic value of the task completion time was calculated by the static model according to different values of the four factors manipulated and recorded in the experiment. Then, this theoretic value would be statistically compared with the empirical value of task completion time collected in the experiment; in the verification of dynamic model, to verify the predicted direct and indirect causal relationship among those factors, the path analytic method was adopted as it enabled the statistical control of some variables that were difficult to be controlled directly (Guo, 1999). Consequently, one factor within-subject Latin square design was adopted in the experiment, which could also counterbalance the effect of sequence of the segmentation time (Wang, 1990).

In this study, there were six levels of the segmentation time: 190, 440, 740, 1040, 1340 and 1640 ms. This minimum segmentation time (190 ms) was set by a pilot study (Wu and Zhang, 2000), subjects had much difficulty to keep up with the system when the segmentation time was less than the minimum value; the maximum segmentation time (1640 ms) was set according to the research done by Youmans (1983) (Shneiderman, 1992), because user will not accept system response time longer than 2 s and the longest total response time in this study reached 1700 ms (segmentation time (1640 ms) + recognition time (60 ms) = 1700 ms).

Experimental procedure: Every subject will experience 12 trials, which means that he or she will experience each level of the segmentation time twice. The detailed process of the experiment included: (1) Introduction of the use of the system and practices of the repair of the characters: all the subjects were taught how to use the new systems and experienced enough exercise so that they had no extra trouble in using the system. (2) Practicing the repair of the characters, subjects were asked to repair the wrongly recognized characters with the light pen. And their average repairing time was recorded and analysed immediately. When their speed of repair became stable, the subjects were asked to take the pretest. (3) Experiment instructions and the pretest, subjects were asked to copy the sentence on the screen as quickly as possible with their own style of cursive, and repair the wrongly recognized characters immediately after they appeared. In the pretest, subjects were asked to copy two sentences chosen from the pilot study, which took almost same duration in copying. After 3 min rest, subjects would go on to the major tests. Subjects were told that the systems were different in each trial. (4) Major tests, 12 trials were given in the major tests, and after each trial dependent variables were measured like the pretest. Subjects took a rest of 3 min between each trial.

4.1.2. Results of Experiment 1

4.1.2.1. Verification of the static performance model. In each trial, subjects were asked to copy all the 24 characters correctly; consequently, the corresponding task completion time (D) was

$$D = 24[(1/\mathrm{RA})(T + \mathrm{WT} + \mathrm{ST} + R) + 439 - T] + \mathrm{Error}.$$

Overall, the mean of the task completion time/duration in the experiment was 93589.9 ms, S.D. $= 25888.77$; mean of the expected task completion time calculated from the static model was 91872.81 ms, S.D. $= 23236.97$. With the increasing of segmentation time, different values of expected and observed task completion time were compared in Fig. 6.

To examine whether the task completion time observed significantly differed from the one estimated by the static model or not, T test for two independent samples were conducted since the two distributions of the time were normal distributions (Kolmogorov–Smirnov value = 0.092 for observed task completion time, df $= 69$, $p > 0.05$; Kolmogorov–Smirnov value = 0.082 for expected task completion time, df $= 72$, $p > 0.05$).

$t = 0.415$ (df $= 139$) ($p = 0.679 > 0.05$) and Levene's test for equality of variances ($F = 0.576$ $p > 0.05$) indicated that there was no significant difference between the mean of the observed and expected task completion time.

Moreover, to measure the degree of fitness of the static model, a linear regression model was set up to describe the relationship between observed and expected task completion time (Fig. 7).

Table 2 indicated that there was a significant linear regression between the observed and the expected task completion time (unstandardized coefficients

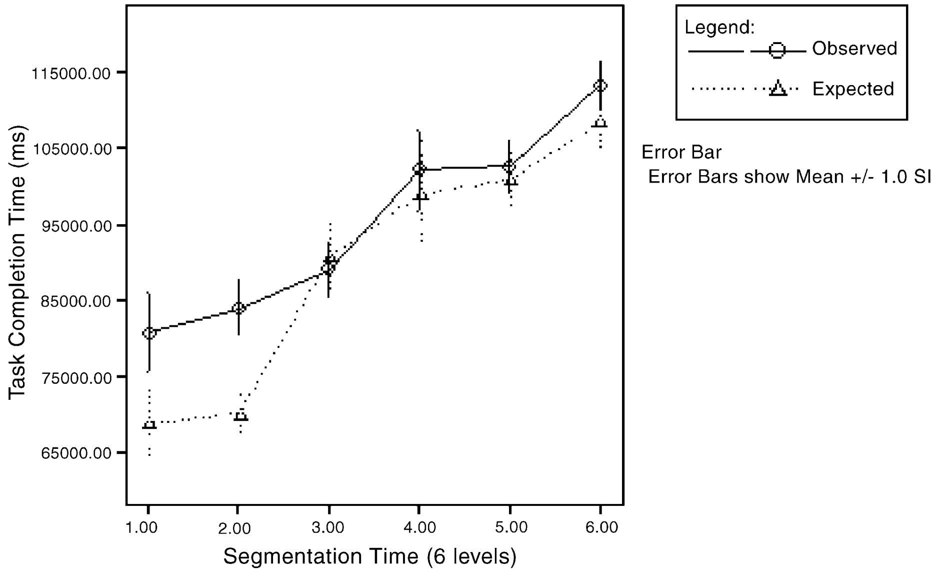

Fig. 6. Comparison between the observed and expected task completion time.

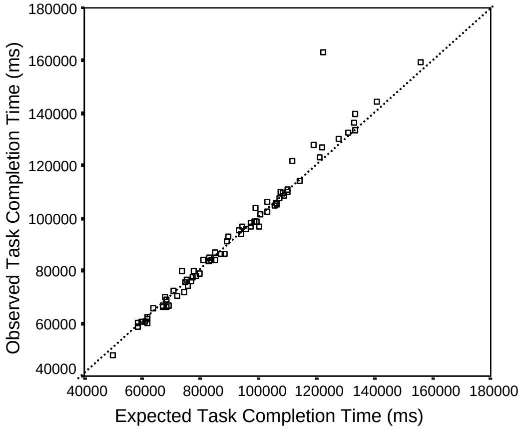

Fig. 7. The scatter plot of the observed and expected task completion time.

$B = 1.07$, $t = 40.368$, $p < 0.0001$). Moreover, since the adjusted $R^2 = 0.960$, it suggested that the static model could explain 96% of the variance of the observed data. In addition, since the slope of the regression line (unstandardized coefficients B in Table 2) approximately equaled 1, only the constant in the regression can be served as the error term in the static model to modify it:

$$D = 24[(1/\mathrm{RA})(T + \mathrm{WT} + \mathrm{ST} + R) + 439 - T] - 4973. \tag{9}$$

Table 2
Coefficients in the regression models[a]

	Unstandardized coefficients B	S.E.	Standardized coefficient betas	T	Sig
(Constant)	−4973.25	2425.54		−1.961	0.054
Expected task completion time	1.07	0.026	0.981	40.368	0.000

[a] Dependent variable: observed task completion time.

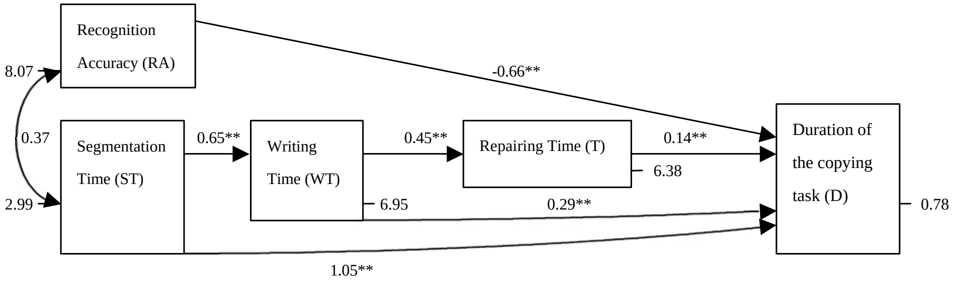

Fig. 8. The path analytic model of the dynamic performance model (**$p<0.01$).

When user delay was replaced with $439 \pm 48 \times 2$ ($M \pm 2$ S.D.) ms, model (9) varied very little (user delay = 535 ms: adjusted $R^2 = 0.960$, *constant* = −7436; user delay = 343, adjusted $R^2 = 0.960$, *constant* = −2509).

4.1.2.2. Verification and modification of the dynamic performance model. In order to verify and modify the dynamic model, Lisrel 8.3 simplistic causal modeling technique (Joreskog and Sorbom, 1999) was used to conduct a path analysis. This method can examine the direct and indirect causal relationship among the factors (Joreskog and Sorbom, 1999; Guo, 1999). It was also used in human factors studies effectively (Guerrier et al., 1999). Moreover, the method permits the statistical control of some variables that are difficult to be controlled directly (Guo, 1999). The verification of the theoretical model is shown in Fig. 8.

In this model, the goodness of fit index (GFI) equaled 0.979, adjusted goodness of fit index (AGFI) = 0.897, root mean square error of approximation (RMSEA) = 0.052, and the p value for test of close fit (RMSEA <0.05) = $0.387 > 0.05$, which indicated that the model generally fitted the empirical data (Guo, 1999; Hou and Meng, 2001) except that there was no significant linear relationship between RA and WT (the path B assumed by the dynamic model in Fig. 5). The corresponding unstandardized regression equations were calculated as follows:

$$T = 0.816\,\mathrm{WT} - 51.94\ (\text{regression coefficient}: p<0.001,\ R^2 = 0.224), \tag{10}$$

$$\mathrm{WT} = 0.123\,\mathrm{ST} + 1377.38\ (\text{regression coefficient}: p<0.01,\ R^2 = 0.06), \tag{11}$$

$$D = -164956.984\,\mathrm{RA} + 29.176\,\mathrm{ST} + 39.814\,\mathrm{WT} + 4.718\,T + 134984.18 \times (\text{regression coefficient} : p<0.001,\ R^2 = 0.939). \quad (12)$$

4.1.2.3. The effect of segmentation time on task completion time. It was found that main effect of the segmentation time on the task completion time was significant (GLM repeat measure, F (5, 55) = 11.626, $p<0.001$). With the within-subjects simple contrasts test, which set the level 1 of segmentation time (190 ms) as the baseline, there were the two stages of the effect of segmentation time on task completion time (see Fig. 6):

Stage 1 (190–740 ms): There was no significant difference of the task completion time when segmentation time was between level 2 (440 ms) and the baseline (190 ms), F (1, 11) = 0.001, $p>0.01$; no significant difference was found between level 3 (740 ms) and the baseline (190 ms), F (1, 11) = 9.03, $p>0.01$; consequently, the prolongation of the segmentation time in this stage will not significantly affect the task completion time.

Stage 2 (190–740 ms): Significant difference of the task completion time was found when segmentation time was between level 4 (1040 ms) and the baseline, F (1, 11) = 22.170, $p<0.01$, level 5(1340 ms) and the baseline, F (1, 11) = 18.294, $p<0.01$, level 6 (1640 ms) and the baseline, F (1, 11) = 63.779, $p<0.01$; that meant when the segmentation time is longer than 1040 ms, user with average writing time 1492 ms per frequently used character (S.D.: 223 ms) would take significant longer time to fulfill the copy task.

4.1.3. Discussion of Experiment 1

The data from Experiment 1 strongly supported the assumed static model and dynamic model because the static model accounted for 96% of the variance of the performance and the GFI of the dynamic path analytic model reached 0.979.

The static model was based on the mathematical inference and data from the previous studies. Compared with the empirical linear regression model to describe human performance in speech recognition system (Spine et al., 1984; Casali et al., 1990) and other empirical studies in handwriting recognizer (Mackenzie and Chang, 1999), this static model may have higher degree of fitness to human performance data in another experiment (see Experiment 2). Moreover, when user delay was replaced with the reaction time of the characters ($M \pm 2$ S.D., $439 \pm 48 \times 2$ ms), the degree of fitness of static model to the experimental data would not change greatly. Since the frequently used characters cover 99% of the text content in daily communication (Beijing Language Research Institute, 1986), it indicated that the model could be used to estimate human performance of handwriting copying task with the recognizer in daily usage. However, the static model did not consider that there is a dynamic direct and indirect causal relationship among the factors.

The dynamic model was constructed from theoretic summary and verified by path analytic model. However, (1) the relative small value of R^2 in regression Eq. (11) ($R^2 = 0.06$) indicated that the variance of writing time was independent of the variance of the segmentation time. Consequently, it is difficult to predict writing time only with the data from segmentation time; (2) one path in the predicted dynamic model (path B) was not verified in the path analytic model. The insignificant correlation in that path indicated that writing time was independent of the recognition accuracy. This result was consistent with the research (Wolf, 1990) in pen interface, which found no evidence that experience with a recognition device caused the subjects to change their writing styles.

Moreover, user performance was not significantly impaired when the segmentation time was prolonged from 190 to 740 ms. However, when the segmentation time was longer than 1040 ms, user performance was significantly impaired. This indicated that the preferred segmentation of handwriting recognizer was shorter than 1 s when users' average writing time is around 1500 ms per frequently used character.

In addition, user will adopt different strategies with the prolongation of system response time (see Introduction). Teal and Rudinicky (1992) suggested and confirmed that there are three strategies of the user in the process: automatic performance, pacing and monitoring strategy; user delay is also changed with the different adoption of the strategies. This research focused only on the process of handwritings when user adopts monitoring strategy, which was partially confirmed by the pilot study (Wu and Zhang, 2000). (It was found that user delay of four subjects out of five were stable (Subjects 1–4: $\chi^2(\mathrm{df} = 87) = 52.029$, $p > 0.05$; subject 5: $\chi^2(\mathrm{df} = 33) = 69.833$, $p < 0.05$.) In that case, subjects relied on the external stimuli to signal system availability (e.g. only on seeing the characters appearing on the screen, did user begin to input the next character), and his or her temporal delay became stable and equal to the user's simple reaction time (Teal and Rudinicky, 1992), which can be describe with the model of open-loop motor control system (Schmidt, 1991; Long, 1976).

It is necessary to minimize the number of factors in the static model to fit the observed data. With the equation derived from Experiment 1, Eq. (10) in the dynamic model can be integrated into the static model (Eq. (9)):

$$D = 24[(1/\mathrm{RA})(T + \mathrm{WT} + \mathrm{ST} + R) + 439 - T] - 4973, \tag{9$'$}$$

$$T = 0.816\,\mathrm{WT} - 51.94, \tag{10$'$}$$

$$\Rightarrow D = 24((1/\mathrm{RA})(1.816\,\mathrm{WT} + \mathrm{ST} + R - 51.94) + 490.94 - 0.816\,\mathrm{WT}) - 4973. \tag{13}$$

Eq. (11) might be used in estimating the copying task completion time with immediate repair.

However, first, the R^2 in Eq. (10) is small (R^2 equaled 0.224); it is necessary that the degree of fitness of the integrated model (13) should be testified by another empirical experiment with user under different training conditions; second, another empirical experiment is required to find out whether the integrated model with mathematical inference will have a higher degree of fitness than the empirical multi-regression model. If the R^2 in integrated model in Experiment 2 is bigger than the one in regression model, it is indicated that mathematical and theoretic inference in this study can be generalized into more real recognizers than an empirical linear model.

4.2. Experiment 2

The main goal of Experiment 2 was to find out whether the integrated modal (Eq. (13)) could fit the observed user performance data well in different training with a commercial handwriting recognizer. One factor between-subjects experimental design was adopted in Experiment 2. The factor is the subjects with different training conditions.

4.2.1. Method

4.2.1.1. Participants with different training conditions and procedures.

(1) *Naive user*: training session was not given before the current formal test. Seven university students (four male, three female), who never used any handwriting recognizer, participated in this training condition. After their initial writing time measurement by the software self-developed in this study, they went to the formal test stage directly.
(2) *Trained user*: There were two sub-groups of trained user:
    (a) *Outdated-trained user*: Training session was given 12 months before the current formal test. Six university students who had participated in Experiment 1(five male, one female) voluntarily joined Experiment 2. After their initial writing time measurement by the software, they went to the formal test stage directly.
    (b) *Immediate-trained user*: training session was given immediately before the formal test. Six university students (three males, three female) who never used any handwriting recognizer participated in this training. They received the training session immediately before the formal test. To balance the amount of training in each subject, they were required to copy a paragraph that contained 46 characters and few inputting errors for all of the subjects were observed after the training session. The training session usually lasted 3–5 min.

In addition to these different training conditions above, all the three groups of subjects experienced the experimental procedure as follows: at the first stage of the experiment, average writing time per character of each subject was recorded. Subjects were asked to copy down the first sentence prepared by the pilot study (Wu and Zhang, 2000) with their own writing style. Then, subjects were introduced to the use of the commercial system. At the formal test stage, subjects were asked to copy the second sentence prepared by the pilot study as quickly as possible with their own style of cursive, and repair the wrongly recognized characters immediately after it appeared. To eliminate the effect of practice in formal test, only one level of segmentation time (1100 ms) and only one trial of test were given to each subject.

All the subjects, 19–26 years old, were not major in statistics and psychology. They were unfamiliar with the content of sentence in the experiment and did not have experience with the commercial handwriting recognizer. To exclude the effects of handedness, subjects copied the sentence using their dominant hands (all were right handed). Moreover, the classification of the user into trained and naive user group was verified by their significant difference between the task completion time in Experiment 2 (see Section 4.2.2 result of Experiment 2).

4.2.1.2. Experimental apparatus. A commercial Chinese character inputting system with its handwriting recognizer was used in Experiment 2, not only because it was a commercial product widely used in China but also because it can manipulate the segmentation time in its setting.

To facilitate the verification of the integrated model, segmentation time was manipulated by the setting in the commercial software. The exact setting of its segmentation time plus recognition time is 1100 ms; writing time was measured by self-developed software. This software could also record the interval of every characters and the interval between the inputs of different characters; to facilitate the application of the performance model in realistic world, average writing time per character in the process of handwriting (Experiment 1) was replaced by the subjects' initial writing speed recorded at the beginning of the experiment (This replacement was tested for insignificance between writing time at the beginning of the whole experiment and the writing time at the beginning of the training session, paired samples test, $t(\mathrm{df} = 6) = 1.512$; $p > 0.05$). The whole procedure of Experiment 2 was recorded by software "Snagit32" (version 4.3) and corresponding video files were created. With these video files, the recognition accuracy, and task completion time were recorded and calculated by the software "Premiere" (version 5.1), which can analyse the detailed process of the video documentation.

4.2.2. Result

(1) *Comparison of the observed and expected task completion time*: The result of a comparison of the observed and expected task completion time in the three groups is summarized in Table 3.
(2) *Comparison of the three groups with different training conditions*: There was significant difference of observed task completion time among these three groups ($F(2, 18) = 3.791$, $p < 0.05$). Further post hoc test (LSD method) found that the source of such difference existed because naive user group took significantly longer time to fulfill their copy task than the other two groups ($p < 0.05$). This also indicated that the method of categorization of subjects into different groups was effective.

Table 3
Comparison of the observed and expected task completion time

User group with different training condition		Descriptive statistics			Comparison of the observed and expected task completion time			
		Observed task completion time (ms) (mean and S.D.)	Recognition accuracy (mean and S.D.)	Average writing time per character (ms/charact) (mean and S.D.)	Compare means: nonparametric test (Wald–Wolfowitz test)		Correlation analysis: Pearson correlation	
Naive user		151 286 (44 271)	0.592 (0.22)	1814 (280)	1.39	$0.922>0.05$	0.98**	0.96
Trained user	Out-dated trained	107 833 (16 666)	0.824 (0.10)	1607 (229)	−1.51	$0.067>0.05$	0.91*	0.82
	Immediate trained	132 769 (41 000)	0.692 (0.19)	1784 (252)	0	$0.522>0.05$	0.86*	0.73

*$p<0.05$.
**$p<0.01$.
S.D. was included in the pairs of parentheses.

Treating the arithmetic difference between observed and expected task completion time as the dependent variable, there was no significant change in the difference among these three groups ($F\ (2, 18) = 1.778, p > 0.05$). It indicated the consistency of the model in fitting the observed data among the subjects with different training conditions.

(3) *Comparison of the integrated performance model with the multiple regression model*: It was found that the integrated performance model (13) inferred from Experiment 1 can still account for 92% of the variance in task completion time in the three groups in Experiment 2 (Pearson correlation coefficient reached 0.959, $p < 0.01$); However, the multiple regression model (12) got from Experiment 1 can only explain 64% of the variance of task completion time in the three groups in Experiment 2 (Pearson correlation coefficient was 0.804, $p < 0.01$).

4.2.3. Discussion of Experiment 2

Nonparametric test and linear regression indicated that the integrated model still fitted the naive and trained user performance data well (R^2 reached 0.96) when subjects used commercial handwriting recognizer fulfilling copying task with immediate repair. Now, with the parameter of segmentation time, recognition time, recognition accuracy, and user's initial writing time, it is possible to calculate user's copying task completion time with immediate repair in interacting with handwriting recognizer.

The R^2 in integrated model was greater than the one in empirical linear model, which indicated that the integrated model with mathematical inference fitted the observed performance data better than the empirical linear model, which also stressed the necessity of mathematical and theoretic inference in this study.

From Table 3, it was found that the model would have lower degree of fitness for the trained user data (R^2 reached 0.82 for the outdated trained user; 0.73 for the immediate trained user) than for the naive user data (R^2 reached 0.96). The reasons may stem from the individual difference in learning efficiency, transference, and maintenance of motor skills (Matthews and Davies, 2001). Therefore, it is possible that there is a greater individual difference among a trained user than a naive user. Example, some subjects acquired the inputting skills quickly and maintained it for a longer duration, but some subjects could not. Thus, it will be more difficult for the integrated model to estimate the trained user performance than the naive user performance. In addition, the small number of participants in this study may bring random variance in the estimation of R^2 value.

5. General discussion

Results in Experiment 1 indicated that both the static model and the dynamic model could fit the observed task completion time well with minor modification. In addition, when users' average writing time was around 1500 ms for each frequently used character, the optimal segmentation time (no longer than 1040 ms) in these temporary segmentation recognizers was suggested in Experiment 1. After combining the static and dynamic models, an integrated model was obtained, which was tested with a real commercial handwriting recognizer in Experiment 2. It was found that it could still fit human performance satisfactorily under three different training conditions. The integrated model with mathematical inference held stronger power in predicting the in human performance of the handwriting recognizer than the empirical linear model.

5.1. Implication of the integrated performance model

(1) Comparison of the weights of the repairing time (T), average writing time per character (WT), segmentation time (ST), and recognition time (R) in the determining the performance when recognition accuracy (RA) and number of input characters (N) are constants

$$\begin{aligned}\text{Eq. (13)} \Rightarrow D &= N[(1/\mathrm{RA})(T + \mathrm{WT} + \mathrm{ST} + R) + 439 - T] \\ &\quad - (N/24)4973 \\ &= N\{[(1/\mathrm{RA}) - 1]T + (1/\mathrm{RA})\mathrm{WT} + (1/\mathrm{RA})\mathrm{ST} \\ &\quad + (1/\mathrm{RA})R + 232\}.\end{aligned}$$

Consequently, when RA and N are constants, there is a multiple linear relationship between T, WT, ST, R and D. Since the units of WT, ST, R and T are the same (ms), similar to the standardized regression coefficients in linear regression model, the coefficients ($[(1/\mathrm{RA}) - 1]$, $(1/\mathrm{RA})$, $(1/\mathrm{RA})$, $(1/\mathrm{RA})$) before each variable represent its respective weight in determining the task completion time (Zhang, 1986).

Since

$$[(1/\mathrm{RA}) - 1] < (1/\mathrm{RA}), \quad \mathrm{RA}(0, 1],$$

the order of the relative weights will be:

$$\mathrm{WT} = \mathrm{ST} = R > T.$$

(a) In determining the human performance, the relative importance of repairing time is less than that of the other three factors. It indicates that the work to reduce repairing time in these recognizers, e.g. gesture recognition, is not as important as the work to reduce the average writing time per character, segmentation time and recognition time of the software. This kind of difference in weights will keep constant in spite of the change of RA, because the difference of the weights is a constant ($(1/\mathrm{RA}) - [(1/\mathrm{RA}) - 1] = 1$).
(b) Reduction of the average repairing time per character (WT) will have the same importance as reduction of the recognition time (R). It indicates the work to smooth the physical interface of the handwriting board and the pen will save the same length of task completion time as the work to decrease the recognition time by improvement of recognition algorithms.
(c) Setting the minimal value of segmentation time is also important as the reduction of WT and R. Based on the analysis in the optimal segmentation time

found in Experiment 1, even though user's performance may not be impaired significantly when the ST was in the range from 190 to 740 ms, it was suggested that the ST should be set as minimal as possible.

(2) The relationship between RA and Task Completion Time (D) when T, WT, ST, R, N are constants can be shown as

$$\begin{aligned}(13) \Rightarrow D &= N\{[(1/\mathrm{RA})(T + \mathrm{WT} + \mathrm{ST} + R) + 439 - T] - 4973/24\} \\ &= N[(1/\mathrm{RA})(T + \mathrm{WT} + \mathrm{ST} + R) + 232 - T].\end{aligned}$$

In this case, there is an inverse relationship between recognition accuracy (RA) and task completion time (D)

$$\Delta(D) = \partial D/\partial \mathrm{RA} = -N(T + \mathrm{WT} + \mathrm{ST} + R)(\mathrm{RA}) - 2,$$

$$\mathrm{d}^2(D)/\mathrm{d}(\mathrm{RA})^2 = 2N(T + \mathrm{WT} + \mathrm{ST} + R)(\mathrm{RA})^{-3}.$$

Since

$$\mathrm{RA}(0, 1]; \quad N, T, \mathrm{WT}, \mathrm{ST}, \quad R > 0,$$

we have

$$2N(T + \mathrm{WT} + \mathrm{ST} + R)(\mathrm{RA})^{-3} > 0$$

and

$$\mathrm{d}^2(D)/\mathrm{d}(\mathrm{RA})^2 > 0, \quad \mathrm{RA} \in (0, 1].$$

Consequently, there is no turning point of the function between D and RA (Applied Mathematics Department, 1987). The shape of the function is a concave line (see Fig. 9). With the increase of RA, the reduction of the task completion time will be decreased (see Fig. 10).

(a) When recognition accuracy is from 50% to 93%, every 1% increase of RA will reduce task completion time from 1989 to 1915 ms.
(b) When recognition accuracy is from 94% to 100%, every 1% increase of RA will reduce D from 1895 to 1392 ms.

(3) *Some generalization of the integrated model*: It is possible that the elements of the user delay (UD) and repairing time (T) in the integrated model can be replaced by the other corresponding factors when part of tasks, user interface of handwriting recognizer and characters of the user changed.

User delay (UD) refers to the duration between the time when the system was ready for the next system input and the time when the user actually began to enter the next activity into the system. It can be replaced by the other cognitive operating time in other tasks in the domain of text editors. For example, in proofreading task, UD can be replaced by their average reaction time in pointing out the wrong character.

Repairing time (T), in this study, is the duration between the time when the wrongly recognized character appears on the screen and the time when the user

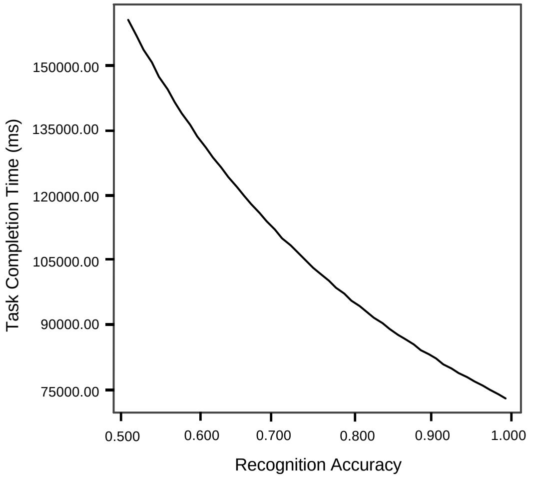

Fig. 9. The inverse relationship between recognition accuracy and task completion time (D).

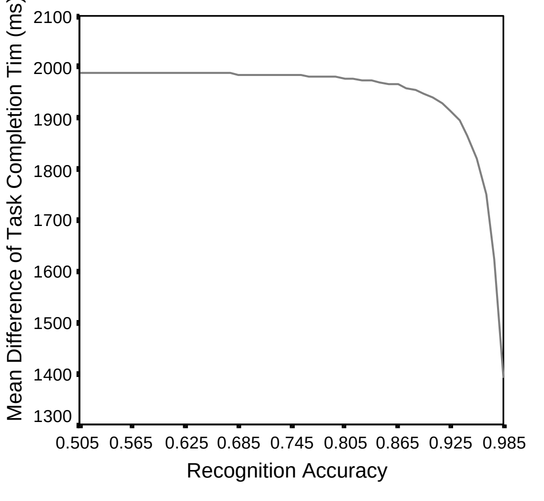

Fig. 10. The relationship between recognition accuracy and mean difference of task completion time.

pressed the "back space" key in the keyboard. It can also be replaced by the other action operating time when the user uses different strategies in repairing or the user interface offers different repairing methods. For instance, some handwriting recognizers offer a direct repairing method by the handwriting pen (e.g. pointing a button on the screen with the pen directly rather than pressing the "back space" key). Consequently, repairing time (T) in these recognizers would be the duration between when the wrongly recognized character appeared on the screen and when the user pressed that repairing button on the screen with the handwriting pen.

Meanwhile, both UD and T can be modified when different age and reaction time of the user changed.

However, even though there is some generalization made of the integrated model in this study, the integrated model is only concerned with inputting and repairing process of handwriting. In the whole process of inputting and editing, the editing actions of copying, cutting and pasting, etc. are also important which were studied by KLM and other versions of GOMS (John and Kieras, 1996). Thus, the current model cannot be applied to the whole editing process yet.

In addition, reference of the parameter of user delay (UD) was simplified from view of cognitive psychology. Gao et al. (1995)'s research which is concerned with the simple recognition time of individual characters; however, there was semantic context in each sentence in this study copied by the subjects. Semantic context provided the text priming to the recognition of the character, which will reduce the recognition time of individual characters (Sharkey and Sharkey, 1992; Zhang and Peng, 1990). Therefore, further empirical studies in getting the accurate parameter of UD in these recognizers and their various kinds of task domain are needed. Of course, the parameter of UD may play a less important role in determining the user performance because the degree of fitness of static model to the experimental data would not change greatly with the variance of the use delay ($M \pm 2$ s.d., $439 \pm 48 \times 2$ ms).

Acknowledgements

This research was funded by the Intel Research Funds of China Intel Research Center. The authors thank Dr. Yang Lei, who was the leader of the main project, for his great support during the research. Dr. Sun Xianhong also gave some valuable comments in this research. Many thanks to Ms. Wang Yunxia, Zhejiang Forest University, who spared no efforts to revise and reedit the manuscript.

References

Applied Mathematics Department, 1987. Advanced Mathematics. High Education Press, Tongji University, Beijing, pp. 264–287.

Beijing Language Research Institute, 1986. The Frequency of Usage Dictionary for Chinese characters. Beijing Language Research Institute.

Buskirk, R.V., LaLomia, M., 1995. The just noticeable difference of speech recognition accuracy interactive posters. Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems, Denver, Colorado, USA, May 7–11, Vol. 2, p. 95.

Casali, S., Williges, B.H., Dryden, R.D., 1990. Effects of recognition accuracy and vocabulary size of a speech recognition system on task performance and user acceptance. Human Factors 32 (2), 183–196.

Cheng, Y.F., 1996. Skills and Theory for Inputting Chinese Characters into Computer with Keyboard. Chinese Institute of Information, Tsing Hua University Press, Beijing, China, pp. 156–212.

Chien, L.F., 1995. Fast and quasi-natural language search for gigabits of Chinese texts natural language processing. Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA, USA, July 9–13, pp. 112–120.

Cho, Y., Morita, T., Higaki, S., Moriya, S., 1993. A pen-based system to input correct answers to assist in the development of recognition and understanding algorithms of ink data I. Software interfaces. Proceedings of the Fifth International Conference on Human–Computer Interaction, Orlando, FL, USA, August 8–13, Vol. 2, pp. 68–73.

Davis, R.C., Lin, J., Brotherton, J.A., Landay, J.A., Price, M.N., Schilit, B.N., 1998. A framework for sharing handwritten notes demonstrations. Proceedings of the ACM Symposium on User Interface Software and Technology, San Francisco, CA, USA, November 1–4, pp. 119–120.

Frankish, C., Hull, R., Morgan, 1995. Recognition accuracy and user acceptance of pen interfaces papers: pen interfaces. Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems, Denver, CO, USA, May 7–11, Vol. 1, pp. 503–510.

Gao, D.G., Zhong, Y., Ceng, L.J., 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.

Guerrier, J.H., Manivannan, S.N., 1999. The role of working memory, field dependence, visual search, and reaction time in the left turn performance of older female drivers. Applied Ergonomics 30 (2), 101–119.

Guo, Z.G., 1999. The Statistical Research Methods in Social Sciences. Chinese People's University Publisher, Beijing, China, pp. 8–171.

Hone, K.S., Baber, C., 1999. Modeling the effects of constraint upon speech-based human–computer interaction. International Journal of Human–Computer Studies 50 (1), 85–107.

Hou, J.T., Meng, Q.M., 2001. ANCOVA and Multivariable Analysis in Education and Psychology. Beijing Normal University Press, Beijing, pp. 84–87.

John, B.E., Kieras, D.E., 1996. Using GOMS for user interface design and evaluation: which techniques? ACM Transactions on Computer–Human Interaction 34, 287–319.

Joreskog, K., Sorbom, D., 1999. LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. Scientific Software International, Chicago, pp. 3–120.

Kalaswsky, R.S., 1999. VRUSE—a computerized diagnostic tool: for usability evaluation of virtual; synthetic environment systems. Applied Ergonomics 30 (1), 11–25.

Karl, L.R., 1993. Speech versus mouse commands for word processing: an empirical evaluation. International Journal of Man–Machine Studies 39, 667–687.

Krishnan, S.N., Moriya, S., 1993. Segmentation of handwritten text and editing-symbols from ink-data V. Hardware interfaces. Proceedings of the Fifth International Conference on Human–Computer Interaction, Orlando, FL, USA, August 8–13, Vol. 2, pp. 1010–1015.

LaLomia, M.J., 1994. User acceptance of handwritten recognition accuracy INTERACTIVE POSTERS. Proceedings of ACM CHI'94 Conference on Human Factors in Computing Systems, Boston, MA, USA, April 24–28, Vol. 2, p. 107.

Long, J., 1976. Effects of delayed irregular feedback on unskilled and skilled keying. Performance Ergonomics 19 (2), 183–202.

Lu, Y.S., Zhan, S.Y., Zhang, H., 1995. Design and implement of an algorithm for online recognition of hand writing Chinese character system. Journal of Beijing Institute of Technology 15 (5), 1–5.

Mackenzie, I.S., Chang, L., 1999. A performance comparison of two handwriting recognizers. Interacting with Computers 11 (3), 283–297.

Matthews, G., Davies, D.R., 2001. Differences in ability and performance. In: Matthews, G., Westerholm, S., Stammers, R. (Eds.), Human Performance: Cognition, Stress and Individual Differences. Psychology Press, Taylor & Francis Group, London, pp. 255–259.

Oomika, K.A.N., Nakagawa, M., 1997. Idea memo PDA in scalable handwriting interfaces interface design and evaluation. Proceedings of the Seventh International Conference on Human–Computer Interaction, San Francisco, CA, USA, August 24–29, Vol. 2, pp. 454–458.

Paddy, O., Draper, S.W., 1996. How machine delays change user strategies temporal aspects of usability. ACM SIGCHI Bulletin 28 (2), 39–42.

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

Schmidt, R.A., 1991. Principle of motor control and movement accuracy. In: Richard, A.S. (Ed.), Motor Learning and Performance Human Kinetics Books. Champaign, IL, pp. 101–126, 139–141.

Sharkey, J.C., Sharkey, N.E., 1992. Weak contextual constraints in text priming and word priming. Journal of Experimental Psychology: Learning Memory, and Cognition 18, 543–570.

Shneiderman, B., 1984. Response time and display in human performance with computer. Computing Suverys 16 (3), 265–285.

Shneiderman, B., 1992. Response time and display rate. In: Ben (Ed.), Designing the User Interface Strategies for Effective Human–Computer Interaction. Addison-Wesley Publishing Company, Chinese Version, pp. 205–223.

Spine, T.M., Williges, B.H., Maynard, J.F., 1984. An economical approach to modeling speech recognition accuracy. International Journal of Man–Machine Studies 21 (3), 191–202.

Teal, S., Rudnicky, A., 1992. A performance model of system delay and user strategy selection, CHI'92, Montegery, CA, USA, May 3–7, pp. 295–305.

Wang, Z.M., 1990. Research Methods in Psychology. Chinese People's Education Press, Beijing, pp. 110–125.

Wolf, C.G., 1990. Understanding handwriting recognition from the user's perspective computer systems: modeling. Proceedings of the Human Factors Society 34th Annual Meeting, Orlando, FL, USA, Vol. 1, pp. 249–253.

Wolf, C.G., Glasser, A.R., Fujisaki, T., 1991. An evaluation of recognition accuracy for discrete and run-on writing computer systems: handwriting, speech, and other input techniques. Proceedings of the Human Factors Society 35th Annual Meeting, Santa Monica, CA, USA, Vol. 1, pp. 359–363.

Wolfgang, H., Yang, J., 1998. Interactive error repair for an online handwriting interface. CHI'98, Los Angeles, CA, USA, April 18–23, pp. 353–354.

Wu, D.X., 2001. Inputting short messages with handwriting recognizer. Journal of Chinese Computer-world 1, available online at: http://www/pcworld.com.cn/2001/back_issues/2122/2202.b.asp

Wu, C.X., Zhang, K., 2000. Technique Report of the Engineering Psychology and Human Factors Lab. The Study of Pen–pencil Copy Task: Pilot study for the Project of the Handwriting recognition system CAS: (No: 20001101).

Youmans, D.M., 1993. The effects of system response time on users of interactive computer systems, In: Shneiderman (Ed.), Designing the User Interface Strategies for Effective Human Computer Interaction. Addison-Wesley Publishing Company, Chinese version, pp. 205–223.

Zhang, H.C., 1986. The Statistics in Psychological Research. Beijing Normal University Press, Beijing, pp. 308–340.

Zhang, K., 1992. A cognitive model of typing Chinese into computer. Journal of Chinese Information Processing 4, 5–399.

Zhang, H., Peng, D.L., 1990. The effect of context on the retrieval of abstract and concrete. Sentence Acta Sina Psycholgica 4, 391–396.