Considering Subjective Factors in Performance Models for Human-Computer Interface Design and Evaluation

Jing Kong

A dissertation submitted to
Kochi University of Technology
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Graduate School of Engineering
Kochi University of Technology
Kochi, Japan

March 2006

Abstract

Considering Subjective Factors in Performance Models for Human-Computer Interface Design and Evaluation

Jing Kong

Pointing tasks are basic and important for interaction in human-computer interfaces. A large numbers of candidate interfaces and devices for pointing tasks require reliable models for device evaluation and design guidance.

Fitts' law is one of the most famous models for pointing task in HCI and has been used widely. Unfortunately, it is established on a thin theoretic base and its adequacy remains debatable. The subjective factors or the humans individual factors cannot be indicated reliably by Fitts' law. Moreover, the two existing forms of Fitts' law complicate the application situation. We regard that there are two layers of speed-accuracy tradeoffs affecting the pointing performance, the task layer and the subjective layer. Through a series of special manipulated experiments, we analyzed the two forms of Fitts' law from the aspect of reconciling the two layers of speed-accuracy tradeoffs. Our investigation reveals the nature and relationship of the two layers of speed-accuracy tradeoffs in pointing tasks. Moreover, the analysis gives a comprehensive comparison of the two forms of Fitts' law to clean away the doubts and hesitation existing in model selection all along. This work helps us to realize that it is impossible to model the two layers of speed-accuracy tradeoffs in pointing tasks by the simple relationship described in Fitts' law models accurately and completely.

Then we established a new model for pointing tasks through analyzing the data of the performance time. We named the new model as SH-Model, which takes the human factors or, the subjective factors into consideration from the aspect of performance successful rates. A statistical tool, AIC, was applied for model evaluation. According to the AIC values, the SH-Model is better than the traditional models.

Around the topic of subjective factors in modeling, we have also studied the optimal effective target width calculation method of Fitts' law. Using the data from the experiment where the subjects were instructed to perform with different speed-accuracy tradeoffs, we explored the information processing or transmission rate of the human motor system. The application research of evaluation models on colors effects on pointing tasks had also been studied.

This work will contribute to considering the human or subjective factors in modeling

the performance. All these results will be beneficial for the future interface design and device evaluation in HCI.

key words   Human-computer interaction, pointing task, model, evaluation, Fitts' law, speed-accuracy tradeoff

Contents

Contents

List of Figures

List of Figures

List of Tables

Chapter 1

Introduction

1.1　Research Motivation

Human computer interaction (HCI) is a scientific branch of computer science also intertwining with the knowledge of physiology, psychology, motor control, philosophy, etc. Some typical performances, such as pointing, dragging and drawing, help users to accomplish the complex task of interacting with computers. Therefore, researchers in HCI have concentrated on devices design for these performances for a long time. For pointing task along, as a result of tremendous developing speed of computing technologies, numerous new devices appear to the market (such as mouse, pen, joystick, touchpad, trackball, eyetracker, etc.). Even the physiological organs like finger and eye can be adapted as input devices [48]. Actually, pointing tasks are not limited to selecting a target item from the screen. Tapping on keyboard and buttons on the mobile phones can also be studied in this category (see Figure 1.1).
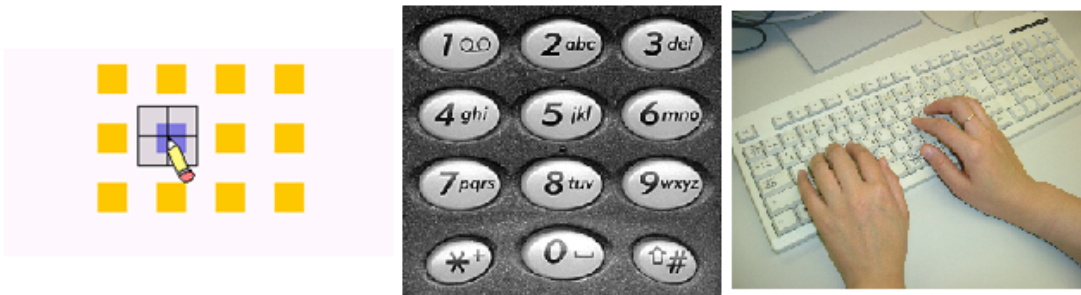
Fig. 1.1　Examples of pointing tasks

Facing the expatiatory catalogs of the devices and omnifarious choices candidates, not only the users, but also the system designers are puzzled. Some people prefer one certain kind of design, others may like another kind. We don't know whether the selections, or preferences are based on scientific researches, or purely user's own habit. Actually, some habits are really harmful. Hence, it is necessary to be sure that the applied tools are suitable for certain application, that means, people can maneuver the tool without great effort to adapt to the features of this tool, because the important thing

is that the design of a tool should be capable of inspiring human potential and helping them work happily and efficiently. We need tools congruent with human physiological and psychological characters. One of the most important tasks of the researchers in human computer interaction is to create tools friendly for human beings, not create tools which force people to be trained very hard. There should also be considerate designs for the handicapped. Therefore, it is crucial for us to find useful models to predict and evaluate people's performance using different devices and to test the feasibility and efficiency of the devices on interfaces of different systems.

A model is a simplification of reality [37], and it can be applied to afford guidance for design, propose standard for evaluation or provide a basis for understanding the motor behaviors. Unfortunately, modeling the performance in human computer interaction has always been a tough job. It is one reason that until now the most famous model for pointing tasks is still Fitts' law [18], which has been used to predict and evaluate the performance of rapid and aimed reciprocal movements in HCI since 1978 [9]. Nevertheless, the studies on the modeling works for pointing tasks will be beneficial not only for device evaluation, but also for affording guidelines for interface design. Therefore, considering the great importance of models in HCI, the studies on this topic are still worthy of our imperative efforts.

1.2   Background Knowledge

1.2.1   Fitts' Law and Speed-Accuracy Tradeoff

Most models in HCI can be categorized into two groups: the descriptive models with metaphoric characteristics (such as Guiard's model of bimanual control [21]) and the predictive models with mathematics rigors. Simply speaking, "the descriptive models provide a framework or context for thinking about or describing a problem or situation"[37]. It is not the focus in this dissertation.

Fitts' law belongs to the second group. Like Guiard's model, Fitts' law also emerged from basic research area in motor control, and like the other predictive models in HCI, it provide metrics to analyze the human performance mathematically without taking time-consuming and resource-intensive experiments.

Acclaimed as one of the most successful human performance models[40], Fitts' law has served as one of the few quantitative foundations for human computer interaction research. In particular, it has been used as a theoretical framework for computer input device evaluation (e.g. Card et al. [9]; ISO [25]; MacKenzie [34]), a tool for optimizing new interfaces (e.g. Lewis et al. [30]; MacKenzie and Zhang [36]; Zhai et al.[53]), as well as a logical basis for modeling more complex human computer interaction tasks (Accot

and Zhai[1]). Fitts' law has also inspired alternative interaction techniques (e.g. Accot and Zhai[3]; Kabbash and Buxton [26]; Zhai et al.[52]) and gained new understandings, expansions, and applications in human computer interaction research in recent years (e.g. Accot and Zhai[4]; Guiard et al.[22]; McGuffin and Balakrishnan[38]; Zhai et al.[54]).

Despite its impressive successes and critical importance, however, some of the fundamental issues in Fitts' law, either as a general human performance model or as a tool for human computer interaction research, are still not fully understood in the literature. Speed-accuracy tradeoff is one of them [*1]. In essence, Fitts' law is about revealing the rule of speed-accuracy tradeoff in human control performance.

For pointing tasks, Fitts' law precisely models how task precision affects pointing completion time.

Firstly, Fitts was seeking a universal formula to obtain the information capacity of the motor system that could be applied to pointing tasks under a wide variety of conditions. He developed some experiments including one direction pointing tasks and observed the human performance. In Fitts' paradigmatic experiment, subjects used a pen to reciprocally point to two strips separated from each other by some distance on a platform (see Figure 1.2).
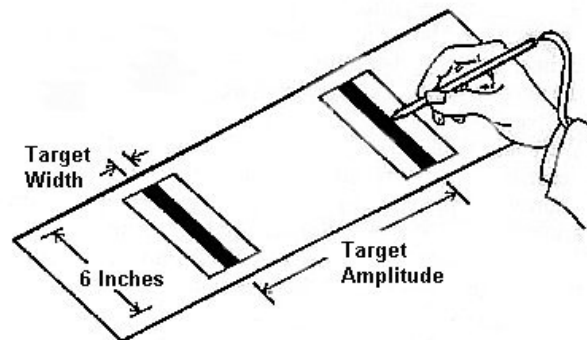
Fig. 1.2   Fitts' reciprocal pointing paradigm [34].

Fitts conceived that there was a constant information processing rate for the response from a certain kind of task, no matter how it was performed. He defined the concept of IP as the index of performance and used it to describe the information processing rate of the pointing task[18]. According to the data derived from the experiment, the first form of Fitts' law was established as follows:

[*1] As in many human-performed tasks, the more precisely the task is to be accomplished, the slower it is. Conversely, the faster the task is completed, the less precisely the task tends to be performed.

– **3** –

$$IP = \frac{ID}{MT} = -\frac{1}{MT}\log_2(\frac{W}{2A}) \tag{1.1}$$

ID: the index of difficulty of the task (measured in information unit bits);

MT: the expected average movement time for task completion;

A: the motor amplitude;

W: the target width limiting pointing accuracy.

After the descendent modifications of Equation.1.1, now the most popular form of Fitts' law is a direct analogy to Shannon theorem 17[31], since it is the most logical one at the boundary condition $A = 0$ [31]:

$$MT = a + b\log_2(\frac{A}{W} + 1) \tag{1.2}$$

where a and b are regression coefficients. We call Equation.1.2 ID model in this paper.

The ID of Fitts' law is changed as follows:

$$ID = \log_2(\frac{A}{W} + 1) \tag{1.3}$$

Note that so far the precision parameters, A and W, are *a priori* task parameters. In this sense the original Fitts' law is about the relationship of temporal (speed) performance and task precision. Ideally the human performer uses all of the precision tolerance (W) that the task specifies, no more, no less. Statistically, this means the spread of the endpoints (hits) of Fitts' aimed movements corresponds to the target width.

In actuality, either when performing laboratory experiments or when selecting graphical user interface (GUI) widgets on a computer with a mouse, the human performer (or the computer user in the context HCI) may or may not comply with the task precision as specified by W — the width of the endpoints dispersion may depart from the target width W, causing either over or under utilization of the target area. In other words, the performer may introduce another layer of precision choice relative to the nominal task precision. The performer may be biased towards accuracy and use less area than the target gives, resulting in a more accurate than necessary but slower performance. Conversely, the performer may be biased towards speed and use more area than the target gives, resulting in a faster but more error prone performance. This second layer (or component) of speed-accuracy tradeoff is subjective and personal (hereafter referred as the subjective layer). In contrast, the bottom layer is objective (task specified) and nominal.

1.2.2   Another Form of Fitts' Law Model

The existence of the subjective layer of accuracy complicates Fitts' law. Facing such a complication, students of Fitts' law have implicitly or explicitly taken two views. One is to continue to treat Fitts' law as a task model as in Equation 1.2 based on A and W, two nominal, *a priori*, and deterministic pointing task parameters. Alternatively, Fitts' law can be re-written based on actual behavioral parameters:

$$MT = a + b \log_2(\frac{A_e + W_e}{W_e}) \tag{1.4}$$

correspondingly,

$$ID_e = \log_2(\frac{A_e + W_e}{W_e}) \tag{1.5}$$

where,

$$W_e = 4.133\sigma \tag{1.6}$$

where A_e and W_e are *a posteriori* and statistical measurement of the actual movements [*2]. As practiced in the literature, they are called effective distance and effective target width, although the implication of the term effective is a biased one without proof. For distinction, we refer Equation 1.2 as the task form of Fitts' law and Equation 1.4 as the behavior form of Fitts' law.

In fact the method of using W_e to correct errors or including factual accuracy into consideration could be traced back to Crossman[12], who utilized the information theory that the information in a normal distribution is $\log_2 \sigma\sqrt{2\pi e}$, where σ is the standard deviation of distribution. A range of $\pm\frac{4.133}{2}\sigma$ includes about 96% of the distribution. Therefore, if error rate equals to 4%, $\log_2 W$ is an accurate representation of the information contained in the hits distribution. If the errors exceed 4%, W_e is greater than W. If errors are less than 4%, W_e is smaller than W (see Figure 1.3). Thus, the concept of W_e helps to include accuracy into Fitts' law and make it more reasonable from the behavior point. The behavior form of Fitts' law has been accepted by ISO standards 9241-9 [25].

Although some researchers support using W_e in Fitts' law[34], no study has proved that whether the *a posteriori* measurement is indeed cognitively, or quantitatively, equivalent to *a priori* task specification. Paul Fitts' own view on the theoretical basis of Fitts' law was not rigid. Empirically, there is a scarcity of evidence to prove or disprove whether and how well substituting ID with ID_e could compensate for the influence of

---

[*2] Since the difference of A vs. A_e is comparatively smaller than that of W vs. W_e, this study focuses on the latter.
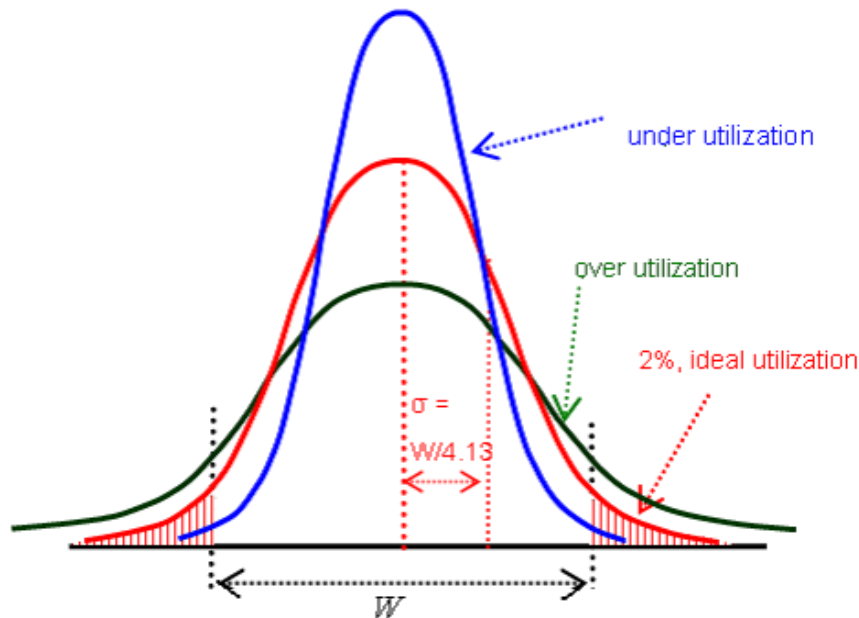
Fig. 1.3   Actual hits distribution indication figure (the real distribution maybe wider or narrower than the target width)

nominal and actual precision mismatch caused by the subjective layer of speed-accuracy tradeoff. It will be empirically informative to measure whether human performance adjusted with effective width is equivalent to the performance under a nominal width of the same size, had the performer complied with the exact target specification.

There are both practical and theoretical implications to each of the two forms of Fitts' law. For practical purposes, particularly for work in HCI, the desirability of task vs. behavior form of Fitts' law isn't clear cut. Often it is desirable to relate a user's performance to the geometry of a graphical user interface. For example, for an interface designer it is important to know a user's average selection time as a function of task parameters such as a GUI widget in certain size and location. For another example, in Fitts' law based stylus keyboard optimization research researchers are concerned tapping time in relation to the geometrical variables – the relative location of keys (e.g. Lewis et al., [30]; MacKenzie and Zhang, [36]; Zhai et al., [53]). A task form of Fitts' law is more desired in these cases. On the other hand, it is logically difficult to expect Fitts' law in its task form to serve as a reliable tool in evaluating the performance of an input device characterized by a and b constants in Fitts' law ([55][34]) if the performer's actual pointing precision deviates too far from the specified task precision. In this case, it is reasonable to expect that Fitts' law in its behavioral form, or at least some modification of the task form of Fitts' law in consideration of the deviation, be more useful.

1.3   Objectives and Research Issues

We will explore the problems existing in modeling pointing tasks from analyzing the human or subjective factors in the performance. Therefore, the two layers of speed-accuracy tradeoffs will be analyzed thoroughly, especially the second layer of the speed-accuracy tradeoff incurred by performers. Through revealing the nature of speed-accuracy tradeoff in pointing tasks and the relationship between system factors and human's factors, we will be able to compare the two models introduced previously (the task form and behavior form of Fitts' law). The comprehensive comparison of these two models will in turn help us to understand pointing tasks more deeply.

Then we will go on to try to establish a model in which system factors as well as human factors can be demonstrated respectively. Such a model should also be able to disencumber the performers or subjects from keeping an error rate of 4% during the task. The application of this model will be testified.

We will also try to resolve other related questions in pointing tasks, such as the optimal calculation methods of the effective target width, the information transmission rate analysis in pointing tasks, and the colors' effects on pointing tasks.

All these studies will afford us an opportunity to understand the features of pointing tasks comprehensively. The results will be instructive to study about other kinds of motor behaviors in HCI.

1.4   Dissertation Structure

The structure of this paper is shown by Figure 1.4.

Chapter 2 in this dissertation scrutinizes the effects of the two forms of Fitts' law (Equation 1.2 and Equation 1.4) based on the analysis of speed-accuracy tradeoff. The results of a series specially controlled experiments will be reported to help us observe the subjects' reciprocal pointing performance with different level of speed and accuracy inclinations incurred by experimenters' instructions. These experiments will testify the existence of the two layers of speed-accuracy tradeoffs and their respective impact on performance. Through these experiments we can also compare the two forms of Fitts' law and know whether the two models can reconcile the two layers of speed-accuracy tradeoff completely.

The study in Chapter 2 inspired us to do the studies in the following chapters (Chapter 3 and Chapter 4). Some basic knowledge of the Fitts' law models is discussed.

Then Chapter 5 will resolve the whole problem from a new horizon. A new model named as SH-Model will be expounded in this chapter. The model will be compared with the traditional Fitts' law models by a statistical method AIC. We will also introduce
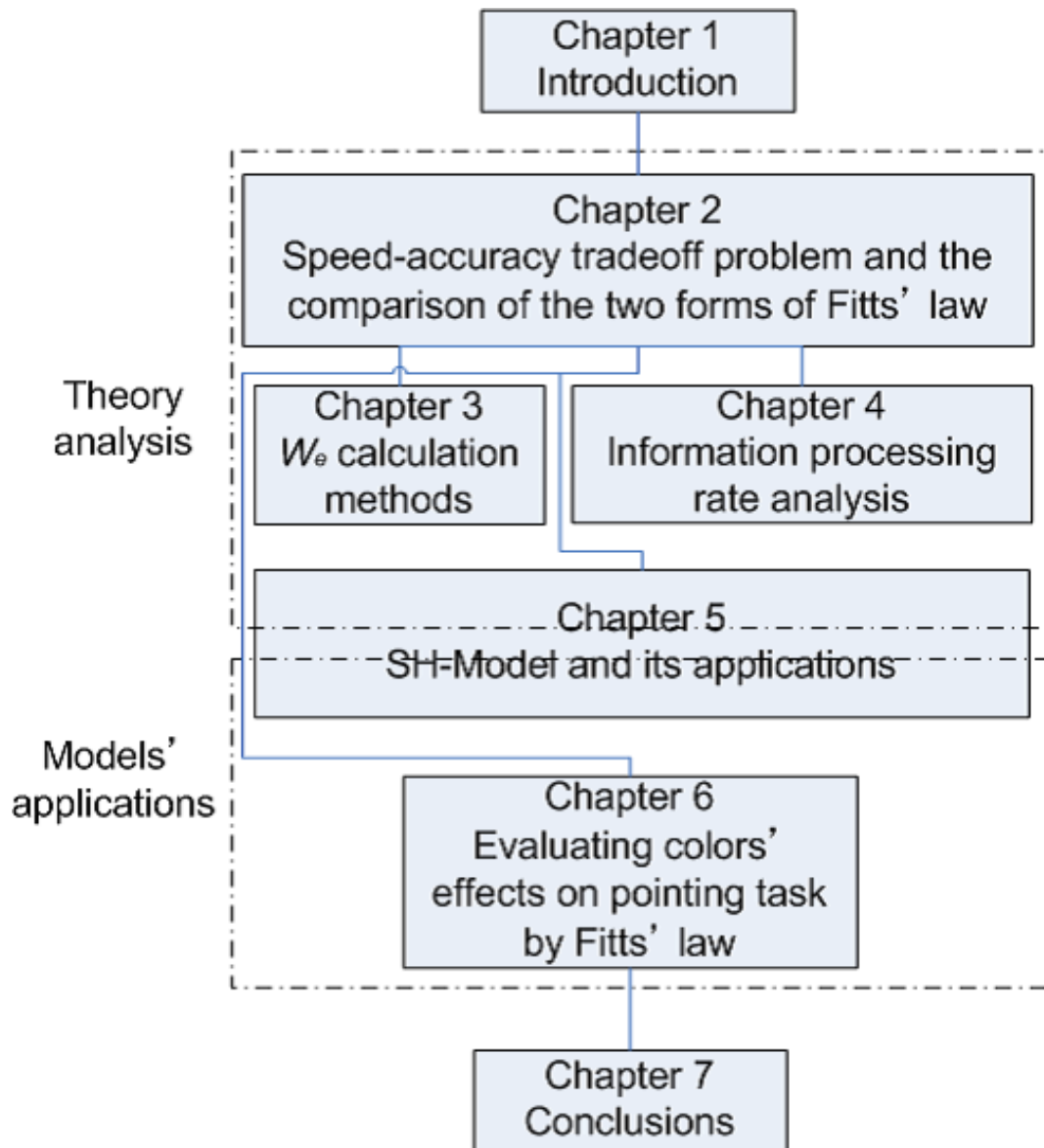
Fig. 1.4   Dissertation structure

the application of this model.

Finally, we will discuss another application of Fitts' law in pointing task—comparing the color's effects on pointing task. This is a first application of Fitts' law model in colors' effects evaluation in HCI.

1.5   Summary

The research reported in this dissertation represents an initial exploration of considering human or subjective factors in modeling the performance in HCI.

1.5 Summary

While a great deal of additional research will be needed to verify and extend the results and guidelines reported in the following chapters, the research conducted here provides an initial framework for further development of the methodologies and guidelines to include human or subjective factors in establishing the HCI performance models.

Since numerous new interfaces applying pointing devices appear to the market, we need to use models to give clear evaluation of the different devices and reliable guidelines for more adaptable interfaces design. The research introduced in this dissertation provides a first step for considering the human's individual subjective factors in models establishment in HCI.

This is a blank page.

Chapter 2

Speed-Accuracy Tradeoff in Fitts' Law Tasks

Pointing tasks in human computer interaction obey certain speed-accuracy trade-off (SAT) rules (refer to Section 1.2.1). In general, the more accurate the task to be accomplished, the longer it takes, and vice versa. Fitts' law models the speed-accuracy tradeoff effect in pointing as imposed by the task parameters, through Fitts' index of difficulty based on the ratio of the nominal movement distance and the size of the target. Operating with different speed or accuracy biases, performers may utilize more or less area than the target specifies, introducing another subjective layer of speed-accuracy tradeoff relative to the task specification. A conventional approach to overcome the impact of the subjective layer of speed-accuracy tradeoff is to use the *a posteriori* "effective" pointing precision W_e in lieu of the nominal target width W. Such an approach has lacked a theoretical or empirical foundation. This chapter investigates the nature and the relationship of the two layers of speed-accuracy tradeoffs by systematically controlling both the index of difficulty and the index of target utilization in a set of three experiments. Their results show that the impacts of the two layers of speed-accuracy tradeoffs are not fundamentally equivalent. The use of W_e could indeed compensate for the difference in target utilization, but not completely. More logical Fitts' law parameter estimates can be obtained by the W_e adjustment, although its use also lowers the correlation between pointing time and the index of difficulty. The study also shows the complex interaction effect between the index of difficulty and the index of target utilization, suggesting that a simple and complete model accommodating both layers of speed-accuracy tradeoffs may not exist.

2.1   Pre-analysis

As we have introduced in Chapter 1, Fitts' law in the task form (Equation 1.2) cannot model the pointing performance accurately in the case of different speed or accuracy inclination incurred from either task requests or subjective tendencies, since the appointed target width is not used ideally with an error rate of 4%. Therefore,

another derivation of Fitts' law (Equation 1.4) is applied to compensate the subjective layer of speed and accuracy tradeoff in pointing task.

For clarity of notations, in this chapter, we denote the index of difficulty based on the nominal target width, as defined by Equation 1.3, I_{dn}, and use I_d as a generic reference to index of difficulty. Meanwhile, we also change the format of ID_e (see Equation 1.5 in Introduction) into I_{de}.

We are trying to see whether there is a potentially more complete and more sophisticated approach is to represent the two layers, or two components, of speed-accuracy tradeoffs in one model. Such a model should relate time (or speed) to two independent factors, both concerning precision. The first is the task precision, as specified by target parameters A and W. The second independent factor is the degree of target area utilization the performer chooses. This is relative to the specified task precision and subjectively introduced by the performer. The performer may choose to be less precise than the task specification and use more area than W and hence gain a faster speed, or choose to be more precise than the task specification and use less area than W, causing a slower completion speed. Another goal of this study is to explore whether there exists a model of Fitts' law that explicitly relate time to both layers of speed-accuracy tradeoffs.

The degree of target area utilization has to do with the risk of missing the target the performer is willing to take. A more risky behavior tends to result in a wider endpoints distribution and cause over-utilization of the target area. A more risk-averse behavior tends to result in a narrower endpoints distribution and cause under-utilization of the target area.

Aiming for the "standard behavior" of approximately 4% error rate, most Fitts' law studies instructed the participants to perform "as fast as possible and as accurately as possible", but do not systematically vary or control experimental participants' actual precision relative to the nominal task precision; hence it is difficult to evaluate the influence of the subjective layer of speed-accuracy tradeoff and the correction effects using I_{de}. There have been very few exceptions ([20][5]) where the subjects' performance was controlled by different requirement on speed and accuracy, but the purposes of these studies were completely different from this study, and the variations on speed and accuracy in the studies are very limited.

The focus of this study is on the modeling aspects of speed-accuracy tradeoff in aimed movements. The first task of is to empirically evaluate the I_{de} based behavior form of Fitts' law, when the target utilization levels are controlled to both sides of the ideal case. For clarity and convenience, we first formally define the notation of target width utilization. It is logical to assume that movement time is a function of both nominal task difficulty as quantified by I_{dn} and the level of target width (over)

utilization by the performer, as quantified by I_u. Without committing to a particular form of the function, we have:

$$MT = f(I_{dn}, I_u) \tag{2.1}$$

where the index of target width (over) utilization I_u is formally defined as:

$$I_u = \log_2(\frac{W_e}{W}) \tag{2.2}$$

The logarithmic transformation here is merely for mathematical convenience and symmetry with I_d. The absolute value of I_u indicates the degree to which the actual spread of the endpoints departs from the specified target width. A positive I_u means the performer over utilizes the target width and misses the target (error) with more than 4% probability. A zero I_u means the performer perfectly utilizes all variability specified by the task, no more, no less, and the error rate is exactly 4%. A negative I_u means that the performer under utilizes the target area, leaving a certain amount of safety margin, and misses the target with less than 4% probability.

In order to thoroughly study speed-accuracy tradeoff in pointing tasks, in particular whether the impact of a non-zero I_u is compensated by replacing the task form of Fitts' law (Equation 1.2) with a behavior form of Fitts' law (Equation 1.4), empirical studies which systematically manipulate both index of difficulty I_{dn} and index of target utilization I_u have to be conducted.

2.2 Experiment SAT1

2.2.1 Set-up and Design

Twelve volunteers, of different gender (9 male and 3 female) and age (21 to 38 years, mean 26), participated in a target pointing experiment on a tablet computer (FUJITSU FMV Stylistic) with a screen size of 21 cm x 15.6 cm. Each pixel on the screen was 0.2055 mm wide. Similar to Fitts' original experiment [18], participants did reciprocal pointing on a pair of vertical strip targets with a stylus. The width (W) of the targets and the center-to-center distances (A) between the two strips were set at $W = 12, 36, 72$ pixels and $A = 120, 360, 840$ pixels. The order of the nine width and distance combinations was randomized. Twelve trials were presented in each W, A combination, with the first tap excluded in analysis. If tapped on the outside of the target, an auditory signal was played. Each participant was instructed to repeat the experiment three times with different operating conditions biased toward accuracy or speed: accurate (A), neutral (N), and fast (F). They were instructed to tap the targets "as accurately as possible" in Condition A, "as accurately as possible and as

fast as possible" in Condition N, and "as fast as possible" in Condition F. The goal was to make the participants operate at different levels of target utilization. The order of the A, N, F conditions was balanced by a Latin square pattern across the twelve participants.

2.2.2   Data Processing and Analysis

Occasionally, "accidental clicks" outside the general region of the target were registered, due to either the confusion of the participant, or instrument error. We used two simple and conservative rules to remove these "outliers" from further analysis to prevent their disproportional impact on modeling [43]. The first rule of removal was that the user hit the same target the trial started from or the user landed in the direction opposite to where the destination target was. This was determined by the distance between the hit point and the target center being greater than $A - W/2$. Twenty-five trials were removed by this rule. The second rule was that the distance of the endpoint to the target center was 8 times greater than the target size. Three additional trials were removed by the second rule. A total of 28 trials were removed by these two rules, constituting a small percentage of the total number of trials (3564).

2.2.3   The Index of Target Utilization I_u

The instructions for the operating strategy in the three experimental conditions had an obvious impact on participants' target utilization and error rate. The average error rate in the A, N, F conditions was 3.2%, 10% and 19.4%, respectively. These rates overall were higher than what we hoped. Ideally the error rate in the N condition would be around 4% and A and F conditions be on the two opposite sides. A wide range of target utilization levels was taken in the experiment by the participants (Figure 2.1). I_u varied from -1 to 1.5 bit. Note that 1 bit of I_u change means that the spread of the endpoints is twice or half of the specified target width. The operating conditions (A, N, F) changed the overall I_u level. Furthermore, participants also shifted from target under utilization in low I_{dn} to target over utilization in high I_{dn} trials, regardless of the condition. Even under the N (neutral) condition, I_u was not maintained at the ideal level (zero bit). This could be a deliberate choice of strategy shift by the human performers, or it could be fundamentally difficult to maintain the same level of target utilization facing targets with different I_{dn}. Guiard (2002) explains the same effect from a power constraint vs. precision constraint perspective, albeit in different terminologies. Note also that the amount of shift caused by A and W change were not the same (Table 2.1).
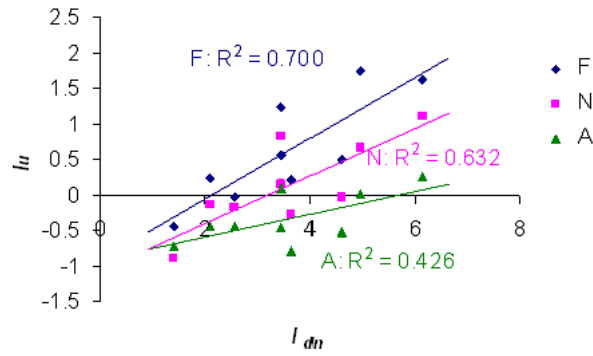
Fig. 2.1   The index of target utilization I_u changes with instruction and I_{dn}.

Table 2.1   Mean I_u values in different levels of A or W

	Instruction Bias			
	A	N	F	Mean
A 120	-0.36	-0.07	0.34	-0.03
A 360	-0.30	0.21	0.76	0.23
A 840	-0.35	0.27	0.78	0.23
W 12	0.12	0.86	1.53	0.84
W 36	-0.47	-0.01	0.43	-0.02
W 72	-0.65	-0.44	-0.08	-0.39
Mean	-0.34	0.14	0.63	0.14

2.2.4   The Nominal I_{dn} Model

As a baseline for further analysis, we first applied the basic task form of Fitts' law (Equation 1.2) to the data collected, using the task's nominal index of difficulty I_{dn}. The result is shown in Figure 2.2.
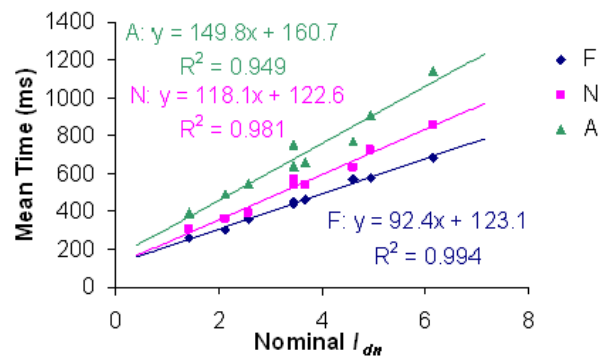
Fig. 2.2   Linear regression MT vs. I_{dn} in Experiment SAT1.

Table 2.2   Summary of MT vs I_{dn} regression in Experiment SAT1

Strategy	a	b	r^2
A	160.7	149.8	0.949
$p(F_{1,7})$	0.013	<0.0001	<0.0001
N	122.6	118.1	0.981
$p(F_{1,7})$	0.001	<0.0001	<0.0001
F	123.1	92.4	0.994
$p(F_{1,7})$	<0.0001	<0.0001	<0.0001
Mixed	135.4	120.1	0.696
$p(F_{1,23})$	0.036	<0.0001	<0.0001

When analyzed separately, within each operating condition MT and I_{dn} correlated strongly. 95% to 99% of variance in MT could be accounted for by the change of I_{dn}. There was a tendency that the more risky (faster-paced) the operating condition was, the stronger the correlation between MT and I_{dn} was. The robustness of I_{dn} prediction here is quite remarkable given the very different operating biases (different levels of target width utilization) in these conditions. However, the coefficients of the regression results (or a and b in Equation 1.2) varied from one condition to another. Table 2.2 summarizes the results. The statistical significance levels (p, based on F tests) are also listed under each parameter, and the overall regression significance is listed under the r^2 value of the regression.

Results in Table 2.2 clearly show that Fitts' law regression coefficients a and b using the I_{dn} model were influenced by the operating conditions (biases). b varied 62% from Condition F to A in this experiment.

If we perform the same Fitts' law regression based on the data from all conditions mixed, while still keeping the same unit of analysis 8 (I_{dn} levels) x 3 (operating conditions), we obtain a much weaker correlation ($r^2 = 0.696$), although the correlation is still statistically significant.

2.2.5   The Effective I_{de} Model

We now test if the use of effective width, the behavior form of Fitts' law, would be able to compensate for the difference of operating conditions (strategy biases). The linear regression results between mean trial completion time and the effective index of difficulty I_{de} are shown in Figure 2.3 and Table 2.3.

Comparing Figure 2.2 vs. 2.3 and Table 2.2 vs.2.3, the following observations can be made on the use of I_{de}.

First, the regression coefficients under different operating conditions were much closer to each other with I_{de}, showing the effect of compensation for the different levels
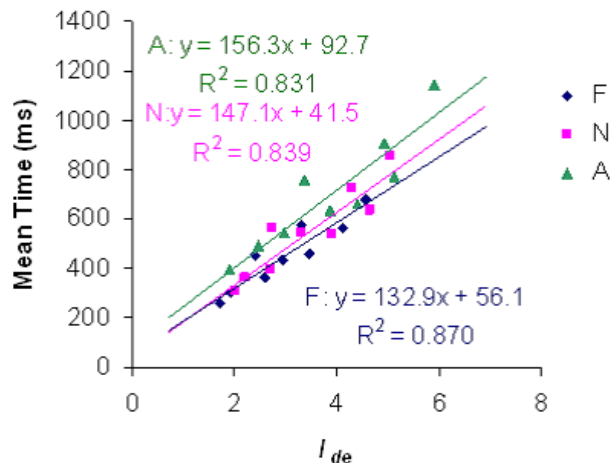
Fig. 2.3 MT vs. I_{de} regression in Experiment SAT1.

Table 2.3 Summary of MT vs I_{de} regression in Experiment SAT1

Strategy	a	b	r^2
A	92.7	156.3	0.831
$p(\mathrm{F}_{1,8})$	0.42	<0.001	<0.001
N	41.5	147.1	0.839
$p(\mathrm{F}_{1,8})$	0.65	<0.001	<0.001
F	56.1	132.9	0.870
$p(\mathrm{F}_{1,8})$	0.39	<0.001	<0.001
Mixed	13.2	161.0	0.825
$p(\mathrm{F}_{1,25})$	0.81	<0.0001	<0.0001

of target utilization. The a values from different conditions were all reduced and the b values from different conditions became much closer to each other. From Conditions F to A, the b values' change was now 17.6% (in comparison to 62%).

Second, within each operating condition, between MT and I_{de} decreased from the corresponding r^2 between MT and I_{dn}. Only 83% to 87% of the MT variance could be accounted for by I_{de} (in comparison to 95% to 99% by I_{dn}). The same trend of r^2 reduction can be observed in the data of [20]. Why this was true will be discussed later. A counter example to this trend was seen in MacKenzie's recalculation of Fitts' 1954 data [32], which found a slight increase in r^2 by using I_{de} (from 0.983 to 0.99 and 0.98 to 0.988 respectively). Note that there the initial correlation was very high, and the change of correlation was small. Furthermore, since Fitts' 1954 data did not have endpoints location recordings, the recalculation of W_e was based on Z-score conversion from error rates—an imprecise or arbitrary estimation method when the error rate is low or zero.

Third, overall there was a shrinkage of the range of the independent variable from I_{dn} to I_{de}. Within the same condition, particularly for the more risky condition F, there was a counter-clockwise rotation of the regression line.

Fourth, if we use data from all operating conditions together, the r^2 value between MT and I_{de} increased from the corresponding r^2 value between MT and I_{dn} regression (0.723 to 0.825), showing a stronger regularity of $MT = f(I_{de})$ than $MT = f(I_{dn})$ in modeling pointing time in the presence of a wide range of target utilization.

In summary, the use of I_{de} demonstrated both benefits and drawbacks. It compensated operating biases (more converging a and b parameters), but not completely. Its robustness as a determinant of pointing time as measured by r^2 decreased within each operating condition but increased across conditions.

The utilization level in this experiment overall tilted to over utilization (positive I_u, see Figure 2.1). Even in Condition A, the overall error rate was 3.2%. It will be informative to also observe the more conservative under utilization side, as in the next experiment.

2.3   Experiment SAT2

The experiment had two operating conditions: A and F. In Condition F, the instruction was, "Move as fast as possible. It is okay if a few errors are made". A gentle "ding" sound was played when an error was made. In Condition A, which in fact could be called the EA (extremely accurate) condition, participants were instructed to "Try to avoid any errors", and a loud "ding" sound was played when a target was missed.

Eleven people of different gender (4 female and rest male) and age (20's to 50's), who had not been in the first experiment, participated in this experiment with both conditions. They were alternated between A to F and F to A order. An LCD display with stylus touch-sensitive surface (Wacom LCD Tablet Model PL-400) was used as the experiment apparatus. The rest of the experiment setup remained the same as Experiment SAT1.

Due to the quality of the tablet used in this experiment, many more erroneous trials were found. If the stylus struck the surface of the tablet too hard (and quickly bounced up), no click was registered. This meant the next click could be aimed at the "wrong" target. Using the same two rules outlined in Experiment SAT1, out of a total of 2178 trials in this experiment, 160 trials were removed by the first rule (tapped on the wrong side). Another 74 trials were removed by the second rule.

Under the instruction given in this experiment, participants indeed exhibited more conservative (risk-averse) behavior. Figure 2.4 shows the index of target utilization (cf. Figure 2.1). The I_u values at almost all I_{dn} levels were negative (the average error rates

2.3 Experiment SAT2

Table 2.4   Fitts' law regression results in Experiment SAT2

	Bias	a	b	r^2	p
$MT = a + bI_{dn}$	A	110.5	172.6	0.932	<0.0001
	F	111.0	119.0	0.921	<0.0001
	Mixed	110.8	145.8	0.747	<0.0001
$MT = a + bI_{de}$	A	-28.1	182.2	0.896	<0.0001
	F	48.0	130.7	0.907	<0.0001
	Mixed	-19.1	165.1	0.825	<0.0001

were 0 and 0.5% respectively for A and F conditions, see Table 2.4), but participants were clearly more conservative in Condition A than in Condition F. We tried to bias the participants in this experiment to the direction opposite to Experiment SAT1 that was overall on the risky side. Note that the operating condition labels (A, F, etc) are relative within each experiment.
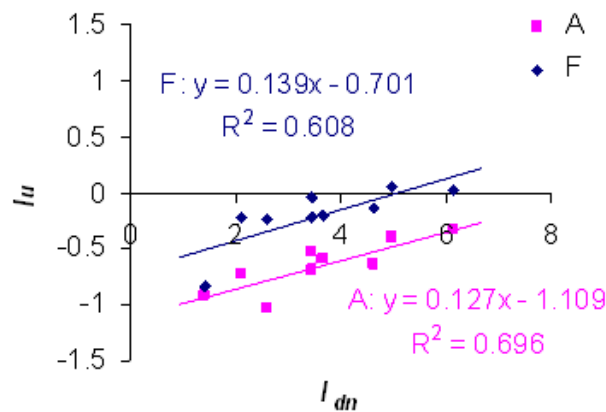
Fig. 2.4   The index of target utilization I_u in Experiment SAT2.
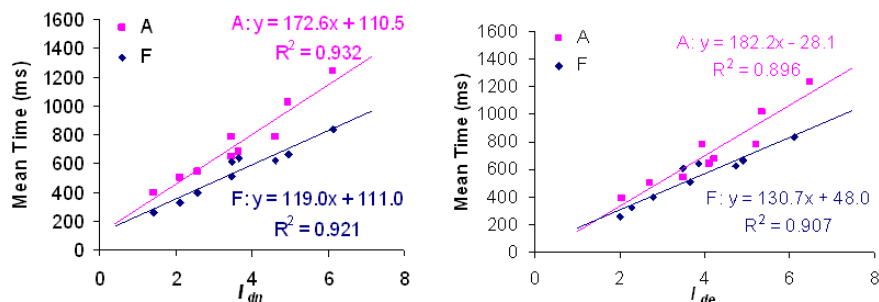
Fig. 2.5 MT vs. I_{dn} and MT vs. I_{de} linear regressions results of Experiment SAT2.

Table 2.4 and Figure 2.5 show I_{dn} and I_{de} model regression results. Observations similar to those in Experiment SAT1 can be made here, but to a lesser degree: (1) As indicated by r^2 values I_{dn} is a (slightly) more robust pointing time determinant than I_{de} within each condition; (2) I_{de} is more robust than I_{dn} across conditions. (3) The range of I_{de} shrank from that of I_{dn}, but only from the low end of the index of difficulty this time. In fact for the A condition, the high end of I_{de} was further extended to the right. (4) I_{de} yielded more converging a and b parameters between the conditions than I_{dn}, hence it compensated for the different target utilization levels in different operating bias conditions (but not completely and to a lesser degree than in Experiment SAT1). Overall the changes caused by substituting I_{dn} with I_{de} are lesser in this experiment.

To verify the observations in these two experiments, we decided to conduct a more comprehensive experiment that covers a wide target utilization range on both sides.

2.4   Experiment SAT3

2.4.1   Set-up and Design

Fifteen volunteers, 5 female and 10 male, aged 20 to 36 years old, participated in Experiment SAT3, in which the same experimental apparatus, software, and procedure as in Experiment SAT1 were used. The difference is that a greater range of target utilization is inducted: each participant was instructed to repeat the experiment five times with different operational strategies: extremely accurate (EA), accurate (A), neutral (N), fast (F) and extremely fast (EF). The following verbal instructions corresponding to each task were given by the experimenter to the participants: Perform as accurately as possible and dont worry about time or speed; try to avoid any error in Condition EA; as accurately as possible but keep some speed in Condition A; as accurately as possible and as fast as possible in Condition N; as fast as possible but keep some accuracy in Condition F; and as fast as possible and some errors are acceptable in Condition EF.

2.4.2   Data Processing and Basic Results

Similar to Experiments SAT1 and SAT2, 16 accidental trials were removed from the data pool out of a total of 7425 trials.

The error rates of the five conditions varied according to the verbal instructions, and the overall error rates in the EA, A, N, F, EF conditions were 0%, 1%, 4%, 9%, 22%, respectively, which were rather ideal because in condition N the error rate was at the standard 4% and the rest of the conditions were distributed symmetrically. As shown in Table 2.5, the overall I_u levels of the first two experiments were either tilted to positive (Experiment SAT1) or negative (Experiment SAT2). This experiment is

Table 2.5   The I_u range and error rate in each operating condition of the first three experiments

Bias	Experiment SAT1			Experiment SAT2			Experiment SAT3		
	Error rate (%)	I_u max (bit)	I_u min (bit)	Error rate (%)	I_u max (bit)	I_u min (bit)	Error rate (%)	I_u max (bit)	I_u min (bit)
EA							0	-0.63	-1.33
A	3.2	0.24	-0.78	0	-0.33	-1.04	1	-0.17	-1.22
N	10	1.11	-0.88				4	0.23	-0.61
F	19.4	1.74	-0.44	0.5	-0.05	0.83	9	0.81	-0.73
EF							22	2.04	-0.59

similar to a combination of the first two.

The basic results of Experiment SAT3 with regard to the impacts of I_{de} vs. I_{dn} as the determinant of mean trial completion time are summarized in Figure 2.6 and Table 2.5. The results in this more comprehensive experiment verified the trends observed in the first two experiments. First, I_{dn} was once again shown to be a remarkably robust determinant of the mean pointing time within each condition. In spite of the very different instructions and hence the very different levels of overall target utilization, the r^2 values of MT vs. I_{dn} linear regression were all above 0.9. There was a trend of increasing r^2 value from the more risk-averse conditions to the more risky conditions, consistent with the first two experiments. The r^2 values of MT vs. I_{de} linear regression in each condition were uniformly lower than their corresponding r^2 values of MT vs. I_{dn} in the same condition (Figure 2.6). Within the same operating condition, I_{dn} was clearly a stronger determinant of mean pointing time than I_{de}.
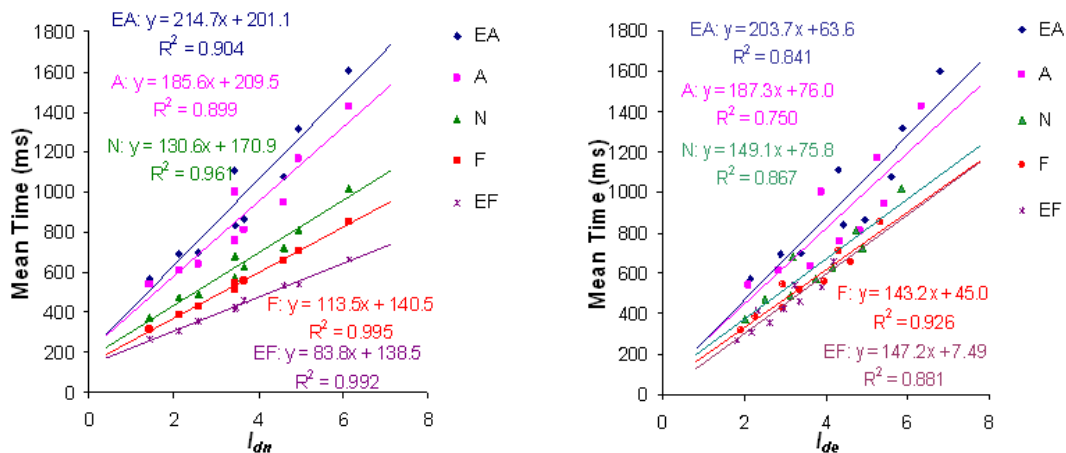
Fig. 2.6 MT vs. I_{dn} (left) and MT vs. I_{de} (right) linear regression in Conditions EA, A, N, F and EF in Experiment SAT3.

Second, in contrast to the strength of I_{dn} within each condition, I_{de} is a stronger determinant than I_{dn} when data from all conditions were merged in one regression.

Table 2.6   Summary of MT vs. I_{dn} and MT vs. I_{de} linear regression in Experiment SAT3

Tasks	Error	I_{dn} Model			I_{de} Model		
	(%)	a	b	r^2	a	b	r^2
EA	0	201.1	214.7	0.904	63.6	203.7	0.841
p		0.09	<0.0001	<0.0001	0.7	<0.001	<0.001
A	1	209.5	185.6	0.899	76.0	187.3	0.750
p		0.05	<0.0001	<0.0001	0.69	0.0025	0.0025
N	4	170.9	130.6	0.961	75.8	149.0	0.867
p		0.003	<0.0001	<0.0001	0.886	<0.001	<0.001
F	9	140.5	113.5	0.995	45.0	143.2	0.926
p		<0.0001	<0.0001	<0.0001	0.45	<0.0001	<0.0001
EF	22	138.5	83.8	0.992	7.49	147.2	0.881
p		<0.0001	<0.0001	<0.0001	0.90	<0.001	<0.001
Mixed		172.1	145.6	0.460	-92.4	207.5	0.783
p		0.07	<0.0001	<0.0001	0.17	<0.0001	<0.0001

I_{de} accounted for 78% of the variance of mean trial completion time caused by both different levels of index of difficulty and the very different five operating strategies (See Figure 2.7). In comparison I_{dn} could account for only 46%. In this sense I_{de} clearly has the ability to convert some impact of the different levels of target utilization to index of difficulty.

Third, there was an overall rightward shift of I_{de} values from their corresponding I_{dn} values at the low (left) end of index of difficulty, but on the high end, the shift depended on the operating condition. For Condition EF and F, the high end I_{de} points moved towards left, for Condition EA and A the high end I_{de} points actually further extended to the right, for condition N there was little change.

Fourth, and perhaps most strikingly, the regression lines of MT vs. I_{de} were much closer across different conditions than those of MT vs. I_{dn}, particularly for the more risky (faster) conditions (N, F, EF) (See Figure 2.6). The regression coefficients a and b hence were more converging between the conditions with Ide than with I_{dn}. Across the five conditions (from EA to FF), the a values were between 139 ms and 209 ms for the I_{dn} model and between 7.5 ms and 76 ms for the I_{de} model. The b values were between 83 ms/bit and 214 ms/bit (158% difference) for the I_{dn} model and between 143 ms/bit and 203 ms/bit (42% difference) for the I_{de} model. This also supports that I_{de} could at least partially overcome the different levels of target utilization due to operating biases and produce more stable estimates of Fitts' law coefficients.
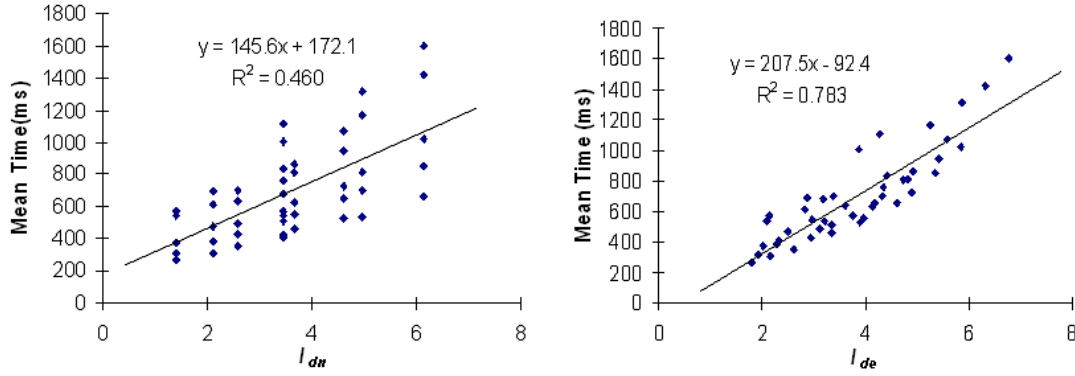
Fig. 2.7 MT vs. I_{dn} and MT vs. I_{de} of all conditions combined.

2.4.3   The Interplay of I_u, I_{dn}, and W_e

We now examine the distribution of I_u, which is related to the third point above on the range and location of I_{de} shift. Figure 2.8 shows I_u across different I_{dn} values in each of the five conditions in Experiment SAT3. As a result of different operating conditions, target utilization levels as measured by I_u shifted up or down over a wide range (-1.33 to +2.04 bit) as a result of the instruction bias. Furthermore, I_u also changed with I_{dn}. While overall I_u was correlated with I_{dn} (the higher the I_{dn}, the higher the level of over utilization the participants tended to make), the degree of such a dependency changed with the operating condition. The more risky (faster-paced) the overall strategy was, the stronger the dependency was. In Condition EF the dependency was the strongest, with $r^2 = 0.831$. In Condition EA, in contrast, participants were quite consistently risk-averse, keeping I_u well below 0 across all I_{dn} values ($r^2 = 0.14$).
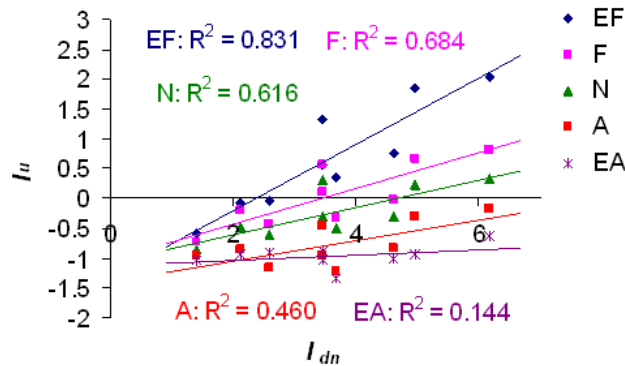
Fig. 2.8   The index of target utilization I_u changes with instruction and I_{dn}.

As we can see in Figure 2.8, even within the same operating condition and at the same I_{dn} level, I_u could still be very different. This leads us to examine the influence on

I_u separately by W and A (Figure 2.9, see also Table 2.7). W was shown to have much stronger influence on I_u than A, as indicated by the slopes and r^2 values. This gives rise to an explanation of the fact that I_{de} was a weaker determinant than I_{dn} of completion time MT within each condition. Since MT and I_{dn} form a very strong correlation within each condition, any adjustment of I_d will only weaken the strength of the correlation, unless all I_d's were changed with the same proportion. The I_{de} adjustment, however, depends on W_e (relative to W), which is influenced more by W than by A. This means the Fitts' law regression points at the same or similar I_{dn}, determined by $(A + W)/W$ ratio, may shift laterally to a different extent in I_{de} adjustment. Figure 2.10 shows the relative shifts in one experimental condition (N) of the regression points when I_{dn} (diamonds) is changed to I_{de} (squares).
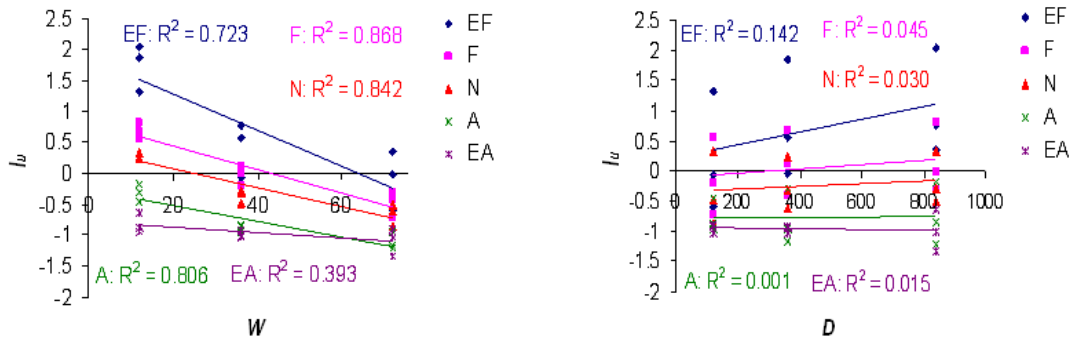
Fig. 2.9   The index of target utilization I_u as a function of W (left), A (right), and instruction condition MT vs. I_{dn} or I_{de} of Condition N.
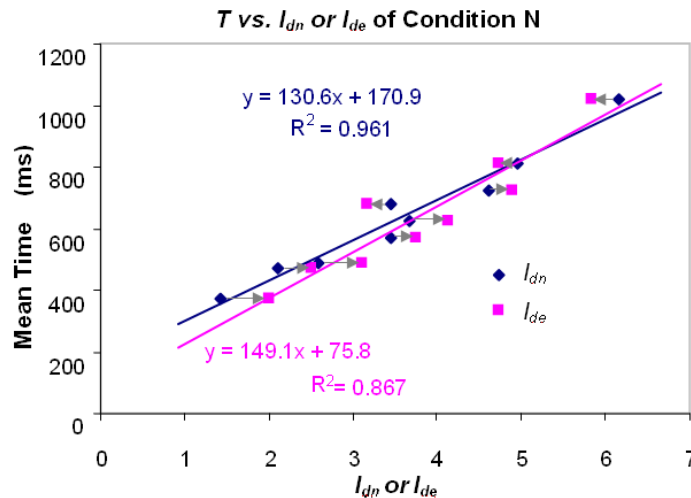
Fig. 2.10 MT vs. I_{dn} points (diamonds) shift different amount to MT vs. I_{de} points (squares), causing lower correlation.

Table 2.7 The standard deviation of endpoints at different A and W

Condition	A	W		
		12	36	72
EA	120	1.58	4.59	8.57
	360	1.52	4.27	9.28
	840	1.87	4.31	6.92
A	120	2.12	4.79	9.01
	360	2.36	4.50	7.77
	840	2.58	4.88	7.49
N	120	3.63	6.24	9.56
	360	3.41	7.02	11.40
	840	3.64	7.10	12.23
F	120	4.27	7.56	10.49
	360	4.61	9.41	13.07
	840	5.12	8.65	13.96
EF	120	7.27	8.33	11.59
	360	10.51	12.94	17.00
	840	11.94	14.75	22.34

Table 2.7 gives the standard deviation of the endpoints in Experiment SAT3 under different A and W values. When accuracy was emphasized, standard deviation was mostly decided by W. When accuracy was less emphasized, A began to exert impact on the standard deviation.

2.4.4 Preliminary Conclusions of I_{dn}, I_{de} and the Two Layers of Speed-Accuracy Tradeoffs

These three experiments systematically examined speed-accuracy tradeoffs in Fitts' pointing tasks. They showed that both the nominal task precision and performer's bias in over or under utilizing the given task precision tolerate change pointing completion time. The nominal index of difficulty I_{dn} is a remarkably robust predictor of completion time within each operating condition: the mean task completion time could be well accounted for by I_{dn} within each operating condition, even though the conditions were at very different overall levels of target utilization as quantified by I_u. The problem with the $T = f(I_{dn})$ model, however, is that the regression coefficients (a and b in Fitts' law) change with the overall level of I_u.

The experiments also showed that performers could rarely completely match the nominal task precision specification. I_u is not only affected by the performers overall bias towards speed (over utilization of target) or accuracy (under utilization), it also interacts with I_{dn}. Higher I_{dn} tended to cause over utilization. Even if the performers overall operated at the standard error rate (4%), as in Condition N in Experiment SAT3,

I_u still changed over 1 bit from low I_{dn} to high I_{dn} trials.

Overcoming some of the limitations of the $MT = f(I_{dn})$ model, the $MT = f(I_{de})$ model could compensate for some of the difference of I_u, particularly when the average I_u was around 0 or negative. As a result of using I_{de} the a and b estimates were much less affected by the operation conditions, and the r^2 value between MT and I_{de} was much higher when both I_{dn} and I_u varied widely (data mixed from all conditions) than the r^2 value between MT and I_{dn} (cf. Figure 2.6).

However, the correction effect of the I_{de} model is at the expense of a weakened MT vs. I_d relationship. Within each operating bias condition, MT vs. I_{de} consistently yielded lower r^2 than MT vs. I_{dn}. This was partly due to the fact that the amount of shift from I_{dn} to I_{de} was influenced more by W than by A, so I_{dn} to I_{de} shifts were not always the same. Given the strong regularity of MT vs. I_{dn} , any uneven change from I_{dn} would only result in a weaker regularity.

In sum, the behavioral model $MT = f(I_{de})$ offers a compromise. It could absorb some of the mismatch between the nominal task specification and the performer's actual pointing precision, at the expense of a strong MT vs. I_d regularity. The results of MT vs. I_{de} regression in terms of coefficients a and b were more stable than those of MT vs. I_{dn} across operating biases, but they still did not completely converge. It appears that W_e as a posteriori adjustment does not fully account for the time performance difference caused by the second and subjective layer of speed-accuracy tradeoff the performance's incompliance with the task specification (a none-zero or varying I_u) resulting in a overall faster or slower speed.

2.5   Experiment SAT4

To observe more directly the exact extent of I_u impact on time performance, we conducted yet another experiment that systematically controlled effective width W_e and nominal target width W to a similar amount in two experimental conditions. If W_e could compensate I_u variance, we would expect the MT vs. W_e relationship in the presence of varying I_u to be identical or similar to the MT vs. W relationship when the performers obediently complied with the target size specification W with no or little I_u variance.

2.5.1   Set-up and Experimental Design

Ten volunteers, eight males and two females (averaging 24.2 years old), participated in this experiment. Some of them had participated in Experiment SAT1 or Experiment SAT2. The experiment was conducted on the same apparatus as in Experiments SAT1

and SAT3 with a similar experimental procedure. It consisted of two parts (or schemes): part A (the target width incompliant scheme) and part B (the target width compliant scheme). In both parts, the participants performed reciprocal target tapping with a fixed distance A of 400 pixels. In part A ("target incompliant"), W was fixed at 20 pixels. Participants performed under the five sets of strategy instructions as in Experiment SAT3 (EA, A, N, F, EF). Under each instruction set, they performed 14 trials. There was a total of 700 trials collected (= 5 instructions x 14 trials x 10 participants). No accidental trials were observed. The goal of part A was to produce a set of time measurements under the same nominal target width, but very different effective target width W_e due to different levels of target width utilization. Based on the experience of Experiment SAT3, I_u was about -1, -0.5, 0, 0.5 and 1 bits in EA, A, N, F and EF bias conditions, respectively, which map 20 pixels (W) to five different W_e values to approximately 10, 14, 20, 28 and 40 pixels.

In part B (target compliant), W was set at 10, 14, 20, 28 and 40 pixels, corresponding to the expected W_e values in part A. The goal of Part B was to produce a set of time measurements when participants obediently complied with the given target widths to an (almost) ideal extent: the $\mid I_u \mid$ value should be less than 0.1 bit (i.e. W_e matches W within 7% margin). To achieve that, we used a target width enforcement method inspired by and refined from the verbal feedback method of Guiard and colleagues [54]. During the experiment (after the first 5 trials in each block), if the running I_u value was greater than 0.1 (i.e. $W_e > 1.072W$), which meant that the participant took too much risk, a sign appeared in the middle of the two target strips to remind the performer to slow down. In contrast, if I_u was less than -0.1 (i.e. $W_e < 0.933W$), a sign of different color appeared to remind the participant to speed up. If no sign was displayed, it meant the participants current endpoints dispersion corresponded to W within a 7% margin so the participant could keep his or her current pace.

The method of measuring the running I_u (W_e) value was as follows: Before the participant performed the 15th trial in a W condition, the program calculated the standard deviation of the endpoints distribution based on all of the past trials (from 1 to 14). From the 15th trial the program calculated the standard deviation of the endpoints, based on the most recent 14 trials (i.e. a 14-trial moving window was used). The experiment program stopped the current W condition and began the next one once a block of a 14 trials whose $\mid I_u \mid$ value was less than 0.1 was captured (i.e. their W_e matched W by a less than 7% margin). These 14 trials were used in later analysis. The program would have also aborted the current W condition if the participant had performed 30 trials without reaching a 14 trial block that met the requirement. In the actual experiment, none of the participants needed to use up the maximum 30 trials. We analyzed the endpoints of the last 14 trials and confirmed that they were normally

Table 2.8   Results of Experiment SAT4

	EA	A	N	F	EF
part A					
W	20	20	20	20	20
error %	0	1.4	4.2	17.8	25
W_e	9.54	14.98	18.20	30.58	37.69
Time	992.5	797.2	634.4	593.3	489.2
part B					
W	10	14	20	28	40
W_e	11.03	14.54	21.90	30.96	44.36
Time	860.8	744.8	639.2	611.8	526.8

distributed. The order of the W condition was randomized in our experiment.

Therefore, through part A, we could observe the relationship between mean trial completion time and the effective target widths under different levels of target utilization as a result of the participants' disrespect of the nominal target width to varying directions and extent. Through part B, we could observe the relationship between mean trial completion time and the nominal target width when the participants obediently and effectively complied with the accuracy tolerance specified by the target. Comparing the target width compliant scheme and the target width incompliant scheme we could directly examine whether and how well W_e reconciles the two layers of speed-accuracy trade-offs.

2.5.2   Results

Table 2.8 shows the results of Experiment SAT4. Figure 2.11 (left) shows the logarithmic regression results from the two experiment schemes (Part A and B). Figure 2.11 (right) show the Fitts' law regression of Part A and Part B, using I_{de} and I_{dn} respectively. The two sets of scatter plots (and regression lines and curves) show that the relationship between completion time and W_e in the target incompliant scheme (part A) and the relationship between completion time and W in the target compliant scheme (part B) were consistent in direction, but different in extent. In the near-zero range of I_u ($< |0.5|$ bits, or $W/2^{0.5} < W_e < W/2^{-0.5}$), the difference was relatively small. When I_u was beyond such a range, this difference increased rapidly. The greater $| I_u |$ was, the greater such a difference was.

The results of this experiment shed more light on the effect that I_{de} only partially compensates for the time variance caused by different operating biases (different target utilization levels): while the MT vs. W_e relationship in the presence of I_u variance was similar to MT vs. W relationship when the performers obediently complied with the
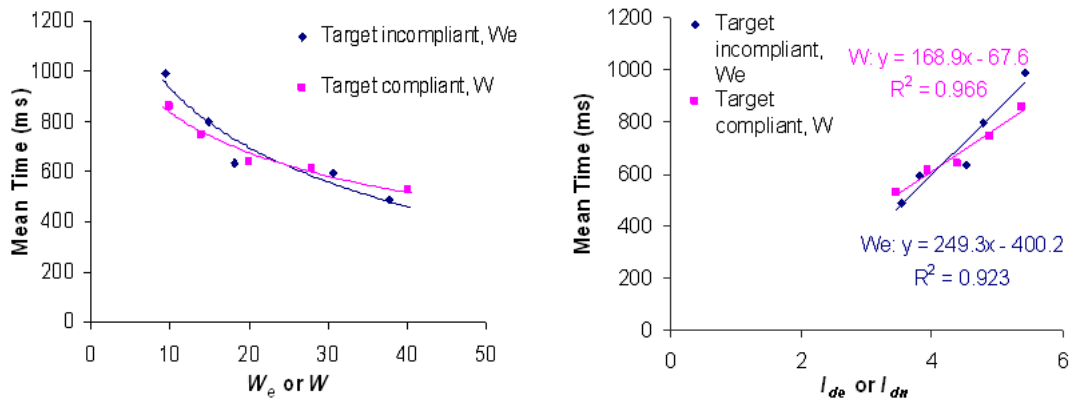
Fig. 2.11   The match between MT vs. W_e in the target incompliant scheme and MT vs. W in the target compliant scheme (left) and their corresponding Fitts' law regressions (right).

target width specification (with no or little I_u variance), they did not exactly match in extent. The impact of the W_e adjustment lagged behind the impact of W changes in the same amount.

2.6   Further Analyses

The foregoing analyses clearly indicate that the use of W_e is an imperfect and insufficient adjustment of W to overcome I_u variances in Fitts' law tasks, although the direction of adjustment was empirically correct. Is it possible to find an adjustment that more fully compensates the impact of I_u variance? Given that W_e under-corrects when the endpoints dispersion deviated significantly from the nominal width W (shown in Figure 12), a more exaggerated effective width could possibly account for more of the remaining difference in completion time. This suggests the following modified effective width is worth investigating:

$$W_m = W(\frac{W_e}{W})^\alpha \tag{2.3}$$

i.e.

$$W_m = W(\frac{4.133\sigma}{W})^\alpha \tag{2.4}$$

Where σ is the standard deviation of the endpoints distribution. α should be greater than 1 in order to cause W_m to nonlinearly exaggerate the impact of deviation from W: When W_e is close to W, W_m has a similar value to W_e; When $W_e \gg W$ or $W_e \ll W$, W_m is much greater or much smaller than W_e. The greater the α value is, the

more pronounced difference there is between W_m and W_e. When α is 1, W_m reduces to W_e. Accordingly,

$$I_{dm} = log_2(\frac{D + W_m}{W_m}) \tag{2.5}$$

and

$$MT = a + bI_{dm} \tag{2.6}$$

Based on the empirical data from Experiment SAT4, 1.5 is a good estimate of α. Figure 2.12 (left) shows the mean trial completion time as a function of W_m with $\alpha = 1.5$ (the target incompliant condition of Experiment SAT4). The relationship between time and W_m matches almost exactly the relationship between time and nominal W when participants closely obeyed the target width (the target compliant condition of Experiment SAT4). Figure 2.12 (right) shows the corresponding Fitts' law regressions using I_{dn} and I_{dm}.
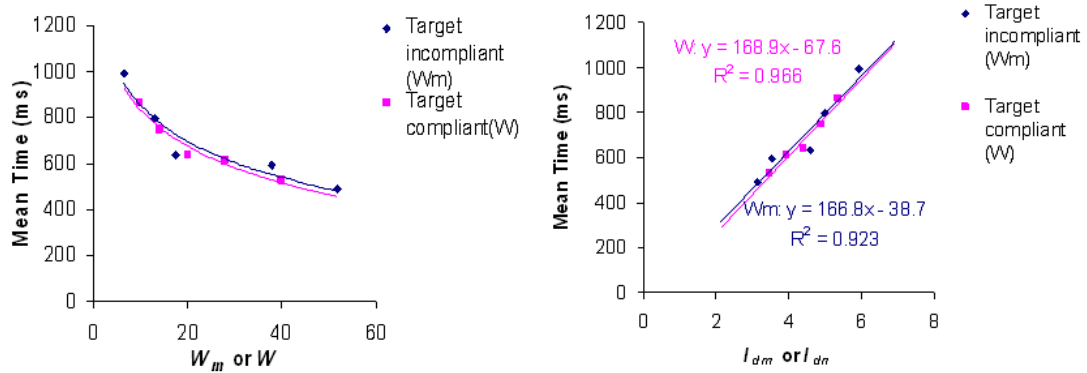
Fig. 2.12   The match between MT vs. W_m in the target incompliant scheme and MT vs. W in the target compliant scheme (left) and their corresponding Fitts' law regressions (right).

Of course, the value of α based on one small experiment (Experiment SAT4) is not likely to be an accurate estimate for other data sets, although it is plausible that a better compensation may be achieved by W_m with certain α value. A potential difficulty that W_m faces, however, is the limitations of W_e adjustment that we see in the first three experiments: while the W_e adjustment reduced the discrepancy of Fitts' law coefficients measured under different operating biases, it also weakened the strong regularity within each operating condition as modeled by $MT = f(I_{dn})$. Since W_m is a more exaggerated version of W_e, the correlation of MT vs. I_{dm} may be reduced further from that of MT vs. I_{dn} within each operating condition. Another obvious weakness of the notion of W_m is that since it is based on W, W_e and another parameter α, its direct definition is

lacking or to be discovered in the future.

To test the possibly stronger compensation power of W_m and its likely drawbacks, we reanalyzed the first three experiments, substituting I_{de} with I_{dm} (with $\alpha = 1.5$). The results are summarized in Figures 2.13 to 2.15 and Table 2.9.
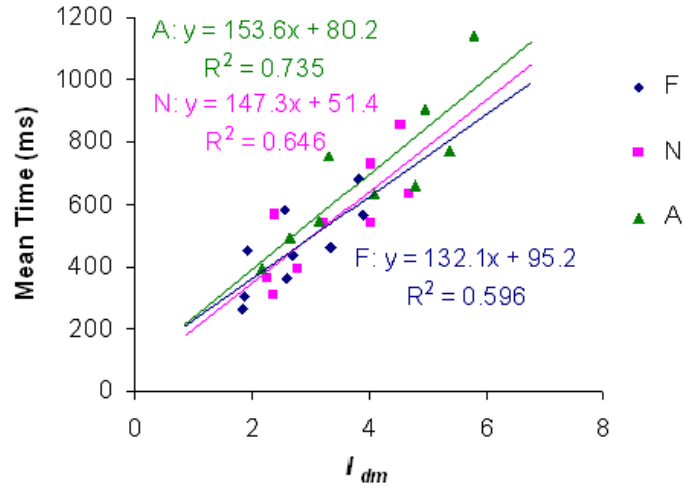
Fig. 2.13 MT vs. I_{dm} regression of Experiment SAT1 (Compare with Figures 2.2, 2.3).
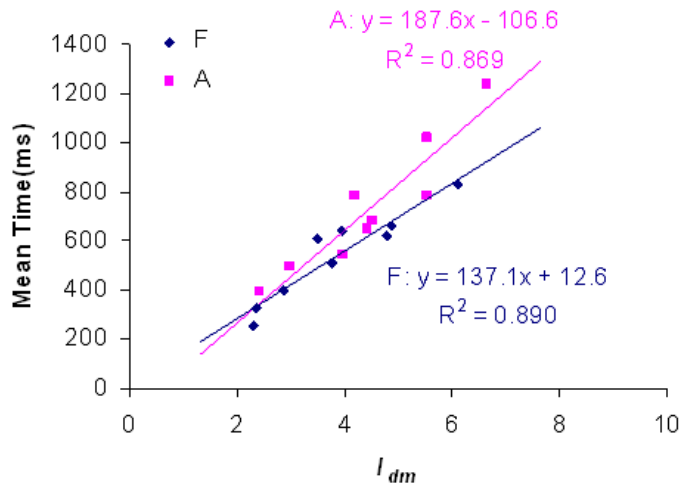
Fig. 2.14 MT vs. I_{dm} regression of Experiment SAT2 (Compare with Figure 2.5).

Based on the degree of convergence of the regression coefficients (see Column 9 of Table 2.9), I_{dm} compensates for the different levels of I_u slightly better (Experiment SAT1, Experiment SAT2) or better (Experiment SAT3) than I_{de}. Taking the b value in Experiment SAT3 (the most comprehensive and balanced experiment) as an example, under the EA, A, N, F, EF operating conditions b was 215, 186, 131, 113, 83 ms/bit
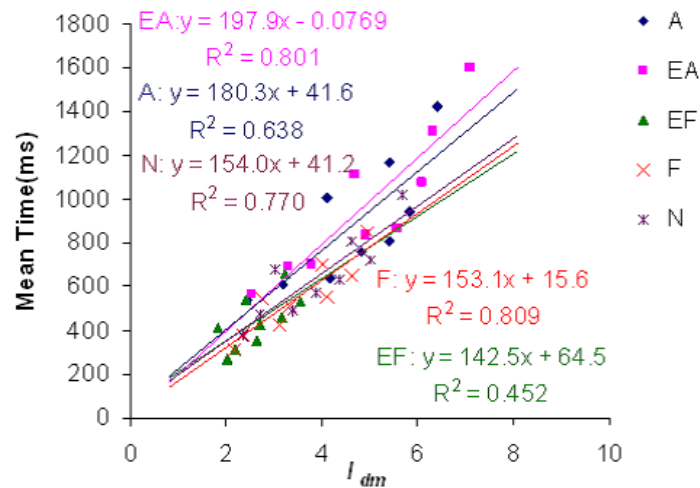
Fig. 2.15 MT vs. I_{dm} regression of Experiment SAT3 (Compare with Figure 2.6).

with I_{dn}; and 204, 187, 149, 143, 147 ms/bit with I_{de} . With I_{dm}, b was 198, 180, 154, 153, 143 ms/bit. If we use the estimates under Condition N as the best possible estimate, the b values can be written as percentage changes from its best estimate, as shown in Table 2.10. I_{dm} appears to offer better correction of operating biases than I_{de} (See also Figure 2.15 in comparison to Figure 2.6).

However, the limitations of I_{dm} were also apparent. First, it still could not completely compensate for the speed accuracy tradeoff caused by different levels of target utilization. The coefficients measured under different operating biases still did not fully converge. Second, since W_m is a nonlinear amplification (or reduction) of W_e , it weakened the regularity found in MT vs. I_d even more than W_e, as indicated by the further decreased r^2 values within each condition (Table 2.9). When mixing all biased conditions, the r^2 of MT vs. I_{dm} was similar to the r^2 of MT vs. I_{de}: weaker in Experiment SAT1, but stronger in Experiments SAT2 and SAT3.

One could argue that stronger results with I_{dm} could be achieved with different α values. As Table 2.11 shows, this was indeed true, but there was not a α value that is optimal for all experiments with different ranges and sets of operating biases.

Table 2.9 Comparison of I_{dn}, I_{de} and I_{dm} in Experiments SAT1, SAT2 and SAT3

1	2 I_d	3 EA a, b, r^2	4 A a, b, r^2	5 N a, b, r^2	6 F a, b, r^2	7 EF a, b, r^2	8 Mixed a, b, r^2	9 $\Delta b/b_{min}$
Expt SAT1	I_{dn}		161, 150, 0.95	123, 118, 0.98	123, 92.4, 0.99		135, 120, 0.70	49%
	I_{de}		92.7, 156, 0.83	41.5, 147, 0.84	56.1, 133, 0.87		13.2, 161, 0.83	16%
	I_{dm}		80.2, 154, 0.73	51.4, 147, 0.65	95.2, 132, 0.60		38.3, 157, 0.75	15%
Expt SAT2	I_{dn}		111, 173, 0.93		111, 119, 0.92		111, 146, 0.75	45%
	I_{de}		-28.1, 182, 0.90		48, 131, 0.91		76.2, 141, 0.60	39%
	I_{dm}		-107, 188, 0.87		12.6, 137, 0.89		-80.9, 172, 0.85	37%
Expt SAT3	I_{dn}	201, 215, 0.90	210, 186, 0.90	171, 131, 0.96	141, 113, 0.996	138, 83.8, 0.99	172, 146, 0.46	100%
	I_{de}	63.6, 204, 0.84	76.0, 187, 0.75	75.8, 149, 0.87	45, 143, 0.93	7.5, 147, 0.88	-92.4, 208, 0.78	41%
	I_{dm}	-0.08, 198, 0.80	41.6, 180, 0.64	41.2, 154, 0.77	15.6, 153, 0.81	64.5, 143, 0.45	-82.1, 199, 0.81	36%

Table 2.10 b estimate as percentage variation from the neutral condition (N) in Experiment SAT3

	EA (%)	A (%)	N (base)	F (%)	EF(%)
I_{dn}	64.1	42	131	-13.7	-36
I_{de}	36.9	25.5	149	-4	-1.3
I_{dm}	28.6	16.9	154	-0.6	-7.1

Table 2.11   Regression results with mixed conditions at different α values

	$\alpha = 1.2$			$\alpha = 1.3$			$\alpha = 1.5$			$\alpha = 1.8$		
	a	b	r^2	a	b	r^2	a	b	r^2	a	b	r^2
EXP. SAT1	13.9	162.1	0.808	19.0	161.3	0.792	31.2	157.9	0.745	86.2	144.1	0.656
EXP. SAT2	-44.6	168.4	0.835	-57.1	169.8	0.839	-80.9	172.4	0.845	-113.1	175.3	0.849
EXP. SAT3	-100.7	207.3	0.806	-98.2	205.4	0.811	-82.1	198.8	0. 807	-37.4	184	0.778

Mathematically, the α term in Equations 2.3 to 2.6 can be separated from I_{dn}. Considering that A is typically greater or much greater than W, and when W_e/W is not too distant from 1, the following equation approximates Equation 2.6:

$$MT = a + b\log_2(\frac{A+W}{W}) - b\alpha\log_2(\frac{W_e}{W}) \tag{2.7}$$

or

$$MT = a + bI_d - b_u \tag{2.8}$$

or

$$MT = a + bI_d + cI_u \tag{2.9}$$

where $c = -b\alpha$.

Multiple regression results (a, b, and c) based on the first three experiments, however, tended to be highly dependent on which experimental conditions were included in the regression, suggesting the complex interactive nature of the I_{dn} and I_u effect.

Figure 2.16 shows the mean of the completion time MT as a function of I_{dn} and I_u based on data from Experiment SAT3 (all conditions mixed). As we can see, while MT increased with I_{dn} and decreased with I_u, in general, they did not form a strictly monotonic function, suggesting complex interaction effects between I_{dn} and I_u. This means that it is difficult, if not impossible, to establish a model that captures both layers of speed-accuracy tradeoffs in a complete and yet simple (linear) manner.

2.7   General Discussions and Conclusions

We have systematically explored the two layers of speed-accuracy trade-offs in Fitts' aimed movement tasks. Fitts' law in its original form reveals the speed-accuracy tradeoff
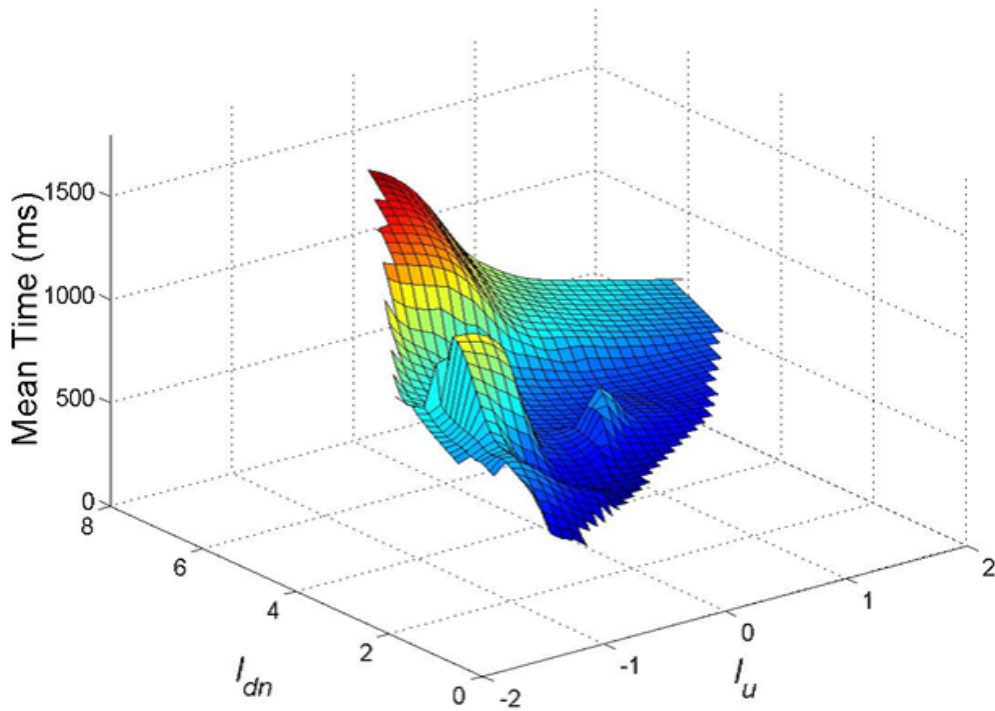
Fig. 2.16   3D perspective illustration of $MT = f(I_{dn}, I_u)$ based on data from all conditions in Experiment SAT3.

relationship between pointing completion time and task precision based on the nominal, objective, geometrical parameters of the target (A and W). However, in actuality speed-accuracy relationship in pointing also contains another subjective layer which depends on how obediently the performer complies with the specified target width and what bias the performer takes (toward either accuracy or speed). This performer introduced accuracy layer causes a discrepancy between the nominal task precision and the actual behavior precision. We defined an index of target utilization, $I_u = \log_2(4.133\sigma/W)$, to quantify the degree of this mismatch. As is implicitly realized in the Fitts' law literature, and as this study has explicitly and systematically shown, Fitts' law tasks tend to involve both layers of speed-accuracy tradeoffs. Our study shows I_u is never constant in an experiment, even within the same instruction set, such as "as accurately as possible and as fast as possible", except when an enforcement method is applied, as in Experiment SAT4. The overall I_u level can be influenced by the experimental instruction in the laboratory, or by performers' preference and task strategy in real world tasks.

This study clearly demonstrated that varied I_u values influence Fitts' law regression modeling, resulting in different a, b coefficients which in the context of human-computer interactions are often used to characterize an input system's efficiency. The classic

approach to correct the influence of varying I_u is the so-called "effective target width" method. This approach takes the performer's actual behavioral parameter, W_e, rather than the nominal target width W, as the basis of index of difficulty calculation. Such an approach has not had a strong theoretical or empirical foundation in the literature.

In a set of four experiments, we deliberately manipulated I_u over a wide range or controlled to specific levels through instructions and feedback control, which enabled us to investigate the two layers of speed-accuracy tradeoffs systematically. Our investigation has led to the following conclusions.

First, the task form of Fitts' law $MT = f(I_{dn})$ is a very strong model. The nominal I_{dn} is an impressive determinant of mean movement time, accounting for up to 99% time variance within an experimental condition. Revisions of index difficulty to a behavior form, either through W_e, or its more aggressive version W_m, consistently weaken the regularity within an particular operating bias condition. On the other hand the resulting coefficients of MT vs. I_{dn} can be easily swung by I_u levels. In the context of HCI, this poses a serious challenge to assessing the quality of various input systems [55].

Second, I_{de} partially incorporates the second, subjective accuracy layer into Fitts' law model by adopting an actual and behavior parameter W_e. The first three experiments showed that adopting W_e reduced the discrepancy of a and b estimates between different experimental conditions. Experiment SAT4 shows that, although not completely congruent, the impact of the a posteriori effective target width was similar to that of a priori nominal target width with which the performer obediently complied. The compensation effect of W_e was also shown by the higher r^2 value of MT vs. I_{de} regression across different operating biases (mixed data from all conditions) than the r^2 value of MT vs. I_{dn} regression. This study, to our knowledge, provided the first systematic empirical foundation for the use of W_e. However, the compensation effect of W_e is gained at the cost of weakened regularity within each experimental condition.

Third, I_{dm}, the more aggressive version of I_{de}, takes a step further than I_{de}: it more fully compensates for the I_u impact but further weakens regularity within each operating strategy condition.

Fourth, in the absence of the subjective layer of speed-accuracy tradeoff (the performer completely complies with task specification, keeping I_u at or near zero), both I_{de} and I_{dm} revert to the nominal I_{dn}. In that sense, no harm can be done by adopting I_{de} and I_{dm}.

Fifth, the level of target utilization in Fitts' law tasks, as measured by the index of target utilization $I_u = \log_2(\frac{W_e}{W})$, can be influenced by three factors: the overall operating bias (in the lab by instruction, in reality by user's strategic choice); nominal target W; and target distance A. However, W and A have very different degrees of impact on I_u, with the former being far greater than the latter. This means that the

change from I_{dn} points (determined by A and W) to I_{de} points (determined by A and W_e) is not uniform. Given that MT and I_{dn} forms a very strong correlation within each strategy, the use of I_{de} or I_{dm} would only weaken this correlation. A compromise has to be made between the goodness of fit within each condition and better estimates of a and b. Fundamentally, the use of I_{de} or I_{dm} is to incorporate a different layer of speed-accurate tradeoff (the subjective layer) into a very strong time and task accuracy relationship as modeled by $MT = f(I_{dn})$.

Sixth and finally, the two layers of speed accuracy tradeoff in pointing interact in a complex manner. They do not cause a simple additive effect, hence the difficulty of using $MT = f(I_{de})$ or $MT = f(I_{dm})$ and any other potential relationship as a strong general model.

In summary, this systematic investigation reveals the nature and the relationship of speed and accuracy in pointing. The implication of this study depends on the specific purpose of use. Theoretically we now know that the two layers of speed-accuracy tradeoffs, the objective and task layer and the subjective and behavior layer have different impact on task performance. The impact of actual pointing precision W_e and the impact of nominal pointing precision W are not equivalent, but are numerically similar, particularly when W_e is not too distant from W. The findings in this work also have more general implications to human motor control theory. This study shows that even for a low level tapping task, both external visual feedback and internal bias settings contribute to the control process. In a manual stabilization task, Pew [41] showed that humans not only react instantaneously to the position of a controlled object, but can also adjust higher level control parameters to achieve stabilization. This experiment shows that in pointing tasks performers could adjust their overall bias towards speed or accuracy and integrate such an internal high level setting with the external low level visual feedback to manage a pointing process.

Practically, the findings in this study suggest that in order to accurately measure Fitts' law parameters, I_u should be kept as close to zero as possible and its variance should be kept as low as possible. One possible method of controlling I_u in Fitts' law studies is to use endpoints standard deviation-based feedback, as we did in Experiment SAT4. When I_u is highly varied or when it is not near zero, however, this study provides an empirical foundation for the application of W_e or its more aggressive and more complete version, W_m, to adjust for I_u changes. These adjustments consistently yield more logical Fitts' law parameter estimates, a and b, although one should also be aware of the limitations and side effects of W_e or W_m, including reduced correlation between pointing time and index of difficulty within each operating strategy and their incomplete compensation for the subjective layer of speed-accuracy tradeoff.

This is a blank page.

Chapter 3

The Optimal Calculation Method of the Effective Target Width of Fitts' Law

The use of effective target width (W_e) in Fitts' law has been a widespread applied method for one directional pointing task evaluation, especially when different speed or accuracy inclinations incurred by subjective factors exist. However, the concrete methods for calculating W_e have not been formally integrated. Although the ID_e model has been discussed a lot in Chapter 2, the optimal calculation method of W_e is not studied. Therefore, this chapter focuses on resolving this problem. A Testing Experiment and a specially designed and controlled Comparison Experiment are described in this chapter. The experiments and the data analysis results show that after comparing the two existing methods for calculating effective target width, the method of mapping all the abscissa data in one integrated coordinate system to perform the calculation is proven to be better for human computer interface modeling than dividing the data into two groups and mapping them in two separate coordinate systems.

3.1   Background Knowledge and Related Works

Although Fitts' law has been used widely in HCI and advocated by many researchers [34], it is still under suspicion[56]. One problem is that the calculation of the W_e has not been integrated.

In Equation 1.5, $W_e = 4.133SD$. SD is the standard deviation of the hits distribution. For SD calculation, some researchers use one united coordinate system to calculate the average of the x-coordinates to get SD, as in Douglas, Kirkpatrick and MacKenzies research[15]. We call this method the Combined-coordinate-system Method (the CC method) in this chapter. Some others use two sets of coordinate systems to calculate the average of the x-coordinates to get SD, as Isokoski and Raisamo have done in their study[24]. In this method the averages of the x-coordinates need to be calculated for the left and right coordinate systems respectively. We call this method

the Separate-coordinate-system Method (the SC Method) in this chapter.

However, at present, no research has been reported on the preferred method of W_e calculation for the application of the Fitts' law model. Moreover, no comparison has been reported in the ISO standards 9241-9[25]. Therefore, in this study we compare the two methods to see which one is better for calculating W_e. The results derived from this work will be of great help for the further application of Fitts' law to the HCI field.

3.2   Testing Experiment: Testing the Hits' Distribution

The SC method is much more complex than the CC method, but some researchers still support the SC Method because they hold to the hypothesis that with bigger targets the users tend to click near the nearest edge of the rectangular target rather than near the middle of it. They therefore go on to argue that if the SD is calculated in relation to a united center, the off-center click distribution will inflate the standard deviation and bring inaccurate results of W_e [24].

To observe whether the distribution of the input hits is as Isokoski and Raisamo assumed, we developed a pointing task experiment with different A and W (target width) combinations.

3.2.1   Subjects

Ten volunteers, five male and five female (average 28.8 years old), participated in this experiment.

3.2.2   Apparatus

We used a desktop PC with a color LCD monitor, the EIZO FlexScan L567 (screen size 338 mm (H) x 270 mm (V)) in this experiment. The Resolution was 1024 x 768. 1 pixel was 0.264 mm. The input device was the Microsoft Wheel Mouse Optical 1.1A.

3.2.3   Procedure

The experimental procedure was designed according to the ISO 9241-9 standard [25]. During the experiment, participants did reciprocal pointing with a mouse at a pair of vertical strip targets displayed on the screen. The width (W) of the targets and the center-to-center distances (A) between the two strips were set at $W = 12, 36, 72$ pixels and $A =120, 360, 840$ pixels. The order of the nine width and distance combinations was randomized. The start position of the cursor was the center of the screen. Twelve

Table 3.1   The SD and ID_e with the CC method and the SC method in the Testing Experiment.

Methods	Combinations(in pixels)	SD(in pixels)	ID_e(in pixels)
the CC	A=120, W=12	3.44	3.24
Method	A=120, W=36	8.77	2.11
	A=120, W=72	12.11	1.76
	A=360, W=12	3.29	4.78
	A=360, W=36	9.73	3.31
	A=360, W=72	14.57	2.80
	A=840, W=12	3.06	6.07
	A=840, W=36	8.66	4.61
	A=840, W=72	15.32	3.83
the SC	A=120, W=12	3.39	3.26
Method	A=120, W=36	7.75	2.25
	A=120, W=72	10.07	1.96
	A=360, W=12	3.27	4.79
	A=360, W=36	9.55	3.34
	A=360, W=72	13.26	2.92
	A=840, W=12	3.05	6.08
	A=840, W=36	8.42	4.65
	A=840, W=72	15.00	3.86

trials were presented in each W-A combination, with the first tap excluded in analysis. If the user tapped on the outside of the target, the task would not be abandoned and an auditory signal would be played.

The subjects were required to perform the tapping task as fast and accurately as possible, as described in Fitts' paradigm experiment[18]. During the task, except for the sound indicating a mistake had occurred, there was no other feedback to affect the subjects performance[*1].

3.2.4   Results

Table 3.1 shows the SD, ID_e, and the corresponding amplitude and target width combinations in the Testing Experiment.

Figures 3.1 and 3.2 show the input hits distribution of the pointing task in the Testing Experiment. The abscissa values indicate the horizontal distribution range of the hits (E.g., the target width in Figure 3.1 is from -36 to 36, and the target center's position is 0). The ordinate values indicate the distribution density of the hits in corresponding horizontal positions (E.g., in Figure 3.1, 2 hits fall into the area from 24

[*1] In the Comparison Experiment, with each of the subjects taps there was an instant feedback signal appearing in the screen to remind the subjects to slow down or hurry up.
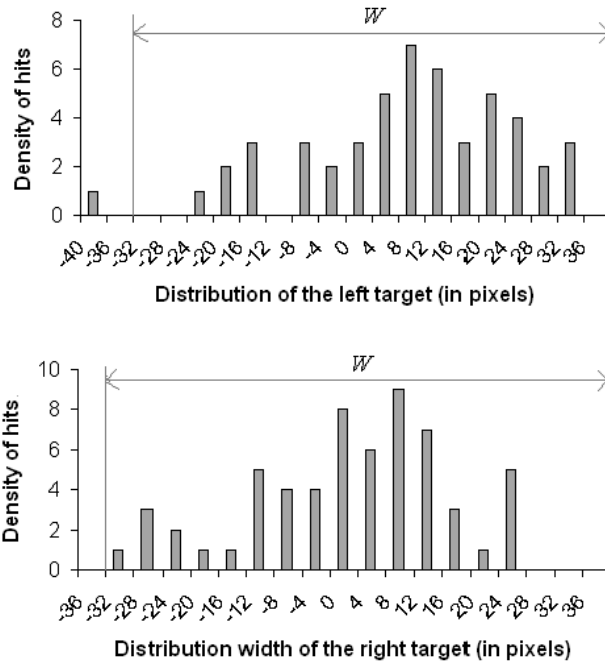
Fig. 3.1   The distribution of the input hits (A=840 pixels W=72 pixels)

to 28, therefore, the value of the ordinate is 2).

3.2.5   Discussion

Table 3.1 shows that the values of SD when using the SC method are less than when using the CC method, which in turn increases the values of ID_e. However the changing amount of SD is uneven. For big target sizes, the SC method decreases the standard deviation more; for small target sizes, the SC method decreases the standard deviation slightly or does not decrease the standard deviation significantly.

We compared the effects of different target sizes in Figures 3.1 and 3.2. The off-center tendency described by Isokoski and Raisamo is not clearly demonstrated with either bigger sizes (W=72 pixels) or smaller sizes (W=36 pixels): for the left target, the distribution of the dots did not lean obviously to the right of the center, meanwhile, the distribution of the dots around the right target did not lean obviously to the left off the center.
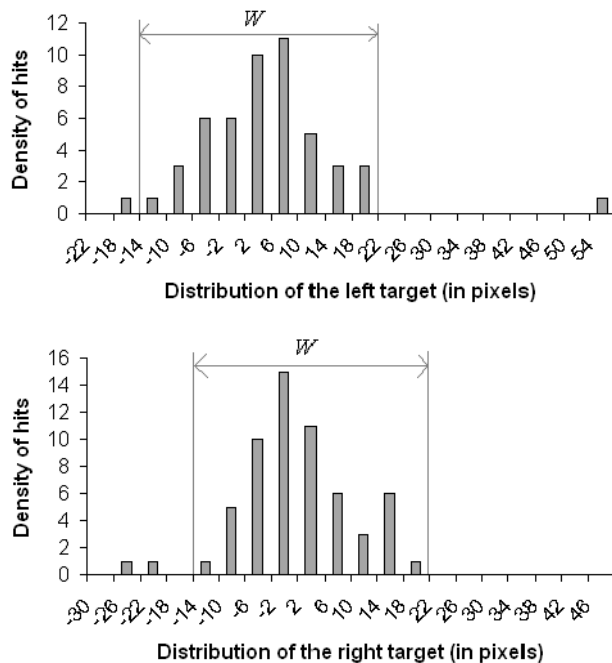
Fig. 3.2   The distribution of the input hits (A=840 pixels W=36 pixels)

3.3   Comparison Experiment: Comparing the CC Method and the SC Method

Although the Testing Experiment has shown that Isokoski and Raisamo's assumption was not supported, a clear comparison between the CC method and SC method could not be given only through the Testing Experiment. Therefore, we were intrigued to develop another experiment to concretely check which method of W_e calculation is better.

To analyze and compare the two methods of W_e calculation accurately, we developed an experiment that could produce a set of time measurements when participants kept their tapping within the given target widths to an almost ideal extent.

Since the results would be obtained from the ideal experimental situation, we expected to see a more precisely defined difference between the two methods.

3.3.1   Subjects

The same subjects in the Testing Experiment participated in the Comparison Experiment.

3.3.2   Apparatus

The same apparatus in the Testing Experiment was applied in the Comparison Experiment, but the program was different because it was designed for different experimental purposes.

3.3.3   Design

The study was partially intrigued by the study of the speed and accuracy tradeoff introduced in the last chapter, so the experiment design was according to Experiment SAT4 (see 2.5). In the Comparison Experiment, participants reciprocally pointed with a mouse on a pair of vertical strips which were at a fixed distance apart A of 400 pixels. W (appointed target width) was set at 10, 14, 20, 28 and 40 pixels. If the outside region of the target was tapped, the task would not be abandoned and an auditory signal would be played as a warning signal. The start position of the cursor for both parts was the center of the screen.

Through this experiment, by observing the ideal input hits distribution, we can see whether either of the methods is superior in modeling a pointing task.

3.3.4   Procedure

We applied the following procedures for the CC method and SC method to calculate SD and control the program for the CC method and SC method.

For the CC method, the program calculated the SD based on a one coordinate system (see Figure 3.3(b)). It meant that the standard deviation (SD) could be calculated by:

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}} \tag{3.1}$$

In Equation 3.1, x_i was the ith of the participant's selection point's x-coordinates (They were mapped into one united coordinate system). $\bar{x}$ was the mean of x-coordinates. n was the number of the trials.

For the SC method, the situation was more complex. The program calculated SD based on two sets of coordinate systems (see Figure 3.3(c)). The concrete steps were as follows: first, to compute the averages of the left and right x-coordinates of the previous 14 trials (or less than this number before the 15th trial), secondly, to get the x_i-$x_{average}$, $(i = 1, 2 \cdots n, n \leq 14)$, here x_i was the ith hit's x-coordinate, and $x_{average}$ was the average of the values of x_i, then there should be 14 numbers of x_i-$x_{average}$. (One point noticeable here was that for the left side hits and right side hits, the values

3.3 Comparison Experiment: Comparing the CC Method and the SC Method
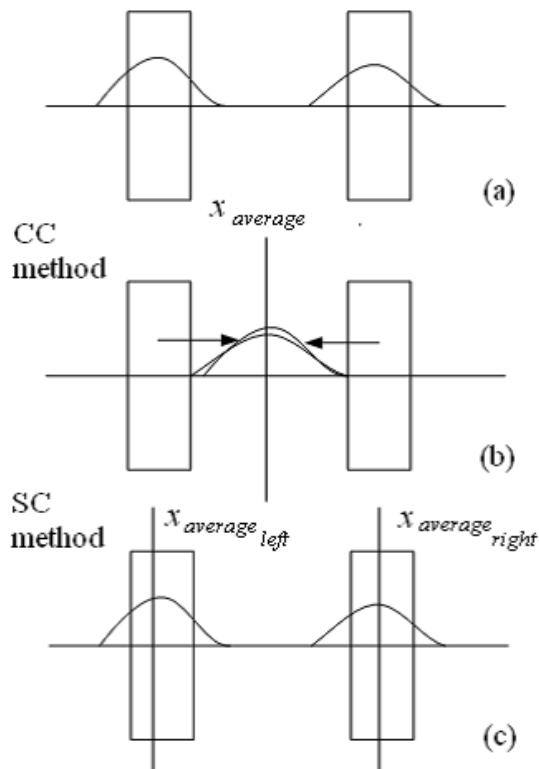
Fig. 3.3 The description figure of the two methods of effective target width calculation. ((a) indicates the hits distribution of the left and right targets; (b) indicates that the CC method was used to calculate the average, SD and W_e; (c) indicates that the SC method was used to calculate the average, SD and W_e.)

of $x_{average}$ were different[*2], here we used x_i-$x_{average}$ only for the convenience of the following narration. The next step was to get the SD of the 14 (x_i-$x_{average}$)s, if $x'_i=x_i$-$x_{average}$, then

$$SD' = \sqrt{\frac{\sum_{i=1}^{n}(x'_i - \bar{x}')^2}{n - 1}} \tag{3.2}$$

For both the CC method and the SC method, the procedure of measuring the running W_e value was as described in Section 2.5.

With either W_e calculation method, the total amount of data for analysis was 700

[*2] for the left side hits, x_i-$x_{average}$ should be written as $x_{i_{left}}$-$x_{average_{left}}$, and for the right side hits, x_i-$x_{average}$ should be written as $x_{i_{right}}$-$x_{average_{right}}$.
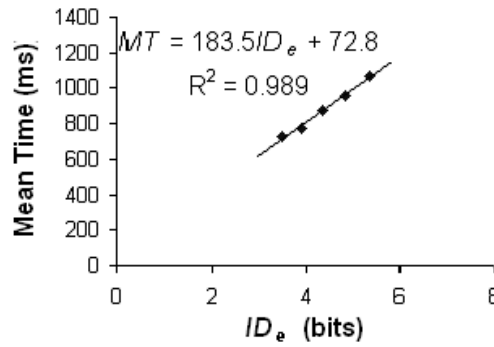
Fig. 3.4   The regression between mean time and ID_e using the CC method
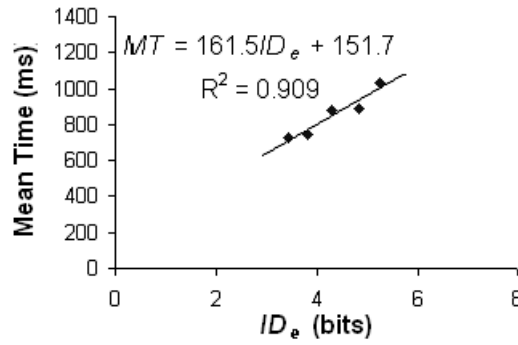
Fig. 3.5   The regression between mean time and ID_e using the SC method

(14 (trials) x 10 (subjects) x 5 (combinations of A and W) = 700).

3.3.5   Results

After the experiment, we collected data and drew the Fitts' law regression lines in Figures 3.4 and 3.5.

In Fitts' law, the relationship between movement time and target width is a logarithm relationship (Equations 1.2 and 1.4). Therefore, a logarithm relation curve between movement time and W_e will be more helpful to compare the effect of the two calculation methods. Therefore, we also made the logarithmic regression lines between the MT and W_e based on the data of the experiment (Fig. 3.6 and Fig 3.7).

3.3.6   Discussion

In Figure 3.4, R^2 of the regression line of the CC method is near to 1 (0.989), which means (that) by using the CC method the regression of Fitts' law is ideal and strong.

3.3 Comparison Experiment: Comparing the CC Method and the SC Method
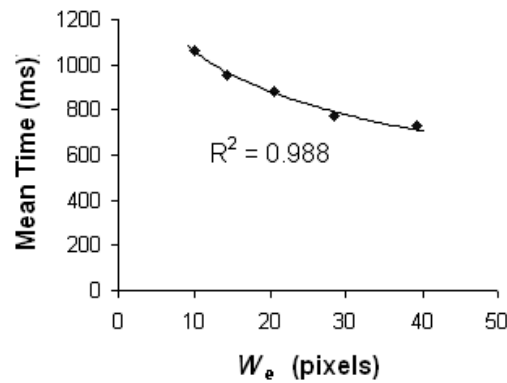
Fig. 3.6   The match between mean time and W_e by using the CC method
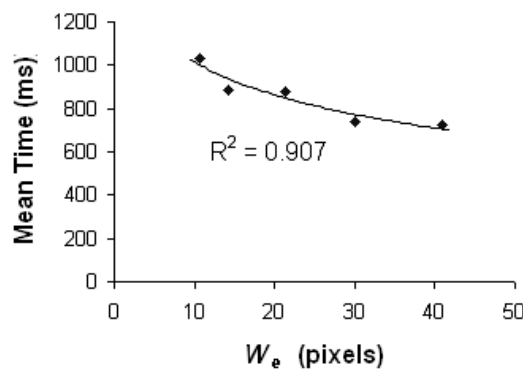
Fig. 3.7   The match between mean time and W_e by using the SC method

The regression of Fitts' law line in Figure 3.5 is still big (0.909), but not as great as indicated by Figure 3.4. This means that the SC method is not as precise as the CC method.

Figure 3.6 shows that with the CC method, the logarithm relationship between movement time and effective target width is obvious and all five dots are restricted to the curve ($R^2 = 0.988$). However, in Figure 3.7, the dots are scattered around the logarithm curve and are not confined tightly to the curve ($R^2 = 0.907$).

Since in the Comparison Experiment, the system gave an immediate response to the subject for each trial, the performance was under almost ideal control, therefore, the regression between MT and ID_e and the regression between MT and W_e was expected to be rather strong. From this point of view, the regression of the Fitts' law line in Figure 3.5 and the logarithmic regression in Figure 3.7 (related to the SC method) are not strong enough.

3.4   General Discussions and Conclusions

The data from the uncontrolled Testing Experiment can help us to investigate the reason for the inadequacies in the SC method.

As explained previously, the values of W_e calculated by the SC method decrease from those values calculated by the CC method, and the changing amount for different combinations of A and W are different (see Table 3.1). For big target sizes, the SC method decreases the standard deviation and W_e more; for small target sizes, the SC method decreases the standard deviation and W_e slightly. This irregular variation of SD or W_e obtained from using the SC method will result in a weaker regression between the mean time and ID_e than the regression obtained by using the CC method. These results show that the use of the SC method produces irregular effects on different target sizes.

In the specially controlled Comparison Experiment, for the SC method, we used the two sets of coordinate systems to calculate the standard deviation, which means the requirements placed on the individual subject were less rigid than if we had used a one coordinate system. Nevertheless, when we analyze the data, we must mix all the subjects data together, and the standard deviation for all the dots will then be inflated. That is the reason why the effective target width obtained from the SC method is bigger than expected.

Based on the above analysis, it is logical to conclude that using one coordinate system to calculate the effective target width is more reliable.

Moreover, the CC calculation method is also much easier and more convenient than the SC method.

Another point worthy of note is that all the subjects in the two experiments were right-handed. Since for the left-handed person, the situation can simply be reversed, we can assume that the preferred hand will not affect the analytical results of this study.

In conclusion, we studied and compared two methods for calculating W_e. The results show that the CC method (Combined-coordinate-system Method) is better than the SC method (Separate-coordinate-system Method), i.e., it is better to map all the abscissa data into one integrated coordinate system to do the calculation, rather than to divide the data into two separate groups according to the corresponding target positions.

We believe that the data shown by this study affords a detailed and reliable comparison of the two methods of W_e calculation based on the information derived from the input hits with different target sizes. The Combined-coordinate-system method recommended in this study will help researchers and developers determine more confidently and precisely the optimum effective target widths calculation method for pointing tasks.

Chapter 4

The Information Processing Rate Analysis and Its Application in Fitts' Law Models Comparison

Fitts' law was established on the assumption that the information processing procedure of human machine systems is analogous to wireless information systems. With a normal distribution of input amplitudes, the information processing rate should reach its maximum, i.e. it should reflect the channel capacity of the information transmission system of the pointing task. The purposely manipulated experiments carried out for Chapter 2 afford us the opportunity to observe the performance under various situations of varying degrees of accuracy and speed, and to testify this assumption and compare models based on this assumption consequently. In this chapter, the input hits distributions in different performance conditions were analyzed in detail. Then through applying the different forms of Fitts' law formulations, we described the varying tendencies of $1/b$ (the reciprocal of the task difficulty coefficient of the regression formulation of Fitts' law) in different performance conditions. Thereafter, two Fitts' law formulations were compared and analyzed on the basis of their ability to describe the information processing rate during tasks with different performance conditions.

4.1   Background Knowledge and Related Works

Although researchers have made a great deal of effort to support Fitts' law from either a theoretical perspective or via its application, the theoretical support for Fitts' law remains incomplete and the actual application of Fitts' law in pointing tasks can only partly support it.

One reason for the incomplete studies on Fitts' law in literature is that during their experiments the researchers did not adopt different performance conditions which would have permitted them (and us) to observe a more comprehensive range of relationships

and influential factors which exist during such tasks.

One exception could be the study introduced in Chapter 2, where a series of experiments with different performance conditions have been performed to reveal the nature and the relationship of speed and accuracy in a pointing task.

However, in Chapter 2, one question still remains untouched: "if the performance conditions vary greatly, can the existing models describe the information transmission of the pointing task precisely?" Actually, the concept of *information processing rate* has almost been ignored since its first appearance in the original paper of Fitts' law[18], and the analysis of the information processing rate has not been discussed much either. Although some researchers still doubt the authenticity of the analogy between the information system and the human system, Chan and Childress have verified from information theory that Fitts' law complies with a unifying noise-velocity relationship based on the analysis of information transmission in human-machine systems[10]. However, their study did not discuss the question proposed here.

Moreover, through this chapter we can also compare different forms of Fitts' law through the analysis of information processing rates. We think that Fitts' law models can be compared from the perspective of their ability to describe the information processing rate. This is the first time that the models have been compared by analyzing their ability to describe the information processing procedure of pointing tasks. Once we are clear about the relationship between the parameters in the models and the information transmission rate , it will not only be easier to decide which model is more applicable, but also be able to help researchers understand some motor control features in pointing tasks. The development of the study of information processing rates will be a new horizon for Fitts' law theory analysis.

4.2   Theoretical Analysis

Reviewing the problem in the light of the origin of Fitts' law, the main reason for the doubts about Fitts' law is that it is not derived from strict mathematical deduction but is based only on a direct analogy of Shannon's information theory[47] (see Equation.4.1).

$$C = B \log_2(\frac{S}{N} + 1) \tag{4.1}$$

According to Shannon, "the capacity of a channel of band W (B) perturbed by white thermal noise power N when the average transmitter power is limited to P (S)" is given by C in Equation.4.1. Shannon also pointed out that information can be transmitted as binary digits at the rate $B \log_2(\frac{S}{N} + 1)$ bits per second, through sufficiently involved encoding systems. " To approximate this limiting rate of transmission the transmitted signals must approximate, in statistical properties, a white noise."

The core function of Shannon Theorem 17 in Fitts' law study is to afford theoretic relations and help to establish a model for the major functional factors in motor tasks.

In Fitts' law models (Equation 1.2), the coefficient b, in ms/bit, can be called the difficulty coefficient, reflecting the information processing rate or, the capacity of the information channel required by the task.

$$\frac{1}{b} = \frac{ID}{MT - a} = \frac{1}{MT - a} \log_2(\frac{A}{W} + 1) \tag{4.2}$$

This part, $1/b$, is similar in calculation method to IP in Equation.1.1. Another term used to indicate information processing rate is TP (throughput)[12] [15] [25][24] [37]. Since b can indicate the information transmission efficiency of a pointing task, with "ideal" performance, the highest information processing rate of the human motor system can be achieved, then with normal distribution of amplitudes, or near to it, the largest value of $1/b$ should be observed.

4.3 Experimental Data Analysis

Since we want to see the information processing rate of different performance conditions, so the varied performance conditions manipulated by the experiment will afford a good chance to observe the information transmission ability with various performance situations.

Therefore, we utilize the data of Experiment SAT3 (refer to Chapter 2) to observe the relationship between different performance situations (different inclinations to speed or accuracy) and the information processing rate. As for the details of the experiment, including the subjects, apparatus and procedures, please refer to Chapter 2.

4.3.1 Results

The horizontal distributions of the input hits are shown in Figure 4.1, 4.2, 4.3, 4.4 and 4.5. The histograms indicate the relative frequency of the real input hits' distribution, and the lines of normal density indicate the normal distribution with the corresponding averages and standard deviations calculated from the real input hits' distribution. For example, in task EA, since about 45% of input hits fell inside the region between 572 and 574, the pole's height is about 0.45 in Figure 4.1. With the average and standard deviation of the input hits, we could also trace the normal density lines, and there are some deviations between the values of the histogram and normal density lines.

In Figures 4.1, 4.2, 4.3, 4.4, and 4.5, the values of the x-coordinate indicate the horizontal position of the input hits in the interface. For the right side, the target center
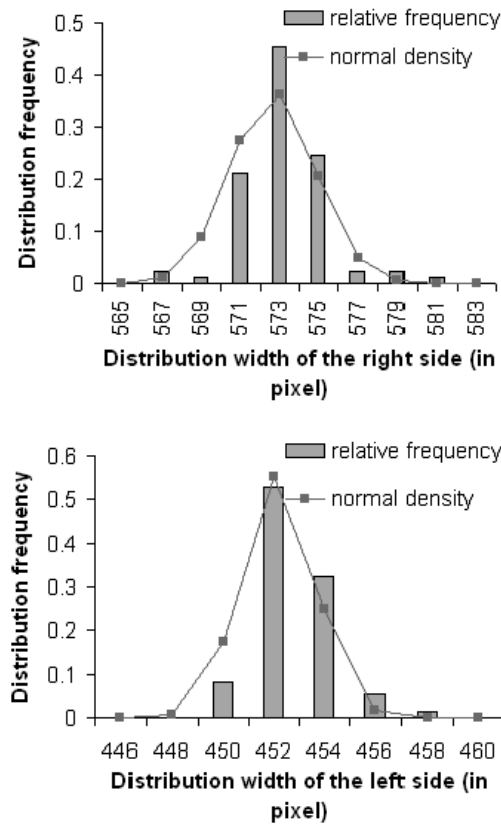
Fig. 4.1   The distribution of input hits in task EA with combination A1.

is 572, and for the left side, the target center is 452. The two borderlines' x-coordinate values of the right target are 566 and 578 respectively, and the two borderlines' x-coordinate values of the left target are 466 and 464 respectively [*1].

4.4   Discussion

Figures 4.1, 4.2, 4.3, 4.4, and 4.5 show that in task N, the input hits' distribution fits the normal distribution better than those in the other tasks. The standard deviations[*2] in Table 4.1 show that the standard deviation of task N is smaller than that of the other task conditions. It means that in task N, the input hits' distribution fits the normal distribution the best of the five task distributions. These data afford us the

---

[*1] We have obtained the distribution figures of all nine $A - W$ combinations and can see similar distribution of other combinations of $A - W$. However, due to the limited space, here we only show one group.

[*2] Here the standard deviations indicate the deviations between the real distributions relative frequency and the normal density, not the standard deviations of the input hits around the target.
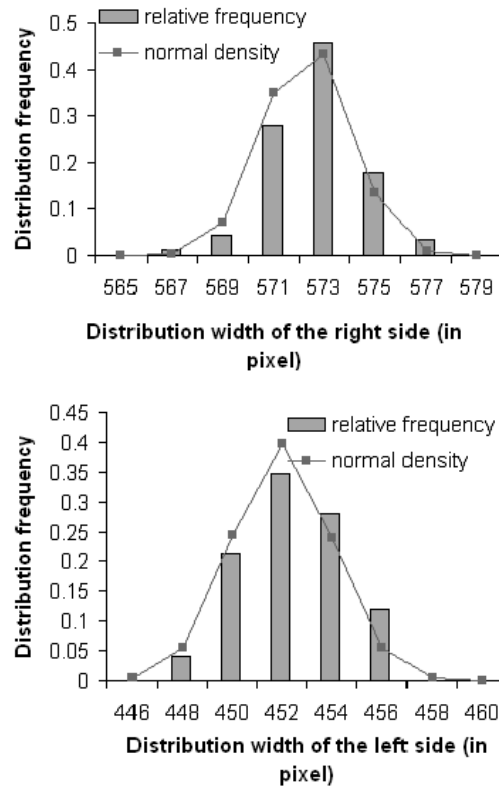
Fig. 4.2   The distribution of input hits in task A with combination A1.

Table 4.1   The standard deviations between the real input hits' distribution of the five tasks and normal distribution

Standard deviations	EA	A	N	F	EF
Right	0.03	0.05	0.01	0.03	0.02
Left	0.05	0.03	0.02	0.02	0.03

chance to observe the information processing rate with ideal performance and with the performance inclinations to speed and accuracy distributed symmetrically around it respectively.

4.4.1   The ID Model

We first applied the basic form of Fitts' law model, i.e. the ID model (Equation 1.2), to the data collected. The result shown in Table 4.2 shows that the slopes of the regression lines varied greatly depending on the different demands on speed and
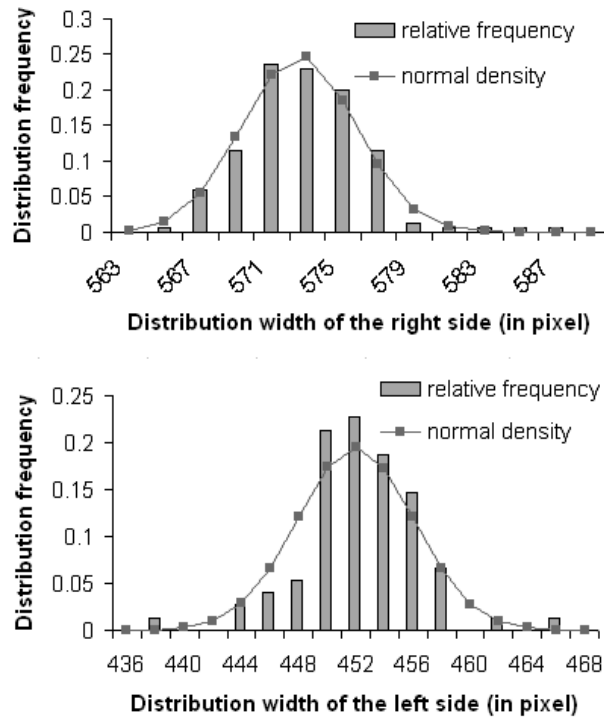
Fig. 4.3   The distribution of input hits in task N with combination A1.

Table 4.2   Linear regression of the five tasks with the ID model

	EA	A	N	F	EF
a	201.1	209.5	170.9	140.5	138.5
b	214.7	185.6	130.6	113.5	83.8
R^2	0.904	0.899	0.961	0.995	0.992

accuracy, i.e., the five tasks (EA, A, N, F, EF). With Table 4.2[*3], it is easy to predict that with greater speed, i.e., from task EA to task EF, the slope of the regression lines (b), will decrease and the regression lines will consequently reach a line parallel to the horizontal line. This means that the ID model cannot include the factor of accuracy properly. When the speed exceeds a certain value, even in different ID conditions, with significantly different error rates, no correction in movement time from different accuracy levels can be integrated by the ID model.

Table 4.2 shows that, although correlations of MT and ID are strong in each task (with R^2 more than 0.90), the coefficients of regression varied greatly from one task

---

[*3] In Table 4.2 and Table 4.3, we only give the regression results of a, but we did not analyze them, because a is a non-information factor of the system[55] and not the main point of discussion in this chapter.
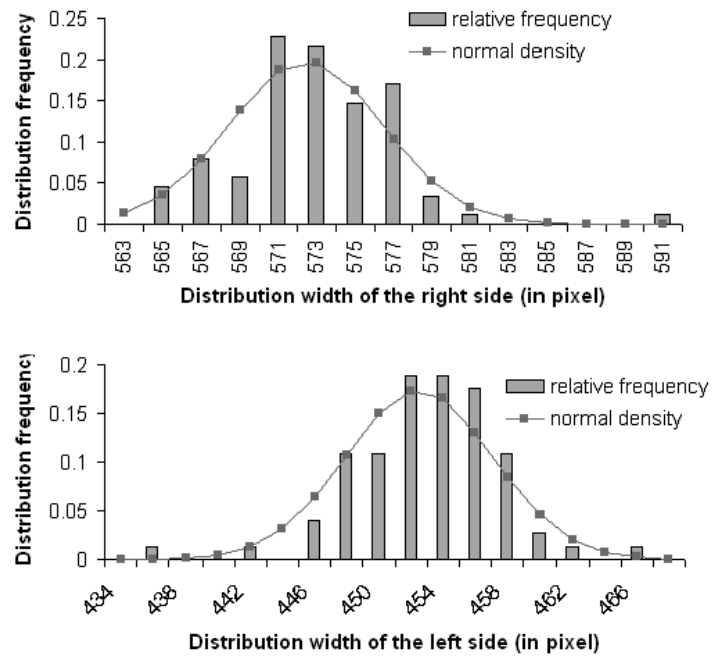
Fig. 4.4   The distribution of input hits in task F with combination A1.

to another. Under different instructions, b changed very noticeably (varied more than 100% from Task EA to EF in this experiment given that we used the middle task, task N, as the standard). With greater emphasis on accuracy b increased accordingly.

Therefore, we produced Figure 4.6 to reveal the complete trend of $1/b$ across the tasks. From Figure 4.6 we can see that with less stress on accuracy, or more emphasis on speed, $1/b$ increases uniformly.

According to Figure 4.6 and the analysis in Sect.1, the biggest value of $1/b$, or the maximum information processing rate (task information capacity) should appear when the signal, or the input hit amplitudes, follow normal distribution. However, here it is obvious that with greater speed, or with more errors, $1/b$ increases if the ID model is applied. This tendency contradicts the analogy of the information theory and seems to indicate that Equation.1.2 is not adequate to describe the information transmission procedure of the human performance in pointing tasks.

4.4.2   The ID_e Model

We then tested the effect of the ID_e model (Equation 1.4). The results of linear regressions between average movement time MT and the effective index of difficulty ID_e are shown in Table 4.3.

In Table 4.3, we can observe that although, as pointed out in [56], R^2 decreased from
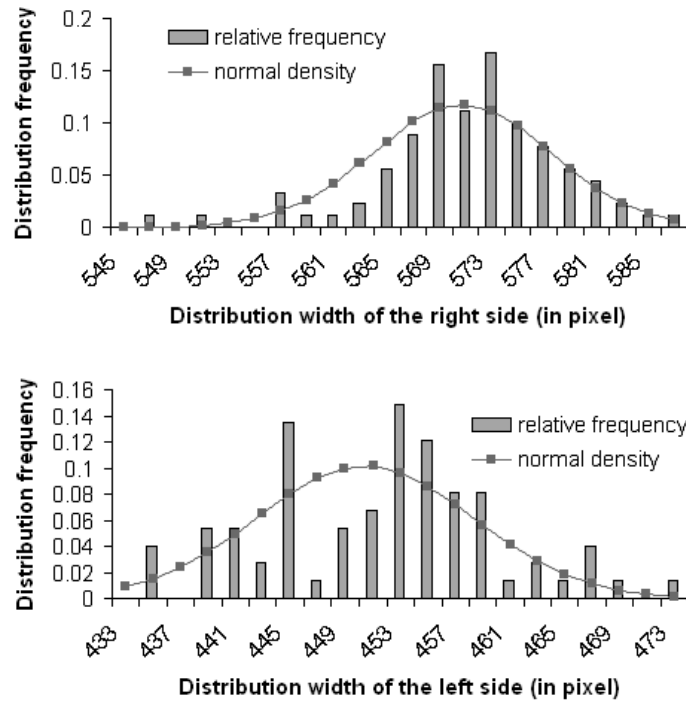
Fig. 4.5   The distribution of input hits in task EF with combination A1.

Table 4.3   Linear regression of the five tasks with the ID_e model

	EA	A	N	F	EF
a	63.6	76.0	75.8	45.0	7.49
b	203.7	187.3	149.1	143.2	147.2
R^2	0.841	0.750	0.867	0.926	0.881

the corresponding regression result using ID for each task, there are also advantages brought by using the ID_e model. First, the variance amplitude of b in different tasks was reduced a lot (varied about 42% from the biggest value to the smallest value using Task N as the comparison basement), and $1/b$, the information transmitting rate for performing the task, was also reduced a lot.

Secondly, Table 4.3 also shows that with greater speed, the slopes of the regression lines decrease slowly and there is a minimum value for the slopes. Beyond that point, the slope will rebound and this implies that no horizontal line will appear. This means that the ID_e model can include the factor of accuracy better than the ID model. When the speed exceeds a certain value, in different conditions, with significantly different error rates, using the ID_e model, correction in movement time with various accuracy levels can be reconciled by the ID_e model.

4.4 Discussion
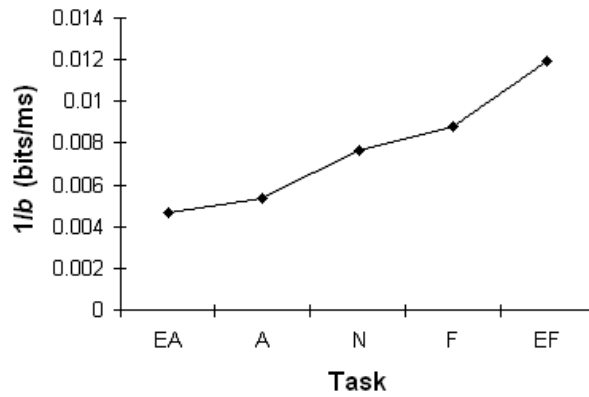
Fig. 4.6   The variation in $1/b$ according to the ID model.
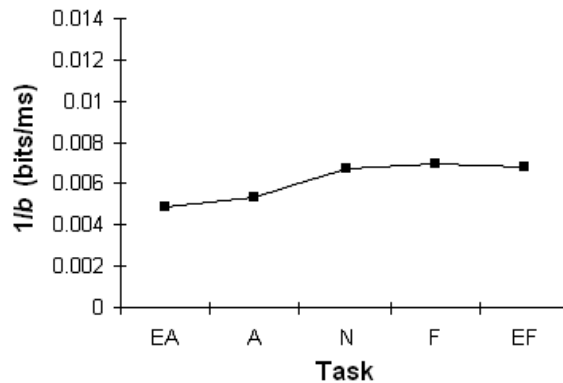
Fig. 4.7   The variation of $1/b$ according to the ID_e model.

We drew a figure showing the variation of $1/b$, Figure 4.7 which shows a mild increase of $1/b$ with the increased emphasis on speed. Moreover, unlike the broken lines in Figure 4.6, there is one potential peak for the broken lines of $1/b$ in Figure 4.7. The peak value of $1/b$, or the maximum information processing rate appears in the case of Task F. In fact, the values of $1/b$ around Task N, F and EF are quite similar to each other, and the $1/b$ value of Task EF does not reach a higher level as it did when using the ID model. As analyzed in Sect.1, when the input hits amplitude follows normal distribution, the information processing rate of the performance system will reach its theoretic maximum value, i.e. the information channel capacity. Although here the peak does not happen in the case of Task N, it shows a big correction when compared with the lines shown in Figure 4.6, where the broken lines reach higher values with more emphasis on speed, which is obviously illogical according to the analogy of the information theory. This figure implies that although the ID_e model is not a perfect

version for the resolution of all problems in Fitts' law, it is an important step in the right direction.

4.5   General Discussions and Conclusions

Although Fitts' law has been applied widely in human computer interaction, it has remained an unclear and unverified analogy with Shannon Theorem 17, and it has not been thoroughly studied. A lot of research has been carried out to verify or improve it. All the efforts are to acquire a reliable model for HCI. Since Fitts' law is derived from the analogy to Shannon's information theory, if Fitts' law really does model performance accurately, according to the information capacity calculation formulation, it should turn out to produce identical characteristics from other aspects of Shannon's Theorem 17, not only from the analogy of "amplitudes to signals" and "variability to noise" [34], i.e., the conceiving of the information processing procedure in a human performance system can be mostly explained in the same way that the information system should be validated.

This study utilized the results of Experiment SAT3 in Chapter 2 which was designed according to the original Fitts' aimed reciprocal pointing task paradigm and in which we had systematically manipulated the performance of the subjects in order to determine (through analysis) whether the information processing procedure expressed in Shannons theory can also be applied to human performance. We consider that this will make the application of Fitts' law more reliable.

By analyzing and comparing the estimates of the regression coefficients, we explored deep into the background theory of the coefficients of the models. This study mainly discussed the variation of b in Fitts' law models and related it to the information processing or transmission rate. Then some points related to the information transmission procedure of the pointing tasks were observed.

With the nominal ID model, $1/b$, the information processing rate varied greatly. This contradicts Fitts' theory according to which the information output of the human motor system in any particular type of task is relatively constant over a range of task conditions. Moreover, the inclination of the broken lines of $1/b$ creates doubts. The biggest value of $1/b$ appeared in Task EF, which means that with error rates of more than 20%, or with an input distribution near to Figure 4.5, the subjects utilized the information channel most efficiently and then the informational transmission rate was able to reach its maximum, i.e. the channel capacity was utilized to the full efficiency. This is illogical and contrary to) the theoretic source of Shannon theorem 17 "to approximate the limiting rate of transmission the transmitted signals must approximate a white noise".

While using the ID_e model, $1/b$ for different tasks remained almost constant, and the biggest value of $1/b$ appeared in task F. Although the peak value of $1/b$ did not appear in task N, the $1/b$ value of task N is near the maximum and there is not a uniform increasing tendency of $1/b$ with more emphasis on speed. The coherence of the tendency of $1/b$ with the changes in performance helps to support the ID_e model better in logic.

These analytical results help us to answer the question proposed early in this chapter: "When the performance conditions vary greatly, can the existing model(s) describe the information transmission in pointing tasks precisely?" The ID model can not describe the performance completely with different speed and accuracy inclinations. On the other hand, the ID_e model can describe the performance much better in this respect.

As we have explained in the earlier part of this chapter, the project introduced by Chapter 2 has included detailed analysis of the comparison of the two models. For the current situation, the two models have different applications because of their different features and advantages. Therefore, the analysis of this chapter is not a mere comparison of the two forms of Fitts' law (Equation 1.2 and Equation 1.4). It points out a direction for better modeling in the human computer interaction field. If we are to continue using the information capacity theory as a background support for human performance measurement, we need to confirm that the variation tendencies of the information processing rate in the human pointing task under varied conditions are consistent with what Shannon theorem 17 implies.

The main contribution of this study is that it discovers another factor which is essential for standard models comparison and evaluation, and it also lends support to the use of the ID_e model of Fitts' law to model the pointing performance in varied conditions. These works will benefit not only Fitts' law researchers in HCI, but also studies in the field of experimental psychology and other related areas.

This is a blank page.

Chapter 5

SH-Model: A Model Based on both System and Human Effects for Pointing Task Evaluation

In the previous chapters, the limitations of the Fitts' law models incurred by the two layers of speed-accuracy tradeoffs have been discussed comprehensively. The traditional methods to resolve the problems are based on the analysis of input hits' distribution (i.e. spatial constraint). In this chapter, we developed a new model (SH-Model) based on temporal distribution to alter the traditional models. The new model and the traditional models are compared in two experiments using AIC (Akaike's Information Criterion), a criterion for statistical model selection. The results show that the new model is better than the traditional ones in performance evaluation. Moreover, the human effects that cannot be observed clearly by using Fitts' law models can be represented by the SH-Model. Therefore, we evaluated four input devices, a mouse, a pen with a big tablet, a trackball and a pen with a small tablet. The comprehensive analysis including the SH-Model, ANOVA analysis and questionnaire can offer a clear comparison of the four input devices. The coefficients of the human factor in the SH-Model show the features of different human performance effects when using different devices. According to our analysis, the mouse is the best for the pointing task designed for our experiment, whereas the trackball is the worst. This chapter not only verifies the application of the SH-Model as a valid evaluation tool for the various devices, but also helps us to observe the human effects separately from the system effects.

5.1   Background Knowledge

The appearance of more and more computer input devices makes designing human computer interfaces a more complex matter. Designers have to choose suitable devices from a lot of candidates. Sometimes the choice can be made comparatively easily, but

in other cases, when more than a few input devices may be applicable, it is not easy to make a final decision. Many factors including physical characteristics (such as mechanical reliability and installation space) and cost have to be considered. Therefore, empirical experiments are necessary if the best selection from a range of input interfaces is to be achieved. Meanwhile, as a basis for empirical analysis in human interface design, researchers use performance evaluation models to afford prediction and evaluation power.

As introduced in the previous chapters, Fitts' law, as a powerful tool for devices evaluation of pointing tasks, are still under suspicion. To resolve the inaccuracy problems brought by the subjective layer of speed-accuracy tradeoff in pointing tasks, the methods proposed by the previous research depend on the spatial distribution of the hits, i.e. researchers are compelled to develop methods which ensure that error rates are limited to 4% (refer to Equation 1.4). Since it has been tested as impossible to establish a simple relationship in the Fitts' law form to describe both the system effects and human effects, here we established a new model based on the concept of temporal distribution which is not limited by spatial constraints.

5.2   The New Model: SH-Model

Our model is based on the general information theory, different from the traditional Fitts' models based on the concept of the capacity of channel of Shannon's theory.

The effects on the performance of a pointing task can be divided into two parts: the system effect and the human effect[*1]. The system effect can be expressed by the condition of a pointing task such as the amplitude between two targets and the target width. The human effect can be indicated by the accuracy of pointing generally.

Regarding the system effect of one-dimensional pointing tasks, assuming the target horizontally and randomly appears within the interaction area, the probability of the target falling into the interface area, P_s, can be represented as

$$P_s = \frac{W}{A + W} \tag{5.1}$$

Considering unstable factors, we use λ, a parameter, to redefine the probability as:

$$P_s = \frac{W}{A + (\lambda + 1)W} \tag{5.2}$$

Thus the self-information of the system is defined as:

[*1] We use "effect" here rather than "factor" because "human factor" has been used with wider meaning.

5.2 The New Model: SH-Model

$$SI_s = \log_2(\frac{1}{P_s}) = \log_2(\frac{A}{W} + \lambda + 1) \tag{5.3}$$

Here SI_s means *Self-Information*. The value of the parameter λ can be estimated by the minimum AIC method (described in Section 5.2.1). To establish a complete and accurate model, we should consider not only system effects but also user performance effects. Thus, we take accuracy in pointing as an indicator of the human effect.

If we use P_h to indicate the probability of hits falling into the target width achieved by the user and call it the "Probability of success", and simultaneously define the ratio of the number of hits falling outside the target width to the total number of hits as the error rate, then P_h + error rate = 1. Thus, Equation 5.4 can be regarded as self-information depending on the probability of success reflecting the effects of human performance.

$$SI_h = \log_2(\frac{1}{P_h}) \tag{5.4}$$

In our calculation, the P_h is calculated by the different combinations of target widths and amplitudes. Since SI_h and SI_s affect the movement time, a linear model which represents the movement time can be stated:

$$MT = a + bSI_s + cSI_h \tag{5.5}$$

MT is the estimation of the real data. a, b and c are the three coefficients.

The reason for taking logarithmic-transformation of MT in Equation 5.5 is that according to observation of the experimental data in the pointing task of the experiments that we had executed previously, the data of the movement time do not follow normal distribution, and most data have the distribution close to the lognormal distribution. One example is shown by Figure 5.1 [*2]. The histogram indicates the relative frequency of the MT data's distribution, and the line of normal density indicates the normal distribution with the corresponding average and standard deviation calculated from MT. Using these experimental data to estimate coefficients may incur biased estimation[49]. According to Everitt[17], after logarithmic-transformation, this kind of data will follow normal distribution (see the data after logarithmic-transformation in Figure 5.2). Therefore, in order to avoid getting biased estimations of the parameters in models, we took the natural logarithm of the data for movement time so that the logarithmic-transformed data followed the normal distribution. Meanwhile, to keep each part of the formulation identical, we took a logarithm of every part (SI_s and SI_h).

[*2] Here in Figures 5.1 and 5.2 we only show the observation of the data of Experiment SAT1 in Chapter 2 as an example because of the limited space. Actually, we can also observe the similar data distribution from other experiments.

Anyway, whether the transformation will provide better estimation of the parameters can be testified by the models evaluation methods (described in the following parts).
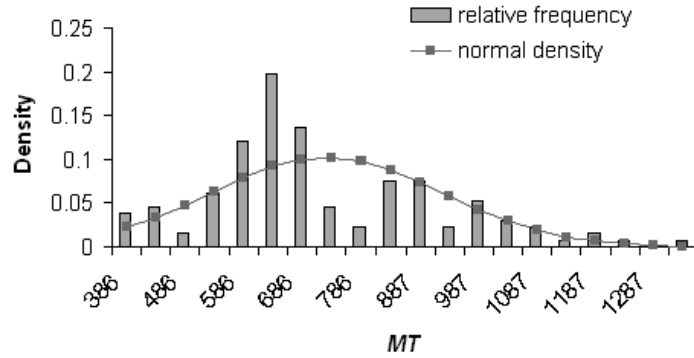
Fig. 5.1   The MT distribution histogram and the normal distribution curve
The standard deviation between the real distribution of MT and normal distribution is 0.036
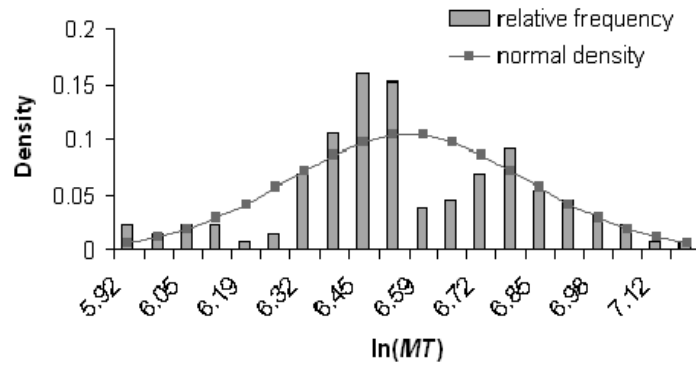
Fig. 5.2   The $\ln(MT)$ distribution histogram and the normal distribution curve
The standard deviation between the real distribution of $\ln(MT)$ and normal distribution is 0.029

Thus, we established the following new model:

$$\ln(MT) = a' + b'\ln(SI_s) + c'\ln(SI_h) \tag{5.6}$$

Here a', b' and c' are also coefficients but probably different from a, b and c in values.

The concept of distribution we discussed here is completely different from the concept in the traditional Fitts' law model researches. In the literature, researchers referred to the spatial distribution of the input hits. This point has been a theoretical and experimental dilemma for researchers of Fitts' law studies as we discussed in the introduction. Contrarily, the concept of distribution in this study was reference to the movement

time (i.e. temporal distribution). We utilize the logarithmic transformation in order to construct a linear model for the logarithmic-transformed movement time data with the normal-distributed error term[*3]. Because Equation 5.6 only gives the part of MT that can be predicted by the model, a more exact expression of the model should be given by adding an error term. Therefore we should add the error term ε at the right-hand of Equation 5.6. As previously discussed in the introduction, here MT is an estimation based on the real data and is recorded as MT_{real}. The real data can be expressed in the following equation.

$$\ln(MT_{real}) = a' + b' \ln(SI_s) + c' \ln(SI_h) + \varepsilon \tag{5.7}$$

Equation 5.7 can be considered to be a regression model for $\ln(MT_{real})$ with $\ln(SI_s)$ and $\ln(SI_h)$ being two independent variables. In this model SI_s shows the effects of the system, such as the effects of different amplitudes and target widths, and SI_h shows the effects of the human. Thus, Equations 5.6 and 5.7 contain complete information of both the system and the human. We call this new model the SH-Model (S indicates the System and H the Human). The variables $\ln(SI_s)$ and $\ln(SI_h)$ are not independent of each other, and their correlation coefficient can be estimated statistically.

When we consider P_h as a parameter in a binomial distribution, its maximum likelihood estimate can be given as follows:

$$P_h = \frac{n}{m} \tag{5.8}$$

where n is the number of the hits falling inside the target, m is the total number of attempts.

If we use Equation 5.8 to calculate P_h, either of two extreme situations could arise. One extreme arises when all the hits fall inside the target, $P_h=1$. The other extreme arises when all the hits fall outside the target, then $P_h=0$. Equation 5.6 could not be applied in either of these situations. We therefore used a Bayesian method to estimate P_h by using a uniform prior distribution [42]. The following equation gives the posterior mean of P_h.

$$P_h = \frac{n+1}{m+2} \tag{5.9}$$

Omitting the error term ε, another form of Equation 5.6 for computing the predictive value of MT is:

$$MT = e^{a'} SI_s^{b'} SI_h^{c'} \tag{5.10}$$

[*3] Here error refers to the difference between the observation of movement time and the estimation of that calculated by corresponding equations.

5.2.1   Model Evaluation by AIC

There are two main ways to evaluate regression models. The traditional one is the use of a coefficient of determination R^2. It indicates the degree of fit of models to the observed data but it cannot represent the predictive ability of models, neither can it be applied to nonlinear models. We usually evaluate models by the descriptive ability and the predictive ability. The former shows how well the model fits the data under analysis, and the latter can indicate how well the model predicts the value of data that can be obtained in future under the same condition. With more parameters, the model's descriptive ability will be improved, but the stability of estimates for parameters will deteriorate so that the predictive ability will decrease. The purpose of statistical modeling is to obtain a model with a strong predictive ability, so the key problem in model selection is how to get a good trade-off between the descriptive ability and the stability of estimates. Thus, it is important to evaluate predictive ability of a model objectively.

Another approach to model evaluation is to use information criteria (ICs), such as AIC and BIC [46]. AIC (Akaike's Information Criterion) is a criterion for model selection [6]. When a number of models are available, we have to select one as *the best* among the alternative models. Akaike's minimum AIC method [6][28][45], is developed for statistical model selection. This method can be interpreted from a maximization of the expected entropy of the predictive distribution approach [7]. It can be applied to comparisons for not only linear but also nonlinear models[8]. It is a better choice for us to compare the new model (SH-Model) with the traditional models (ID model and W_e model)with AIC.

AIC is defined on the basis of the maximum log-likelihood and the number of parameters to be estimated by the maximum likelihood method, i.e., it is defined as follows:

$$AIC = -2M + 2N \tag{5.11}$$

Where, M is maximum log likelihood of the model, N is number of estimated parameters in the model. The term $-2M$ measures the decrease in predictive ability of a model that is contributed to the AIC value by the increase in descriptive ability of a model, and the term $2N$ measures the decrease in predictive ability of a model that is contributed to the model by the increase in the number of parameters of a model (related to the stability of estimates of parameters). Thus, the trade-off between the descriptive ability and the stability of estimates for a model can be obtained by minimizing the value of AIC.

Meanwhile, for the linear regression model with the error term following normal

distribution, the least square estimation agrees with the maximum likelihood estimation. Therefore, by using the method of the least squares, we can not only estimate the parameters in models, but also get the AIC value of different models easily and then compare the effects of different models. For two models that have different numbers of parameters we can estimate the parameters and calculate their AIC values by using the same set of data. Although more parameters can make the model more descriptive, the minimum AIC method itself can reimburse the deviation brought by the parameters before it gives out the final results. That means that AIC can show the consistency between reality and prediction and can test *both* the descriptive ability and predictive ability in a model comprehensively [28]. Overall, the model with the smallest AIC value can be regarded as the best one [6].

Therefore, we decided to use the minimum AIC method to evaluate the models.

5.2.2 Experiment on PDA

To compare the performance of our new model with the traditional models, we used the data from a pointing experiment on a PDA, which was developed according to the one-direction pointing task defined in ISO 9241-9 [25].

(1) Subjects

Twelve subjects (6 male, 6 female, aged from 20 to 22, all right handed) were tested in the experiment.

(2) Apparatus

The PDA used in the experiment was a Psion RevoTM running Windows EPOC, 157 mm (width) x 79 mm (height) x 18 (thickness). The weight of the PDA was 200 g. The display was 480 x 160 pixels (1 pixel is about 0.24 mm). A stylus pen was used as the input device. Experimental software was developed with Java.

(3) Design

The experiment was a 3 x 3 within-subjects factorial design. The factors and levels were as follows:

- Target widths: 10, 20, 40 pixels (2.4, 4.8, 9.6 mm)
- Amplitudes, or distances between the center of targets: 100, 200, 300 pixels (24,

48, 72 mm)

Each subject performed the task in 30 trials in each of nine conditions. There was no rest time between two conditions, because the performance time was so short (within 30 minutes) that no fatigue would be incurred by it. The height of the targets was 90 pixels in all trials. Targets were presented in different order to the various subjects.

Because the actual time slot of the first trial was zero, the total number of data that we processed was 3(targets amplitudes) x 3(target widths) x 29(trials) x 12(subjects) = 3132.

(4) Procedure

In the experiment, two rectangles were shown on the display. One was filled and the other was unfilled. Subjects sat down and held the device with their non-dominant hands. They were instructed not to rest their hands on the table or any other objects during the test. Upon contact the rectangles would switch places and the subjects would again attempt to point to the unfilled rectangle.

Before testing, the subjects were asked to point to the unfilled rectangle (called "target" below) with the input device as fast and accurately as possible. All subjects performed 10 warm-up trials.
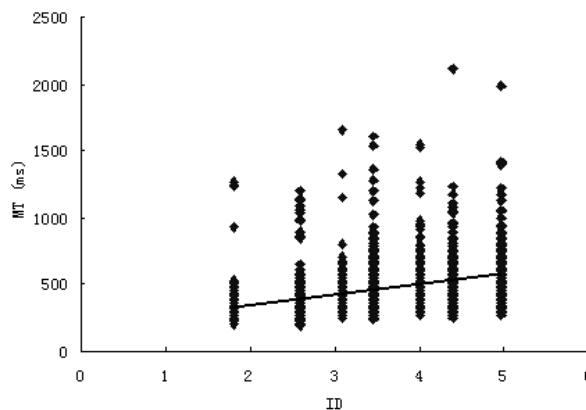
Fig. 5.3   Regression line of the ID model of Exp. on PDA

During the experiment, the subjects accidentally pointed in the wrong direction away from the target (e.g. when the target appeared in the left, the subject pointed to the right). That was related to the inertia and anticipation of the fast movements of the subjects. It was unrelated to the one-dimensional task, so we deleted these accidental hits.   Thus the total valid data is 3132(complete data number)-118(accidental data

5.2 The New Model: SH-Model

Fig. 5.4   Regression line of the ID_e model of Exp. on PDA
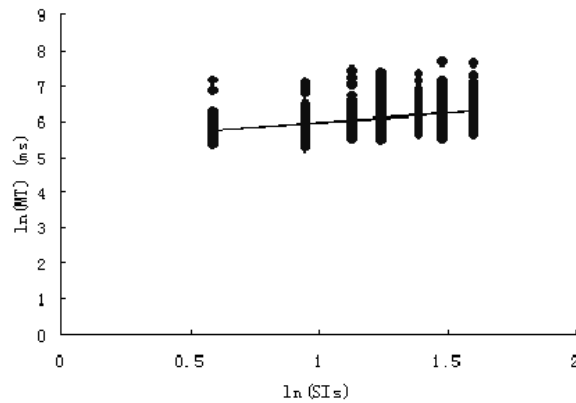
Fig. 5.5   Regression curved surface projection of the SH-Model ($\lambda=0$) of Exp. on PDA

number)=3014.

(5) AIC Values

To test the feasibility of our new model (Equations 5.6,5.10), we applied the experimental data to the ID model (Equation 1.2) and the ID_e model (Equation 1.4) to see whether there was any difference in the effects of different models.

The results of the calculation are shown in Table 5.1. The model with the lowest AIC value will be regarded as the best one (see Section 5.2.1). P_h was calculated by each combination of A and W.

The corresponding AIC value of the ID model (Equation 1.2) is 38927. The regression line is shown in Figure 5.3.

Table 5.1   AIC values of the three models with the data of Experiment on PDA

Model	Formulation	AIC
ID Model	$MT = 197.39 + 75.3\log_2(\dfrac{A}{W} + 1)$	38927
ID_e Model	$MT = -5.05 + 165.3\log_2(\dfrac{A}{W_e} + 1)$	39078
SH-Model	$MT = e^{5.27}\{\log_2(\dfrac{A}{W} + 1)\}^{0.64}\{\log_2(\dfrac{1}{P_h})\}^{-0.03}$	37696

The AIC value of the ID_e model is 39078, which is larger than that of the ID model. The regression line is shown in Figure 5.4.

To compare the effects of the new model with the traditional models, we set the parameter $\lambda = 0$ in the SH-Model (Equation 5.10), then the model is determined as:

$$MT = e^{5.27}\{\log_2(\frac{A}{W} + 1)\}^{0.64}\{\log_2(\frac{1}{P_h})\}^{-0.03} \qquad (5.12)$$

The corresponding AIC value is 37696[*4].

In Figure 5.5, for convenience of contrast with the other two regression lines, we used the two dimension figure for the regression curved surface of the SH-Model, so there is no SI_h in this figure. It can be regarded as a projection of the curved surface on the surface of $\ln(MT)$ and $\ln(SI_s)$. This model includes a negative coefficient in the place of c' (-0.03). It is easy to explain: if the subject performs quickly, he or she may make more mistakes, so P_h will be smaller and SI_h will become bigger, then the value of the MT will be smaller for any occurrences of the negative value of c'.

From the above computation with the PDA experimental data, the SH-Model obtained the lowest AIC (37696). Therefore, this model can be regarded as the best of the three models. The traditional models have bigger AIC values. This indicates that those models cannot describe the data that agree with the real data as accurately as the new model can. These conclusions can effectively test the reasons for the doubts regarding traditional Fitts' models.

As previously noted, the input hits may not be limited by the outer boundaries of the two targets, and may not be 0. Thus we changed the value from $\lambda=0$ to $\lambda=1$, $\lambda=2$, and $\lambda =3$, and the AIC results are shown in Table 5.3.

[*4] Here the AIC value was computed by adding twice the sum of all data to the AIC value of the model for the log-transformed data, so it is comparable with the others [28].

Table 5.2   AIC values of the three models with the data of Experiment on Tablet PC

Model	Formulation	AIC
ID Model	$MT = 136.46 + 119.99\log_2(\dfrac{A}{W} + 1)$	47465
ID_e Model	$MT = 53.52 + 153.05\log_2(\dfrac{A}{W_e} + 1)$	47859
SH-Model	$MT = e^{5.40}\{\log_2(\dfrac{A}{W} + 1)\}^{0.71}\{\log_2(\dfrac{1}{P_h})\}^{-0.00012}$	46077

5.2.3   Experiment on Tablet PC

To make sure our models have universality and are not limited to PDA experimental data, we utilized the data of the Experiment SAT1 (see Section2.2) to see if it indeed supports our conclusions. In this section we called Experiment on Tablet PC for convenience.

The ID model (Equation 1.2), the ID_e model (Equation 1.4), and the SH-Model (Equation 5.6 or 5.10) applied from the experimental data and their AIC values are shown in Table 5.2.

The regression curves of the three models are shown in Figure 5.6 and Figure 5.7. Figure 5.8 shows the regression curved surface projection of the SH-Model. P_h was calculated by each combination of A and W. Therefore, the different functions of different instructions during the tasks can be expressed by Equation 5.6 effectively.
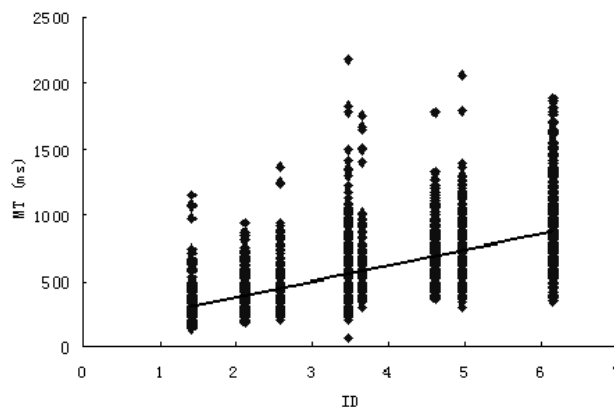
Fig. 5.6   Regression line of the ID model of Exp. on Tablet PC

From the above computation, we can conclude that with the data of Experiment on Tablet PC, the SH-Model still has the lowest AIC (46077) (see Table 5.2). The experimental outcome gave powerful support to our previous conclusion.

The AIC results of the SH-Model with different λs are shown in Table 5.3.
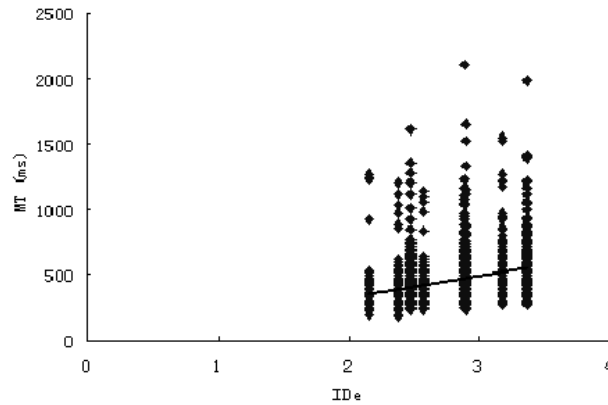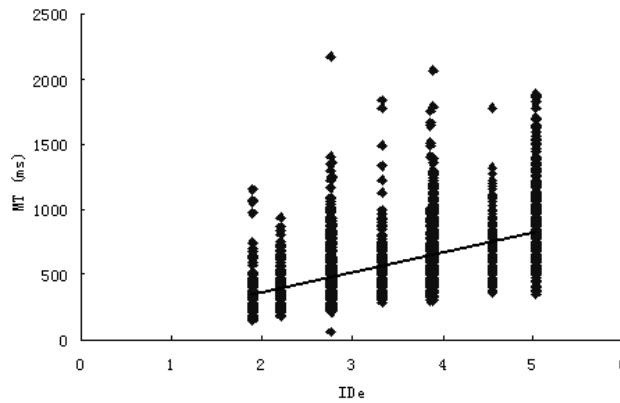
Fig. 5.7 Regression line of the ID_e model of Exp. on Tablet PC
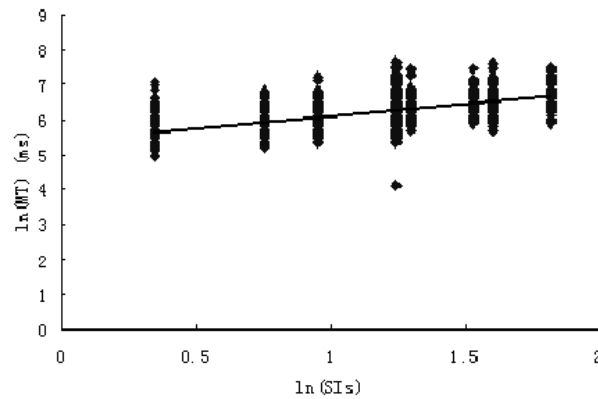
Fig. 5.8 Regression curved surface projection of the SH-Model (λ=0) of Exp. on Tablet PC

5.2.4 Discussion

This study proposed an alternative model, SH-Model, for the development of the solution for Fitts' law's problems. Using the ID model, if error rates have not been considered, in other words, if the experimenter does not control the error rates during the experiment, or if subjects cannot follow the instructions accurately, then the

Table 5.3 AIC values of SH-Model of different λs

Experiment	λ=0	λ=1	λ=2	λ=3
Experiment on PDA	37696	37689	37691	37694
Experiment on Tablet PC	46077	46037	46032	46039

5.2 The New Model: SH-Model

Table 5.4 Error Distribution Percentage

	SH-Model		ID Model		ID_e Model	
	Negative error (%)	Positive error (%)	Negative error (%)	Positive error (%)	Negative error (%)	Positive error (%)
Experiment on PDA	47.8	52.2	54.9	45.1	55.1	44.9
Experiment PC on Tablet	52.8	47.2	58.9	41.1	59.7	40.3

experimental data may not follow the normal distribution and/or keep the error rate of 4%. Using the ID_e model as a post hoc method, though the error rate is modified to be 4%, it is still not certain whether or not the experimental data can follow the normal distribution. This means there may be a difference between the reality and the prediction.

We compared the AIC values of different models including two traditional ones with the new one designed in this study. From Tables 5.1 and 5.2, the AIC results show that the new model is much better than the traditional ones. There is another noteworthy point: the AIC values of the ID_e model were even greater than the ID model. One reason is that Experiment on PDA was developed on the PDA, and the subjects could not rest their hands on knees, tables or any kind of platform. This might have made them produce more mistakes. So the standard deviation of the experimental data of movement time was greater and that made the AIC values larger. In Experiment on Tablet PC, the AIC value of the ID model was still better than that of the ID_e model although the difference in the AIC values between the two models decreased. This may be due to the fact that using W_e to modify the Fitts' law model is from the point of the input hits distribution. It may not contribute to the modification of MT. AIC or similar methods are able to show whether the W_e model can be more advantageous. From the viewpoint of modification of MT, the greater AIC values of MT mean that the ID_e model cannot model the performance better than the ID model. We calculated the error rate for Experiment on PDA (26.11%) and Experiment on Tablet PC (10.94%). With these two kinds of data on the PDA and the tablet PC, the SH-Model always offers the smallest AIC, which means the new model is better than the traditional ones.

The larger AIC values of the traditional models lend support to doubts about the traditional Fitts' law formulations. At the same time, this also testifies to the feasibility of using AIC values to examine different models in human computer interaction. Although we introduce one more parameter here, the AIC results can show that there is a great difference between the new model and the traditional ones. The qualitative difference is greater than twice the number of the new parameters plus 1, e.g. using the data of Experiment on PDA, $\lambda=0$, the AIC value of the SH-Model is 37696, and that of the ID model is 38927. There is only one more parameter in the SH-Model so

to double the sum of 1 parameter plus one is 4, i.e. 2 x (1 parameter + 1) = 4. Then 38927-37696=1231 is much bigger than 4.

Regarding the benefits of the SH-Model, first, it provides the development of the solution for Fitts' law's problems. It is established based on the concept of temporal distribution rather than the traditional concept of spatial constraint. Second, with the new model, we need not keep within the error rate of 4% constantly and strictly, either by controlling experimental conditions or when calculating W_e. Third, it can distinguish between system and human effects.

We propose using "SH-Model" as the name for the new model because we proposed the concept of separating the two parts in one model. Indeed, SI_s in Equation 5.3 is different in its physical meaning from the traditional Fitts' law formulation. It is decided by the situation of the task. Meanwhile, the SI_h in Equation 5.4 is obviously determined by the subjective effect of the performers. The two parts of the information can be observed clearly and distinctly in the SH-Model. From the traditional models, although the system effect and human effect are both considered, they cannot be separately considered and hence they are not easy for others to observe. We need to analyze deeply to find the effects of the two separate parts upon the performance.

In the SH-Model, we add another parameter of P_h to consider the human effect. This means that we need to know the error rate to apply this model. In this situation, the SH-Model has a similar function to the ID_e model (including the behavioral effects or accuracy into movement time). However, because the ID_e model only modifies the error rate to be 4%, we do not know whether the data follow the normal distribution. The SH-Model is established from the viewpoint of temporal distribution (movement time), thus we can fix the error rate and estimate the MT at different levels (i.e. not only 4%). Furthermore, P_h can affect the movement time so that with this information the model can be more reliable. The benefits derived from the increase in complexity in the new model outweigh any inconvenience caused by increased complexity.

We have tested the effects of four different parameters: λ=0, 1, 2, 3. The comparison of the results (see Table 5.3) shows that, for Experiment on PDA, $\lambda = 1$ produced the smallest AIC, for Experiment on Tablet PC, $\lambda = 2$ produced the smallest AIC. Comprehensively, this means that most of the input hits would fall into the range of $(A + 2W)$ to $(A + 3W)$ as shown in Figure 5.9.

The smallest AIC value of the model with $\lambda = 1$ or $\lambda = 2$ shows that most of the hits, including the successful attempts and the misses will fall into the shallow gray area indicated in Figure 5.9. This shows that interfaces with targets should leave at least this much space between any two targets. The reason for the different optimal λ determined by the minimum AIC method in the two experiments is that we used apparatus with different screen sizes. The PDA screen is so small that the subjects

were unable to move their hands freely and also that their attention was focused on a much smaller area. That might make them point to a smaller range. Conversely, the tablet PC's screen is rather big, so it is natural for the subjects to point to a big range. Then the areas of most hits are different and so are the optimal λs, i.e. optimal λs may be expected from different devices. Certainly we can select even more values for parameter λ. Then the corresponding AIC values can indicate whether there are better λs for the models.
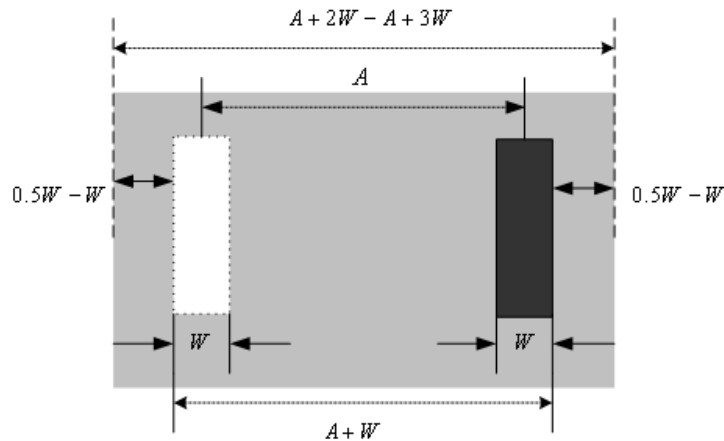
Fig. 5.9   The range of input hits when λ =1 to 2

We analyzed the experimental data through the $ID - MT$ figures (shown in figures 5.3, 5.4 and 5.6, 5.7). Through these figures we can see that the numbers of those hits beside the two parts of the regression lines are significantly different. This means that the distribution of the data is obviously different from a normal distribution. From the SH-Model we can see that the data's distribution beside the two parts of the regression plane are nearly symmetrical (see Figures 5.5 and 5.8). We can also see whether the percentages of the number of errors are greater than 0 or smaller than 0 from Table 4. It is easy to conclude that after logarithm transformation, the data follow normal distribution more accurately and with the new model we can evaluate performance better [17][45]. The estimated MT values can be observed from the trend lines in Figures 5.3 to 5.8[*5].

[*5] A point worthy of noting is that we cannot compare these figures simply, because different units of the axis of the coordinates were applied. Meanwhile, since the movement time's unit is in milliseconds (ms), the values of the Y axis are very big and it is not easy to observe the difference from movement time. Also because the unit of Figures 5.5 and Figure 5.8 is different from Figures 5.3, 5.4, 5.6 and 5.7, the simple comparison of the values of the Y axis will be meaningless.

5.2.5   Conclusions

Our goal in this study is to provide an alternative model for solving the problems of the traditional Fitts' models. In the experiments, we used the data derived from the use of a stylus pen to test the feasibility of different models. We have demonstrated that the SH-Model is better than the traditional models based on the AIC analysis in the PDA and tablet PC experiments.

Our future work includes investigations of various pointing tasks and more pointing devices in order to clarifying the SH-Model's range of application.

5.3   Application of the SH-Model

5.3.1   The Application of Models in Device Comparison

Referring back to the literature on human computer interaction, Fitts' law was first applied to computer input device evaluation by Card and colleagues [9]. A mouse, an isometric joystick, step keys, and text keys were compared in target selection tasks in that research. That study promoted the commercial application of the mouse. After Card and colleagues' avant-garde research which applied Fitts' law to commercially used input devices, some similar studies were carried out using Fitts' law as a basis for comparison [34]. These studies were invaluable in the original testing and evaluation of many commercial computer input devices.

Epps had evaluated six input devices (two kinds of touchpad, a mouse, a trackball and two kinds of joystick) using three different models[16]. In his research, the trackball and the mouse provided the best target acquisition performance of the six devices.

However, in the evaluation study carried out by MacKenzie et al., among the three devices evaluated, the trackball was a poor performer for both pointing task and the dragging task [33]. The performance of the mouse and the stylus with tablet in the pointing task was similar, while that of the stylus with tablet was slightly higher. Based on the introduction of the application of ISO 9241 part 9 standard, Douglas, Kirkpatrick and MacKenzie compared a joystick and a touchpad in their study [15]. They concluded that for the one-directional pointing task, the joystick was slightly better than the touchpad, but the result was not significant.

Another evaluation study was carried out by Poika Isokoski and Roope Raisamo[24]. Six mice were compared in their experiment. They verified that although the mice performed almost equally well, the larger mice performed a little more slowly.

In these studies, all the researchers used a traditional model such as the ID model or the ID_e model. An exception was the comparison study of Accot and Zhai [2]. They used the Steering Law to compare five devices (mouse, stylus, touchpad, trackball and

trackpoint) in a steering task. According to their study, the compared devices could be classified into three groups, and the performance rankings for the groups was 1) the tablet and the mouse, 2) the trackpoint, 3) the touchpad and the trackball.

Although many studies on the use of models to evaluate devices have been developed, the SH-Model, a newly proposed model, has not gained enough attention and support as an effective evaluation tool for input devices. There has been no study on device evaluation based on the SH-Model. The SH-Model is a model derived from the time series analysis; it is not based on the analysis of spacial distribution of the input hits made by the subjects. Moreover, it can clearly show information about the individual humans effects on performance as distinct from the information about the system [44]. This characteristic may help us to observe the devices features more precisely.

5.3.2   Research Purposes

Through this research, we aim to realize two purposes. First, we want to test whether the SH-Model can be generalized, i.e. whether it can be reliably applied to the evaluation of different pointing devices. Although the first study on the SH-Model has afforded the data of an experiment on a tablet personal computer with an electronic pen as the input device[44], the feasibility of using the SH-Model to evaluate other kinds of devices has not been verified. In this part, we will make it clear whether the SH-Model is suitable for other types of devices such as a mouse and other environments such as a desktop personal computer system.

The second purpose of this research is to show how to apply the SH-Model to device selection and how to observe the characteristics of different input devices through the SH-Model. Previously, researchers used to depend on Fitts' Law as a primary input device evaluation tool to do similar analyses.

5.3.3   Comparison Experiment

We executed an experiment according to the Fitts' law paradigm experiment with four different devices to produce the data for device comparison.

(1) Subjects

Twelve subjects, of different genders and ages (3 females and 9 males, 21 to 32 years old, average age 25) participated in the experiment. All the subjects were right hand dominant.

Fig. 5.10　Experimental interface

(2) Apparatus

The experimental apparatus included a desktop personal computer (screen size: 43cm/17.0" Diagonal, pixel pitch: 0.264mmH x 0.263mmV, each pixel on the screen was 0.264 mm wide) (see Fig.5.10), a mouse (Agiler AGM 6124X), a pen with big tablet (WACOM Intuos. Graphics Tablet model i-900 serial), a pen with small tablet (WACOM FAVO Tablet F410 ET0405), and a trackball (Microsoft Trackball Explorer 1.0) (Fig.5.11). The experimental program utilized the full-screen mode as shown in Fig.5.10.

This experiment, however, should not be regarded as an absolute comparison of the four input devices mentioned above. As described in [2], to definitely compare the performance of different types of devices is impracticable, because of a lot of hardware and software implementation details including resolution, sampling frequency, form factors, sensor technology, transfer functions, etc. However, it is still feasible to make an estimate of different devices in performing a certain task for a common user when he or she uses the devices with the default settings. We agree with this opinion and compare four representative devices for the pointing task while keeping all the default values in the system software. In this way we were able to determine which device is the best for pointing when an average user simply uses it as an input tool.

(3) Design

The combinations of different width (W) and amplitude (A) between the two strips were set at W = 12, 36, 72 pixels and A =120, 360, 840 pixels (see Table 3.1). The order of the 9 width and distance combinations was randomized. Twelve trials were presented in each combination, with the first tap excluded in analysis.

5.3 Application of the SH-Model
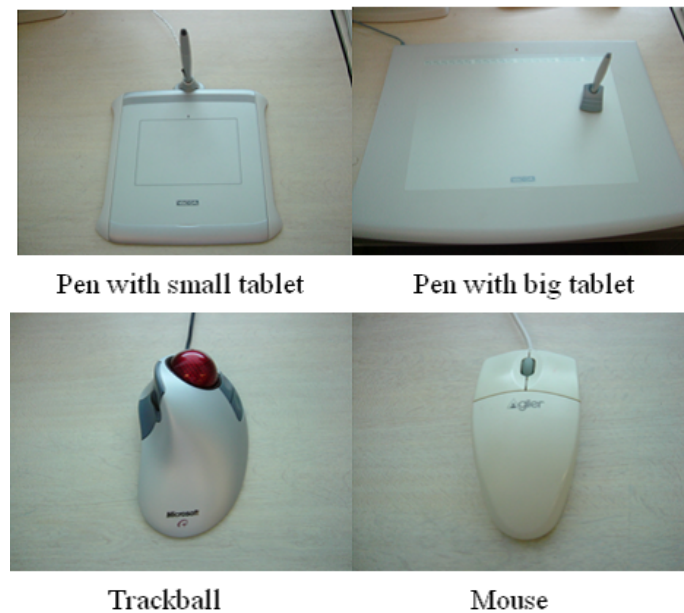
Fig. 5.11   Experimental input devices

(4) Procedure

Similar to the paradigmatic Fitts' experiment[18], participants reciprocally pointed with appointed input devices to a pair of vertical strip targets which appeared on the screen of the PC. Once the white rectangle (the target) was tapped, the position of the white rectangle and black rectangle were reversed, and the subjects were required to tap the current white one as quickly and accurately as possible. A warning beep was played if the subject tapped outside the target.

To reduce the effects of bias due to the users varying degrees of familiarity with the devices, we arranged adaptation practice before the real experiment. The adaptation practice time was 20 minutes for each device and each subject. Since the pointing task is a simple task, it did not take much time for users to become familiar with it. Therefore, we assumed that 20 minutes would be enough for the practice segment. The order of the four devices of the mouse, the pen with big tablet, the pen with small tablet and trackball were balanced by a Latin Square.

After the experiment, we asked each subject to fill in a questionnaire, which helped us to obtain their views on each device. From this we compiled overall usability rankings, reasons for their preferences and other comments. As for the rankings, we asked the subjects to give scores to each of the four devices, the highest score being 4 and the lowest being 1.

Table 5.5   AIC values of the three models for the four devices

Devices	ID model	ID_e model	SH-Model
Mouse	16024.3	16050.5	15631.1
Pen with small tablet	16540.5	16871.1	16031.9
Pen with big tablet	16682.5	17383.1	16118.8
Trackball	18095.7	18103.0	17527.5
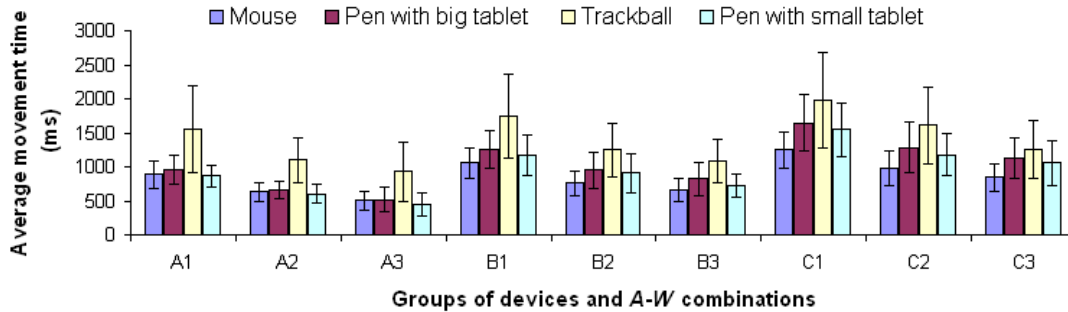
Fig. 5.12   Average movement time of each combination for the four input devices

(5) Results

The AIC values of different models and different devices are listed in Table 5.5. It is obvious that the SH-Model can obtain the smallest AIC values for each of the four devices.

The average movement time for each A and W combination is shown in Fig.5.12 and Table 5.7. In Fig.5.12, A1, A2, ..., C3 represent the nine combinations of different A and W set in the experiment. The respective IDs of the nine combinations (from A1 to C3) are listed in Table 5.7.

Using ANOVA to analyze the data, we were able to determine a significant difference among the four devices on average movement time, (F3,32) = 4.98, p < 0.01.

Fig.5.13 helps us to see what would happen if we use the ID model (Equation (1.2)). However, since Equation (1.2) cannot depict the complex interactions between the effects of the tasks difficulty and the performers' subjective inclination, it is not a completely reliable model for device evaluation.

With partial modification of the ID model, the ID_e model may depict a more reliable picture of the trend lines for the tasks and the four input devices (Fig.5.14). Unfortunately, all R^2 values for the four tasks' regression lines are smaller or much smaller than those derived from the ID model. The R^2 of the regression line for the pen with big tablet is actually too small to be reliable. This means that although the ID_e model helps to observe the reality more clearly, the results brought by it are

5.3 Application of the SH-Model
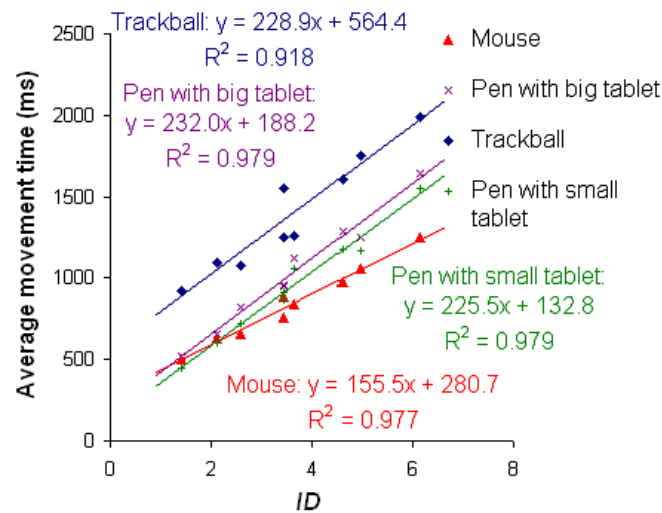
Fig. 5.13   Regression lines for the four tasks using the ID model
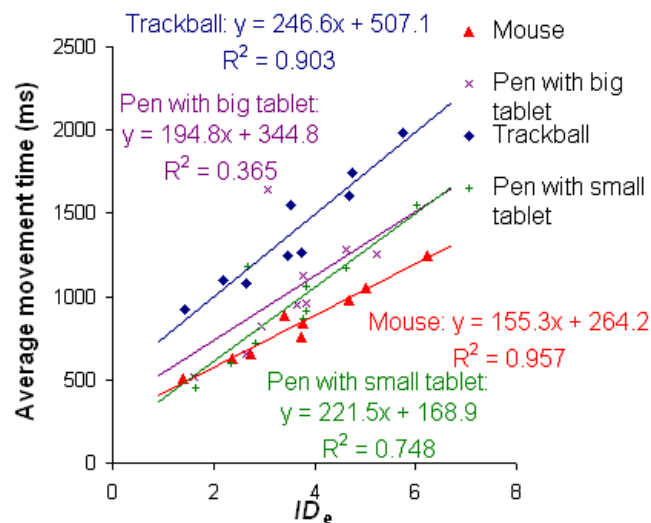
Fig. 5.14   Regression lines for the four tasks using the ID_e model

simultaneously unstable.

Thereafter, to check the feasibility of the SH-Model for the evaluation of the four different input devices, we applied the experimental data to the SH-Model to see whether there was any difference among the effects of different devices and, if there was some difference, which one would be the best one for pointing tasks. Coefficients estimated by the least square method are shown in Table 5.6. Note that although the estimation of c in the SH-Model is comparatively small as an absolute value, the modifying quantity is still significant for the non-linear formulation (Equation (5.10)).

Table 5.6 Coefficients in the SH-Model estimated by the least square method

Coefficient	Mouse	Pen with big tablet	Trackball	Pen with small tablet
a	6.30	5.91	7.02	5.74
b	0.533	0.789	0.461	0.837
c	0.0771	0.000837	0.104	-0.00594
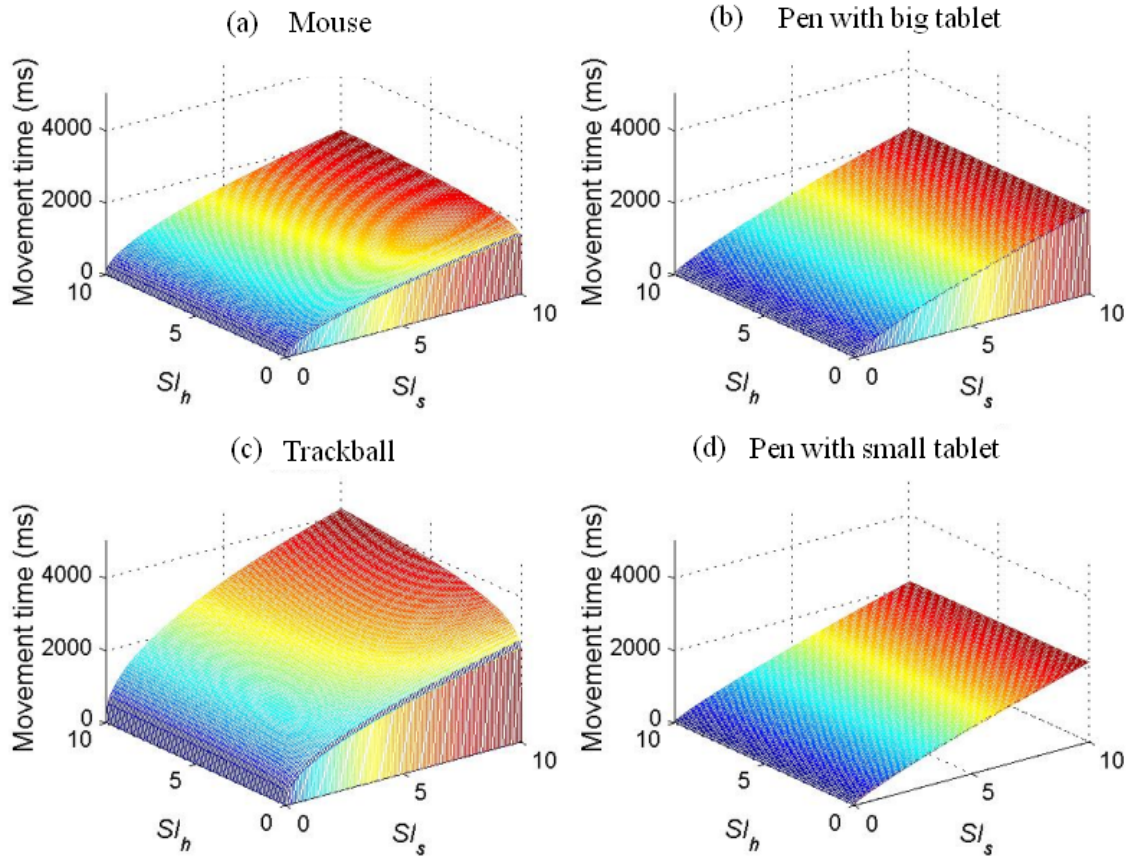
Fig. 5.15 Regression curving surfaces with the SH-Model of the four tasks

Fig.5.15[*6] shows the interaction of the two factors (SI_h and SI_s) in pointing tasks and their effects on movement time. We can see that in most cases, the mouse took the least movement time in the pointing task for the desktop computer. In the mean time, the trackball took the most time of the four devices.

When we used ANOVA to analyze the ranks of the devices assessed by different subjects, we found a significant difference existed among the four input devices, (F3,44) = 14.3, p<0.0001. From Fig.5.16, we can see the comprehensive ranks of the four input

---

[*6] Here the curving surface of the pen with small tablet is different from the others because the value of coefficient c is minus, thus some of the curving surface cannot be shown in this figure.
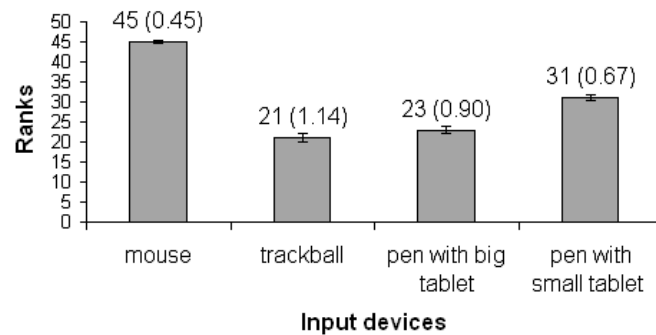
Fig. 5.16   Rankings of the four devices according to the preferences of the subjects

devices. Of the four devices, the mouse obtained the highest score of 45 with a tiny standard deviation of 0.45. The second in ranking which was accepted by most subjects was the pen with small tablet, with a standard deviation of 0.67. The pen with big tablet acquired a score of 23 with a standard deviation of 0.90, and the last one was the trackball, with a comparatively bigger standard deviation of 1.14.

The error rate for the tasks on the four devices are 1.9%, 2.5%, 0.8%, 2.5% respectively for the mouse, the pen with big tablet, the trackball and the pen with small tablet.

5.3.4   Discussion

In the previous part of this chapter, the SH-Model was shown to be better than the traditional models. After this experiment, we also calculated the AIC values of the different models with the data derived from different input devices. Table 5.5 shows the AIC values of three models. The fact that the SH-Model can obtain the smallest AIC values for different input devices not only supports the conclusion in the previous study that the SH-Model can successfully modify the inaccuracy brought by using the traditional Fitts' law models (the ID and ID_e models), but also help to widen the application of the SH-Model to various input devices besides pens. The SH-Model can also be applied to a desktop system which may be different from a tablet PC.

The following paragraphs will first give some ANOVA analyses and some basic analysis based on the traditional evaluation methods, and then we will use the SH-Model to analyze the features of the four devices tested by our experiment. We also compared the evaluation results of this research with those of other related papers.

Table 5.7   Comparison of average movement time of different input devices and
$A - W$ combinations

Combinations	ID	$\ln(SI_s)$	Mouse	Pen with big tablet	Trackball	Pen with small tablet
A3	1.42	0.347	505.4	517.2	925.9	448.1
A2	2.12	0.749	631.5	656.8	1096.5	601.2
B3	2.58	0.950	657.9	821.9	1078.5	721.8
A1	3.46	1.24	886.5	959.9	1547.3	868.6
B2	3.46	1.24	758.1	948.9	1247.3	909.3
C3	3.66	1.30	840.5	1124.8	1259.9	1057.0
C2	4.60	1.53	980.4	1285.2	1606.5	1178.9
B1	4.95	1.60	1054.7	1254.4	1747.6	1170.5
C1	6.15	1.82	1248.7	1643.0	1987.9	1548.0

(1) Analysis Based on Traditional Methods

According to the ANOVA analysis, there is a significant difference in the average movement time corresponding to each A-W combination of four input devices. Based on Table 5.7 and Fig.5.12, we can see that as the difficulty of the tasks (ID and ID_e) increased, more time was needed for completing the tasks for all the devices. This trend can also be observed from Fig.5.13 and Fig.5.14.

However, as discussed in the HCI modeling field, we can never merely consider the average movement time because other factors also interact with the difficulty of any task. For example, in the experiment, when we compare performance based on movement time, the trackball will be the least suitable one. However, when the comparison is based on error rate, the trackball is a good choice. In the SH-Model, we use $\ln(SI_h)$ to express this factor. The effects of this part will be described mainly by the error rate. Amongst the four devices, the subjects made the fewest mistakes when they used the trackball to fulfill the task (error rate = 0.8%). When using the pens (the pen with small tablet and the pen with big tablet), the subjects made the most mistakes.

A different and even contrary result is reached depending on whether error rates or movement times are used as the basis for comparison. This means that we need a more comprehensive, inclusive and reliable method to assess and compare the four devices.

(2) Analysis Based on the SH-Model

Fig.5.13 shows that according to the ID model, in most cases, the mouse costs the least in terms of performance time. With a greater index of difficulty, the difference in performance time is bigger. Nevertheless, Fig.5.13 cannot show the error rate or individual performance situation, therefore it is not adequate for device evaluation for

the pointing tasks.

Since there is a factor of real performance (the effective target with, W_e), Fig.5.14 can show a more comprehensive comparison for the four input devices using ID_e. Fig.5.13 and Fig.5.14 roughly show that for the pointing task with an identical requirement for both speed and accuracy, the mouse is better than the other devices and the trackball performed the worst. However, with the ID_e model, as we have observed, all R^2 values for the four tasks' regression lines are small and sometimes too small to be reliable. Moreover, although the information on individual performers is included in Fig.5.14, we cannot observe it directly from this figure. Therefore, we used the SH-Model to do the device evaluation and to observe the features of different devices.

Fig.5.15 gives a more comprehensive description of the pointing task. In this figure, we can see that the effect from SI_h is obvious. This is seen most clearly by referring to the data for the trackball. When the task situation is fixed, a bigger SI_h derived from smaller P_h will incur a bigger increase in movement time (see Fig.5.15). For the other devices, the increase is not so apparent. This implies that in pointing tasks which require subjects to perform quickly and accurately, it is not easy for the subjects to increase speed when they use the trackball. This agrees with the fact that the subjects were able to obtain greater speed when using the other devices. Considering the effect of SI_s simultaneously, Fig.5.15 (a) shows that the mouse is the most suitable device for the pointing task because when task difficulty (which can be expressed by SI_s) is increased, people do not need to slow down greatly to keep an acceptable degree of accuracy. Furthermore, when using the mouse in any given task situation, it is not very difficult to increase the speed.

The SH-Model can also give us a more precise description of a user's performance through the coefficients of b and c (see Table 5.6). The values of b for the pen with big tablet and the pen with small tablet are comparatively larger than for the other two devices, which means that the difficulty of the task has more effect on movement time with the pens because $\ln(SI_s)$ is a factor decided by the task. When SI_s increases, much more time is needed when using the pens than when using the mouse or the trackball. The value of b for the trackball is the smallest, which means that the effect of the task is not obvious. As for the values of the coefficients for c, with smaller speed or correspondingly smaller error rates, movement time will be decreased more to compensate for the over emphasis exerted on the task by the subjects[*7]. Conversely, with greater speed and usually a correspondingly greater error rate, movement time will be decreased slightly. Sometimes there may even be additional effects if the error rate is too big. We can also see that the value of c for the trackball is the largest, which means

---

[*7] As we have explained previously in this chapter, although the absolute value of c is very small, since the SH-Model is a non-linear model, its function is still significant.

the movement time will be reduced more in accordance with the requirement of the task. On the other hand, almost no amount of time will be reduced for the movement time of the pens.

We can derive coherent conclusions from the subjects' comments on the devices using the SH-Model. From Fig.5.16, it is clear that the mouse is supported by most subjects as the most suitable input device for pointing tasks. On the other hand, most of the subjects did not show any preference for the trackball. For the pens, the subjects' opinions were strongly divided. The pen with small tablet obtained a score of 31 whereas the pen with big tablet only got a score of 23, just a little higher than the trackball.

Comprehensively speaking, through applying the SH-Model (Table 5.6 and Fig.5.15), it is clear that for the pointing task designed for this experiment, the mouse is the most suitable input device, the pen with small tablet ranks second, the pen with big tablet ranks third, and the trackball ranks fourth (last).

However, these results do not mean that trackball is definitely bad for users. Actually it is designed for precise pointing in a small area of screen space. The results we obtained through this study support its ability to acquire low error rates. Simultaneously, the results also show that the trackball is not suitable for target acquisition tasks which require higher rates of speed. Actually, the evaluation results of the four candidate devices may vary in different tasks, such as writing tasks, steering tasks [2], etc.

(3) Comparing the Evaluation Results with Related Works

Here it is necessary to analyze the evaluation results of this study and the related studies. The results in this study support the conclusion of Card's paper[9], however, this study gives a more comprehensive analysis of the performance of the four different devices through the application of the SH-Model.

In the pointing task experiment developed by MacKenzie et al. [33], the performance of the mouse and the stylus with tablet was almost the same. This conclusion is similar to the results in this study. Nevertheless, here we can give a clearer and more precise comparison through the SH-Model (Fig.5.15).

Contrary to this study, for the one-dimensional pointing task, Epps concluded that the trackball performed better than the mouse [16]. The reason for the difference between Epps' paper and this study is that in their task design, the target was a rather small square. Therefore, the emphasis naturally slipped toward accuracy rather than speed. In that situation, the performance of the trackball would be better, as we discussed previously.

Interestingly, Accot and Zhai (1999) have classified the mouse and the pen with

tablet together in the same group in steering tasks [2]. Through this part, it is also possible to class these devices into one group. Future study which will focus on more devices will be instructive and will help us to classify different devices for different purposes.

Certainly, as a model with the *a posteriori* information (P_h), the prediction ability of the SH-Model will be weakened somewhat, which happens also to the ID_e model. However, this cannot affect the evaluation ability of the SH-Model for pointing tasks. As previously discussed in Sect. 1.1, the evaluation models which do not take into account different degrees of accuracy or which are based on the hypothesis that all the subjects always perform with the same error rate are not accurate for device evaluation. The real performance (in the task) has to be included. Moreover, in the SH-Model, we can fix the error rate and estimate the movement time at different levels. The application of a range of data in an experiment can also influence the design of the interface. The comprehensive ability of the SH-Model in prediction and evaluation has been discussed by [44] and all the conclusions are supported by AIC values.

5.3.5 Conclusions

We carried out the experiment presented in this section to make a first effort to evaluate different input devices with the SH-Model. Four commonly used computer input devices were compared and analyzed in the experiment with several indices.

From the analysis of the SH-Model and some other analyses, we can see that the best input device for the pointing task with a requirement for both speed and accuracy as used in this study is the mouse. The second is the pen with small tablet, the third is the pen with big tablet, and the last is the trackball. This order means that for a certain kind of system with human computer interaction, different input devices will affect the users' performance in different ways and it is necessary to evaluate the devices for the system and select the most suitable one for each case.

5.4 General Conclusions

The study in this chapter has the following significant points for the HCI applications.

First, we introduced a new method which applies the general information theory (self-information) and also the probability theory to established pointing performance models.

Second, it is the first attempt to observe the effects of system and human beings distinctly in one model.

Third, we have not only verified the advantages of the SH-Model, but we have also applied the powerful AIC statistical tool to the evaluation of human performance models for the first time in the human computer interaction area.

Fourth, in the device comparison based on the SH-Model, for each of the four devices, the SH-Model obtains the minimum AIC value, which means not only that the SH-Model is better than the other traditional models, but it also supports the idea that the SH-Model can be applied to other kinds of human computer interfaces.

Finally, the SH-Model can effectively evaluate input devices for pointing tasks which require both speed and accuracy. The coefficients estimated by the least square method in the model help us to understand the difference between different input devices.

We have established a new model, the SH-Model, for the pointing task in HCI, and testified its' superiority over the traditional models through AIC analysis. Thereafter, the devices evaluation in different tasks will also contribute to user interface design by affording reliable guidance. We believe we have shown that the SH-Model achieves this better than the traditional models. This agrees with the idea that the establishment of more reliable evaluation models is one of the more important tasks of researchers in the field of human-computer interaction.

Chapter 6

Influence of Colors on Pointing Tasks

Fitts' law has been applied to evaluate the pointing task widely. However, the quantitative effect of using color in the interfaces has not been discussed by literature. This chapter introduces the research on the color effects in pointing task by using Fitts' law as the evaluation method. Different colors and color demonstrating styles are applied in the experiments with similar design of the paradigm Fitts' law pointing task.

The experimental results show that when the subjects use mouse as the input device, there is no significant difference among the mean time of performance with the interface with colored target and white target. The results also reveal that the color demonstrating styles will bring no significant difference to pointing task when the mouse is applied, either.

However, when the tablet personal computer and pen are applied, the subjects without much experience in tablet personal computer usage need more time to perform the task with colored target than with the white target. Especially when the colors are changed randomly during the subjects tapping on the target, the difference is even more obvious. These results are testified by the Checking Experiment and Learning Effect Experiment across different groups of subjects.

6.1 Introduction

Fitts' law (both the ID and ID_e model) has been applied widely in human computer interaction even with the lack of theoretic supports since 1978 [9][36]. However, although the circumstances of human computer interaction are mostly colorful, all the existing researches of the pointing task related with Fitts' law, either theoretic or applications, are based on black and white interfaces. The effects of color have not been considered into the motor tasks related with Fitts' law researches. Therefore, the purpose of this study is to evaluate the pointing task with colored interfaces by Fitts' law models (the ID and ID_e model). This will be a new horizon for Fitts' law applications.

6.2   Color Experiment 1: on Interface with Fixed Colors

First we measured the effect of color in pointing task when the colors of the targets were fixed.

6.2.1   Subjects

Eleven volunteers, of different genders and ages (20 to 29 years, nine males and two females, average 22.3 years old) participated in a target pointing experiment. All the subjects' dominating hands were the right hands.

6.2.2   Apparatus

We used two sets of apparatus to test the influence of color on human performance in HCI interfaces. The first set of apparatus includes a Tablet Computer (FUJITSU FMV Stylistic, with the screen size of 21cm x 15.6cm, each pixel on the screen was 0.2055mm wide) and a plastic pen. The second set of apparatus includes a desktop personal computer (screen size: 43cm/17.0 Diagonal, pixel pitch: 0.264mmH x 0.263mmV, each pixel on the screen was 0.264 mm wide) and a mouse (Agiler AGM 6124X).

Calibration

We measured luminance and chromatic coordinates of the colors that we used as stimuli by spectral radiometer (CS-1000 made by Konica-Minolta). In measurement on tablet PC, we set the angle between the tablet PC and the detector of the radiometer in 66 degree from the horizontal surface, which was obtained as the average of 5 observers' experiments. Distance from the screen and the detector was 30 cm for the tablet PC and 50 cm for the PC (LCD), those were also obtained as the average of viewing distance of these observers.

Table 6.1 shows the luminance and chromatic coordinates of them presented by the PC (LCD) and the tablet PC with Standard Error of the Mean (SEM) and error rate. Figure 6.1 shows the chromatic coordinates plotted in CIE u-v coordinates. As shown in the figure, the colors presented by the tablet PC are much closer to the background white compared to the ones by the PC (LCD). Although red, green and blue shown in the figure are the most saturated colors that can be presented by the table PC, the difference of saturation between PC and tablet PC possibly affects the results of this research. We thought, however, that we should use the most saturated colors for each screen because the aim of this research is to estimate the effects of using colors to human

Table 6.1   Luminance and chromatic coordinates of stimuli used in PC (LCD) and tablet PC.

Color of Stimuli	Values	SEM	Error rate
Red			
PC (LCD) Lv	40.71	+-0.78	1.92
u'	0.4204	+-0.0011	0.27
v'	0.5271	+-0.0007	0.14
Tablet PC Lv	52.81	+-1.09	2.06
u'	0.2829	+-0.0021	0.76
v'	0.5144	+-0.0011	0.22
Green			
PC(LCD) Lv	86.45	+-1.78	2.06
u'	0.1403	+-0.0005	0.34
v'	0.5632	+-0.0013	0.23
Tablet PC Lv	69.64	+-1.12	1.61
u'	0.1979	+-0.0006	0.32
v'	0.5272	+-0.0038	0.72
Blue			
PC(LCD) Lv	16.94	+-0.34	2.01
u'	0.1295	+-0.0012	0.90
v'	0.2886	+-0.0021	0.72
Tablet PC Lv	44.98	+-0.98	2.71
u'	0.1996	+-0.0035	1.76
v'	0.4520	+-0.0054	1.18
Background White			
PC(LCD) Lv	147.24	+-2.47	1.68
u'	0.2060	+-0.0004	0.19
v'	0.4816	+-0.0018	0.37
Tablet PC Lv	93.41	+-0.41	0.44
u'	0.2191	+-0.0002	0.10
v'	0.4940	+-0.0004	0.07

performances on the PC and tablet PC. Also, because we expect that it is most likely to use the most saturated colors instead of desaturated (whitish) colors in the design of interface.

6.2.3   Procedure

Similar to the original Fitts experiment [18], participants did reciprocal pointing on a pair of vertical strip targets with the plastic pen or the mouse according to the experimenters' instructions. The widths (W) of the target were set at $W = 12$, 36, 72 pixels and the center to center distances (D) between the two strips were set at D =120, 360, 840 pixels, the consequent IDs of different $D - W$ combinations designed for the experiments mentioned in this chapter are shown by Table 6.2. The order of the

Table 6.2   Index of Difficulty of different $A - W$ combinations designed for experiments.

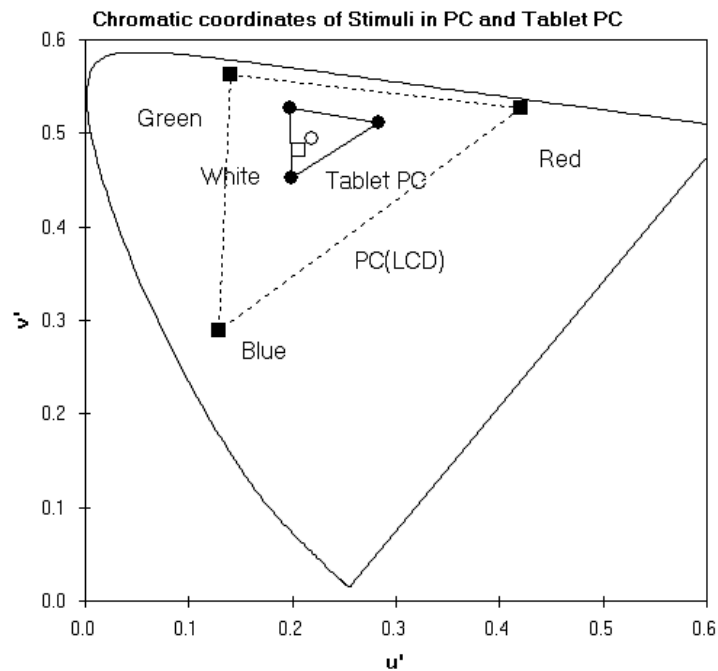| ID	D	W
3.46	120	12
2.12	120	36
1.42	120	72
4.95	360	12
3.46	360	36
2.58	360	72
6.15	840	12
4.60	840	36
3.66	840	72

Fig. 6.1   CIE u'v' chromatic coordinates of stimulus colors in PC(LCD) and tablet PC. Squares and circles denote coordinates of colors by PC(LCD) and by tablet PC, respectively. Open symbols denote background white for each screen.

9 widths and distance combinations was randomized. 12 trials were presented in each pair of targets, with the first tap excluded in analysis. If tapped on the outside of the target, an auditory signal was played.

During the task, in the two rectangles, the non-target one was black, while the colors of target rectangle changed from the regularly used white into one of the tricolors (red, blue and green) in one A-W combination. Once tapped, the positions of the target rectangle and the non-target rectangle would reverse. The appearance of the three colors
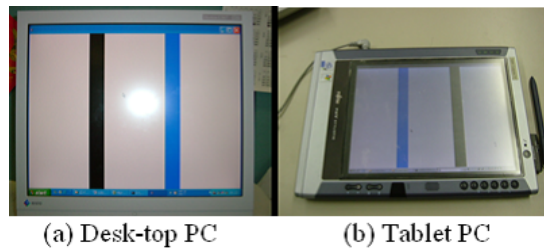
(a) Desk-top PC   (b) Tablet PC

Fig. 6.2   Experiment interfaces (when the target is blue)

Table 6.3   Error rates and accidental trials of Color Exp.1.

Apparatus and colors	Error rates	Accidental trials
Pen (mix)	0.016	5
Pen (red)	0.017	0
Pen (green)	0.019	2
Pen (blue)	0.013	3
Mouse (mix)	0.024	5
Mouse (red)	0.033	1
Mouse (green)	0.023	3
Mouse (blue)	0.018	1

was balanced by a Latin square sequence and set by the experimenter before the subject began to tap. Therefore the total number of the trials afforded for one subject to fulfill is 3 (colors) x 3 (distances) x 3 (widths) x 12 (trials) = 324.

The interface of the experiment tool was shown in Fig.6.2.

6.2.4   Error Rates

During the experiment, due to either the confusion of the participant, or instrument error, accidental clicks outside the general region of the target were registered. The information of the accidental trials and error rates in Color Experiment 1 were listed by Table 6.3.

6.3   Color Experiment 2: on Interface with Randomly Changing Colors

6.3.1   Subjects

The same subjects in Color Experiment 1 also took part in Color Experiment 2.

Table 6.4   Error rates and accidental trials of Color Exp. 2.

Apparatus and colors	Error rates	Accidental trials
Pen (mix)	0.019	18
Pen (red)	0.012	4
Pen (green)	0.027	7
Pen (blue)	0.018	7
Mouse (mix)	0.027	1
Mouse (red)	0.022	0
Mouse (green)	0.027	0
Mouse (blue)	0.032	1

6.3.2   Apparatus

We used the same apparatus of Color Experiment 1 in Color Experiment 2.

6.3.3   Procedure

The procedure of the task was the same with that of Color Experiment 1, except that the colors of the target rectangle were changed randomly during one section of the A and W combination, i.e. while doing the pointing task, the color of the target will change randomly without any warning after each pointing. The total number of the trials afforded for one subjects to fulfill is 3 (distances) x 3 (widths) x 12 (trials) = 108.

6.3.4   Error Rates

The error rates of the subjects with different apparatus and interfaces are shown by Table 6.4.

6.4   Non-color Experiment

To make comparison with the effects of whether using colors in the pointing task, we utilized some data from the experiments we had developed previously [56][29] and called the related experiments as Non-color experiment for conveniences.

One part of the Non-color Experiment using tablet personnel computer and pen was developed by the project of [56], where the two models, ID and ID_e, have both been discussed thoroughly. Fifteen volunteers, 5 female and 10 male, aged 20 to 36 years old, participated in it. We picked up the part of the data of the experiment of Zhai et al. [56] which had the similar procedure as in Color Experiment 1 for comparison. During the experiment of this part, the target rectangle was always white and the non-target rectangle was always black.

Table 6.5   Error rates and accidental trials of Non-Color Exp.

Input device	Error rates	Accidental trials
Pen	0.04	1
Mouse	0.02	0

The error rate of the task was 4%, and one accidental trial was excepted from the data for analysis.

The other part of the Non-color Experiment using regular personnel computer and mouse was developed by the project of [29]. Twelve subjects, of different genders and ages (3 female students and 9 male students, 21 to 32 years old, average 25) participated in the experiment. All the subjects dominant hands were the right hands. The procedure of this part was similar with that of Color Experiment 1. During the experiment of this part, the target rectangle was always white and the non-target rectangle was always black.

6.5   Checking Experiment

Because we used different subjects between the experiments with colored target and the experiments with the white target, to make the comparison reliable, we asked 5 subjects to perform all the experiments mentioned above (Color Exp. 1, Color Exp. 2 and Non-color Exp.) and organized the data as Checking Experiment[*1]. However, in the Checking Experiment, since our purpose was merely to test the reliability of the comparison results of Color Exp.1, Color Exp. 2 and Non-color Exp., we asked the subjects to perform the pointing task only under one $D-W$ combination (D=840 pixels, W=12 pixels). The reason for this choice is that with low level of difficulty, different colors may not incur much difference in performance, only with big task difficulty, the difference can be significant.

6.6   Results and Discussion

Since both the ID model and ID_e model of Fitts' law have obtained supports, we show the comparison results of applying both the two models.

[*1] The purpose of the Checking Exp. is to check whether the comparison results of the experiments with different subjects are identical. The subjects included in the Checking experiment need to perform all the experiments that had been performed by different sets of subjects.
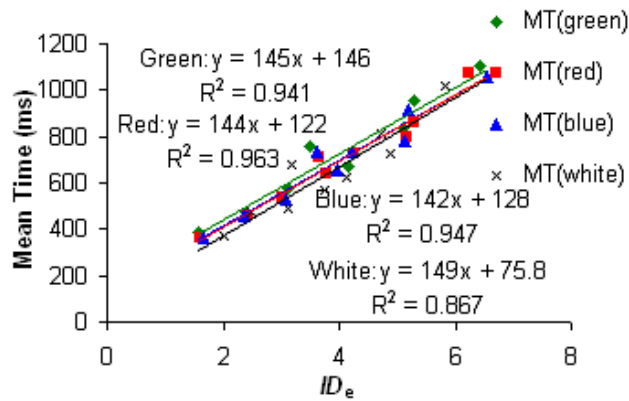
Fig. 6.3 $MT - ID_e$ regression lines with different colors of the target in Color Exp.1 and Non-color Exp. (pen)
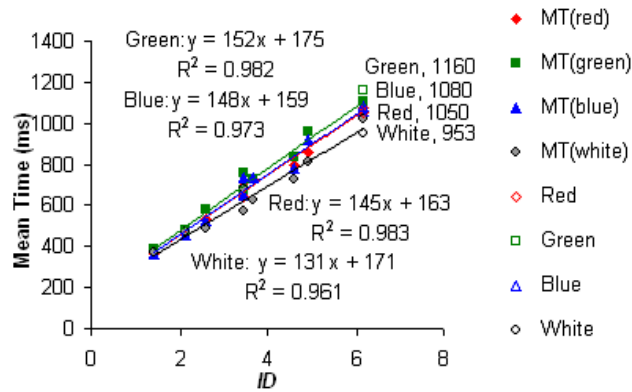

Fig. 6.4 $MT - ID$ regression lines with different colors of the target in Color Exp.1 and Non-color Exp. (pen) and the $MT - ID$ relationship with different colors of the target in Checking Exp. of Color Exp.1 and Non-color Exp. (pen) (The open symbols denote the data points of Checking Experiment.)

6.6.1   Difference Incurred by Different Colors

First, we performed ANOVA and found that there was no significant statistical difference among the mean time of the different colors afforded by the experiment if we used the pen as the input device.

However, from the Fitts' law analysis (based on ID and ID_e), we can still observe that the colored targets cost more time on average than the white target (Figs. 6.3 and 6.4).

Fig. 6.4 also shows the Checking Experiment results of Color Exp.1, through which we can test whether the sequences of the time cost by different colors for different groups of subjects are reliable. We did T-test to test the statistical significance of the difference
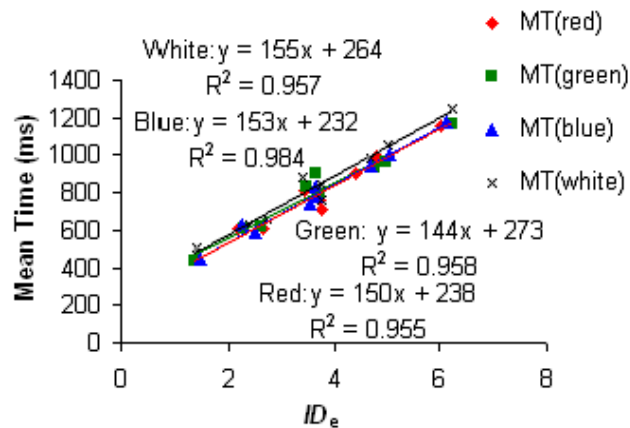
Fig. 6.5 $MT - ID_e$ regression lines with different colors of the target in Color Exp. 1 and Non-color Exp. (Mouse)

between different colors of the Checking Exp. There is significant difference between white and red ($P_{108}(t >= 2.23) < 0.05$), white and green ($P_{108}(t >= 5.13) < 0.0001$), white and blue ($P_{108}(t >= 3.49) < 0.001$), green and blue ($P_{108}(t >= 2.29) < 0.05$), red and green ($P_{108}(t >= 2.48)$

< 0.05), there is no significant difference between blue and red.

Therefore, the results of the Checking Exp. could support the comparison consequences of Color Exp.1 and Non-color Exp.: colored target costs more time than the white target. Nevertheless, the difference between the performance with different colors (red, green and blue) is not significant. Meanwhile, Table 6.3 and the results of error rates of Non-color Exp. show that during the task with different colors, the subjects made less errors with colors compared to the one without color[*2].

ANOVA shows that there was neither significant statistical difference among the mean time of the different colors afforded by the experiment if we used the mouse as the input device.

Fitts' law regression lines (Figs. 6.5 and 6.6) also show that using mouse and desktop PC, targets in different colors bring almost no difference in mean time even though we vary the difficulty of the task.

T-test results of the Checking Exp. only show significant difference between white and red ($P_{108}(t >= 2.19) < 0.05$), white and blue ($P_{108}(t >= 2.61)$

< 0.05), but not between blue and red, white and green, green and blue, red and green.

Therefore, after the Checking Exp., even there is tiny difference between the regression lines in Figs. 6.5 and 6.6 and the T-test analysis, we think that the difference

---

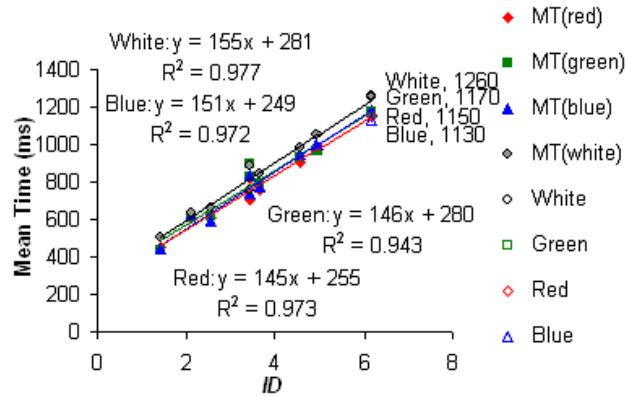[*2] This difference of error rates is included in the ID_e Model

Fig. 6.6 $MT - ID$ regression lines with different colors of the target in Color Exp. 1 and Non-color Exp. (Mouse) and the $MT - ID$ relationship with different colors of the target in the Checking Exp. Of Color Exp.1 and Non-color Exp. (Mouse)

brought by different colors of the target can be ignorable when we use the mouse as the input device.

6.6.2   Differences Incurred by the Color Demonstrating Styles

Next we test the difference brought by the color changing styles.

Analysis of the Data with Pen

According to ANOVA, there is no significant statistical difference among the mean time of different color demonstrating styles with the pen as the input device.

However, Figs. 6.7 and 6.8 coherently show that the targets in randomly changing colors cost more performance time than the targets in fixed color. When the difficulty increases, subjects need more time to track the targets in randomly changing colors. The difference in the data of performance time with the pen is obvious.

However, the Checking Exp. results contradict the comparison results of Color Exp. 1 and 2 (see Fig6.8).

We did T-test to test the statistical significance of the difference between different color changing styles of the Checking Exp. There is significant difference between the data of Random and Fixed task ($P_{218}(t >= 3.77) < 0.001$).

One potential reason for the conflicts between the Checking Exp. and Color Exp. 1 and 2 may be the different pointing task participation experience of the subjects in Color Exp. 1 and 2 and the subjects in the Checking Exp. In both Color Exp. 1 and
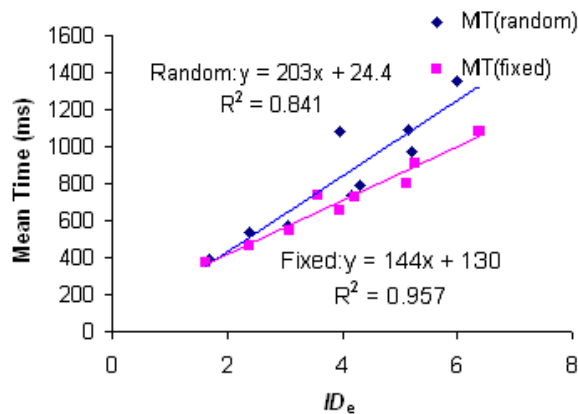
6.6 Results and Discussion

Fig. 6.7 $MT - ID_e$ regression lines of the mixed data of three colors in Color Exp. 1 (fixed) and Color Exp. 2 (random). (pen)
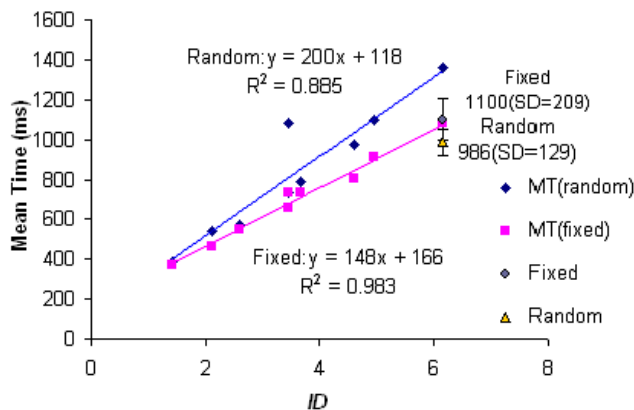

Fig. 6.8 $MT - ID$ regression lines of the mixed data of three colors in Color Exp. 1 (fixed) and Color Exp. 2 (random) (pen) and the $MT - ID$ relationship of the mixed data of three colors in the Checking Exp. Of Color Exp. 1 (fixed) and Color Exp. 2 (random). (pen)

2, five of the total 11 subjects had some experience of using the tablet PC and had taken part into similar pointing task executed in the lab previously, and the others were completely new for tablet PC performance. On the contrary, all the five subjects in the Checking Exp. had experience of the similar pointing task more than one hour, therefore, the conflicts between the Checking Exp. and Color Exp. 1 and 2 may imply some learning effects of the pointing task.

Therefore, we carried out an experiment to check the learning effects.

6.2.1.1 Learning Effect Experiment

Eight subjects without either pointing task participation experience or any experi-
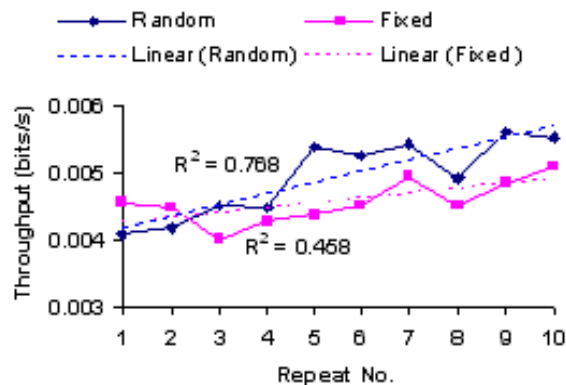
Fig. 6.9   Learning Effect Experiment results of throughput with the Random task of Group A and Fixed Color task of Group B in Day 1

ence of using the tablet PC were involved in the Learning Effect Experiment. Each of the subjects took part in the experiment 20 times altogether in two successive days (20 repeats). The subjects were divided into two groups with four persons in each group. The subjects in Group A performed 10 repeats of the Random task in the first day, and 10 repeats of the Fixed Color task in the second day. For the subjects in Group B, the sequence of the Random task and Fixed Color task was reversed. The time slot between two repeats was 1 hour. During each time, the task procedure was similar with that in Checking Exp. except that in the Random experiment part, the subjects would perform 6 trials for three times in one repeat, and in the Fixed Color experiment part, the subjects would perform each color for 6 trials in one repeat. The total trial number of one subject is 360=6x3x20. Altogether 21 accidents were exempted from analysis.

We checked the learning effect based on $Throughput$ (Equation 6.1), which is decided by both performance time and error rates according to ISO9241-9 standard[25].

$$Throughput = \frac{ID_e}{MT} \tag{6.1}$$

The result shown by Fig. 6.9 shows that the throughput was increased through more practices for both the two tasks, and for the Random task, the improvement was greater.

Through the data of Day 2, we know that after the practice of Day 1, the performance of both tasks intends to be more stable. No more learning effects can be clearly observed through Fig. 6.10.

Thus the results of the Learning Effect Exp. make it easy to explain the conflict happened in the Checking Exp. against Color Exp. 1 and Color Exp. 2. Since in the Checking Exp., all the subjects are experienced in pointing task and tablet PC usage,
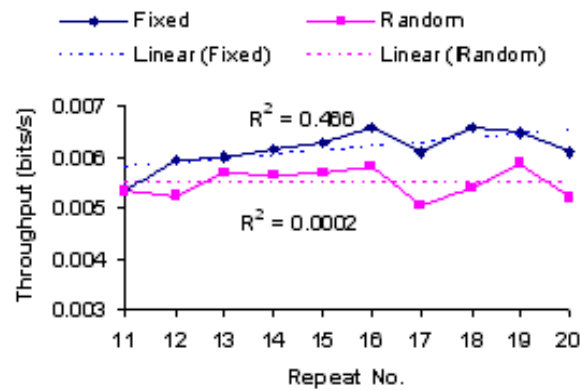
Fig. 6.10   Learning Effect Experiment results of throughput with the Fixed Color task of Group A and Random task of Group B in Day 2

their performance for the two kinds of tasks were close. On the contrary, in Color Exp. 1 and Color Exp. 2, those novice subjects need more time to fulfill the random task than the Fixed Color task. After training, it will cost similar time for the subjects to perform the Random task and the Fixed Color task.

Analysis of the Data with Mouse

According to ANOVA, there is no significant statistical difference among the mean time of different color demonstrating styles with mouse as the input device.

When the mouse was applied as the input device, the Checking Exp. results (T-test results show that there is no significant difference between the Random and Fixed color tasks) are identical with those obtained from Color Exp. 1 and 2, and the time difference of performing with Random and Fixed tasks is smaller than that of using pen as the input device (see Fig. 6.11 and Fig. 6.12), these can be because mouse is a familiar tool for all the subjects. Even though some of the subjects have no experience of using mouse to perform the pointing task designed in this study, their abundant experience of using mouse help them adapt to the task quite easy.

6.6.3   Differences Incurred by the Colored Interfaces

Finally we compared the results of the experiments with and without colors.

The ANOVA results show that there is no significant difference among the mean time of the colored interfaces and the black and white interfaces, either with fixed colors or randomly changing colors, pen or mouse.
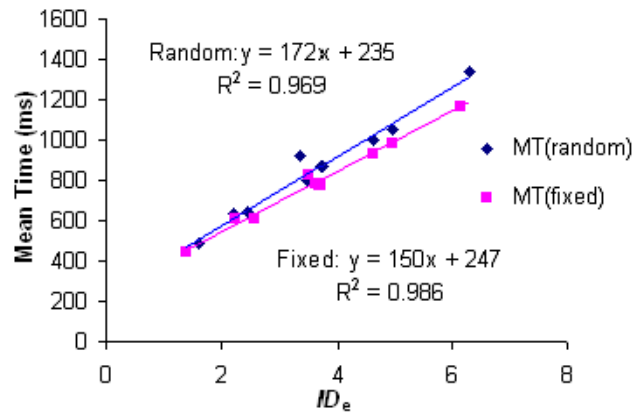
Fig. 6.11 $MT - ID_e$ regression lines of the mixed data of three colors in Color
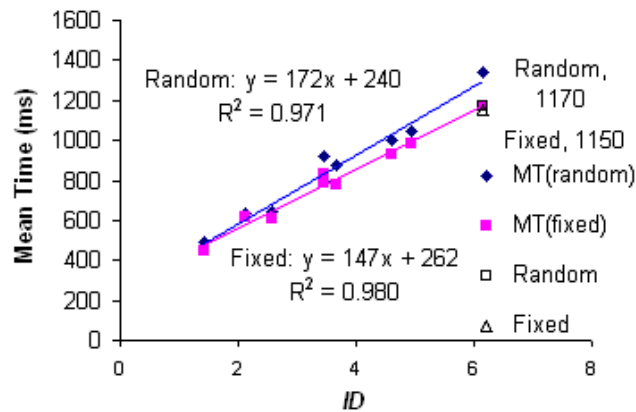Exp. 1 (fixed) and Color Exp. 2 (random). (Mouse)

Fig. 6.12 $MT - ID$ regression lines of the mixed data of three colors in Color
Exp. 1 (fixed) and Color Exp. 2 (random) (Mouse) and the $MT-ID$ relationship
of the mixed data of three colors in the Checking Exp. of Color Exp. 1 (fixed)
and Color Exp. 2 (random). (Mouse)

Nevertheless, Fig. 6.13 and Fig. 6.14 show that with the randomly changing colors
in the interface, subjects need more time to track the target with pen.

The results of Checking Exp. are identical with the comparison results (Fig. 6.14
). T-test results show that there is no significant difference between the Fixed Color
and Random tasks of the Checking Exp.

However, the difference in movement time is not clear when mouse was applied as
the input device (see Fig. 6.15 and Fig. 6.16).

The comparison results of the Checking Exp. are not the same with the comparison
between Color Exp. 2 and Non-color Exp. Nevertheless, the T-test shows no significant
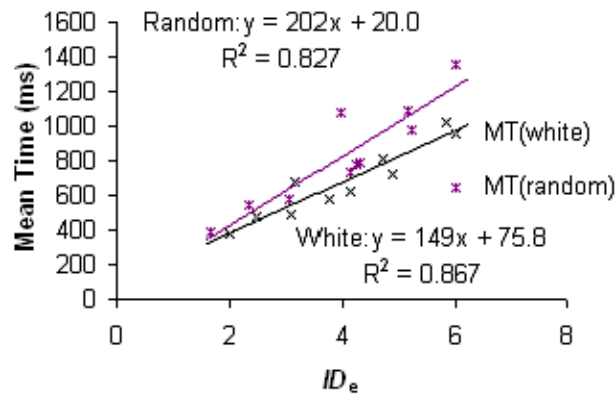
Fig. 6.13 $MT - ID_e$ regression lines of the interface with color (Color Exp. 2) and without color (Non-color Exp.) (pen)
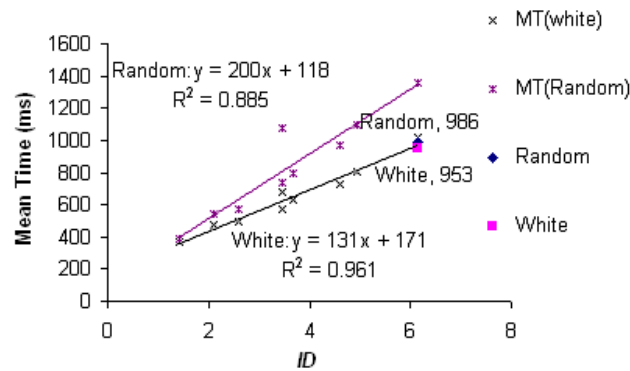

Fig. 6.14 $MT - ID$ regression lines of the interface with color (Color Exp. 2) and without color (Non-color Exp.) (pen) and the $MT - ID$ relationship of the checking Experiment of the interface with color (Color Exp. 2) and without color (Non-color Exp.) (pen)

difference exists between the Non-color and Randomly changing color, therefore, we may ignore the discrepancy.

When the colors of the target are fixed, there is no obvious difference among the colored interfaces and the non-color interface either with pen or mouse (see Fig. 6.17, Fig. 6.18, Fig. 6.19 and Fig. 6.20). Although the comparison results made by Color Exp. 1, Non-color Exp. and Checking Exp. are not the same, since in either group of comparison there is no significant statistical difference and the difference of direct observation of the regression lines is also tiny, the difference between the Non-color task and Fixed Color task can be ignored.
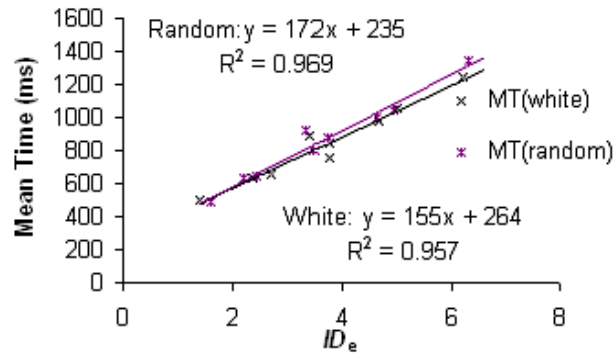
Fig. 6.15 $MT - ID_e$ regression lines of the interface with color (Color Exp. 2) and without color (Non-color Exp.) (Mouse)
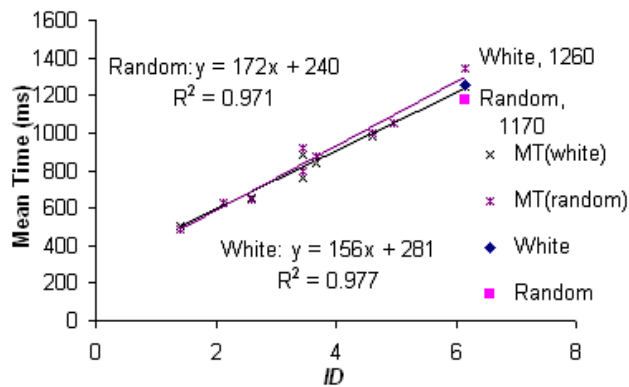

Fig. 6.16 $MT - ID$ regression lines of the interface with color (Color Exp. 2) and without color (Non-color Exp.) (Mouse) and the $MT - ID$ relationship of the Checking Exp. Of the interface with color (Color Exp. 2) and without color (Non-color Exp.) (Mouse)

6.7   Conclusions

In this chapter, we thoroughly compared the effects of whether using colors in the pointing tasks through Fitts' law and ANOVA. Three colors (red, green and blue) are applied in the experiments. Moreover, we also tested the effect of different color demonstrating styles during the pointing task (fixed colors and randomly changing colors). In case that there would be discrepancy brought by different subjects in the color and non-color experiments, we also carried out the Checking experiment for the three experiments (Color Exp. 1, Color Exp. 2 and Non-color Exp.). A learning effect experiment was executed to explain the relationship between the Checking Exp. and the main experiments.
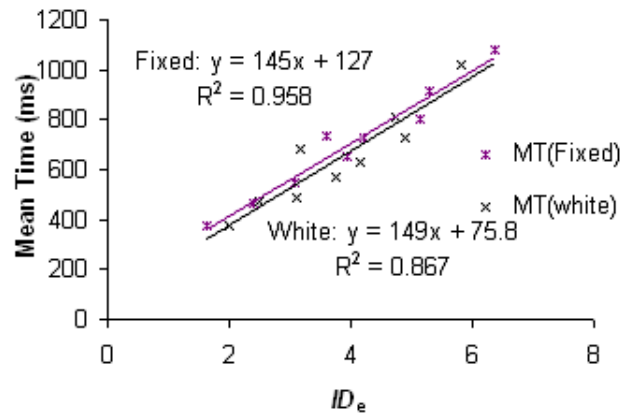
6.7 Conclusions

Fig. 6.17 $MT - ID_e$ regression lines of the interface with color (Color Exp. 1) and without color (Non-color Exp.) (pen)
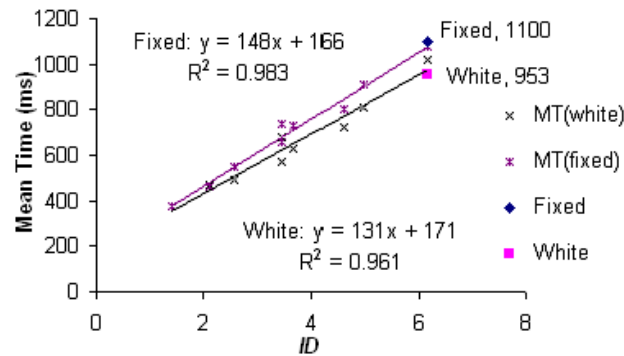

Fig. 6.18 $MT - ID$ regression lines of the interface with color (Color Exp. 1) and without color (Non-color Exp.) (pen) and the $MT - ID$ relationship of the Checking Exp. of the interface with color (Color Exp. 1) and without color (Non-color Exp.) (pen)

The great regression of the relationship of mean time and ID or ID_e demonstrates that Fitts' law is effective for the device evaluation of the interface with colored targets.

With the experimental data, it is not difficult to make conclusions on colors effects that:

1. in pointing task, different colors will not bring significant difference on subjects performance. However, when the subjects used pen, they need more time to perform the color task.

2. in the task with randomly changing colors, for the novice subjects, the performance is worse than in the task with fixed color. However, the difference is ignorable when the subjects used the mouse.
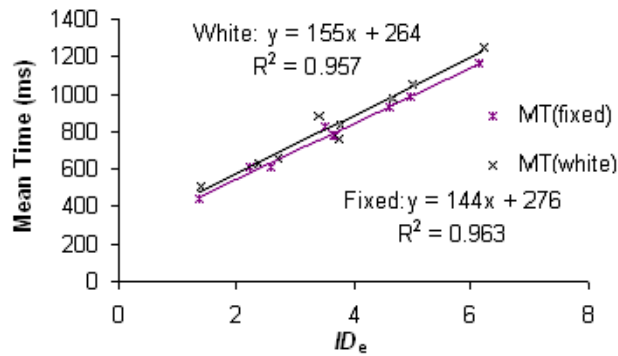
Fig. 6.19 $MT - ID_e$ regression lines of the interface with color (Color Exp. 1) and without color (Non-color Exp.) (Mouse)
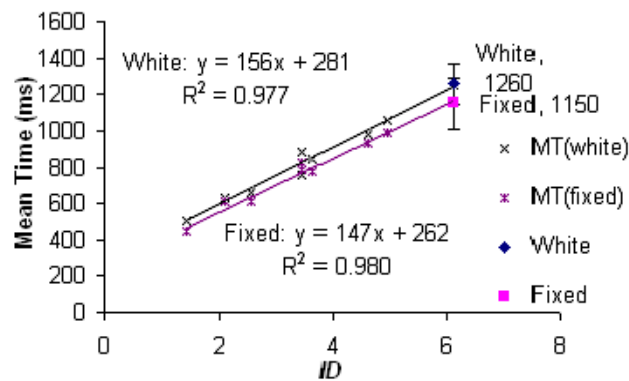

Fig. 6.20 $MT - ID$ regression lines of the interface with color (Color Exp. 1) and without color (Non-color Exp.) (Mouse) and the $MT - ID$ relationship of the Checking Exp. of the interface with color (Color Exp. 1) and without color (Non-color Exp.) (Mouse)

3. in the task with colors(red, green and blue) and without color(white), if the color is fixed when the target is tapped, the subjects performance keeps almost constant.

For the intervened effects of different colors and different input devices, we can conclude that the performance situation will be different, although differences are not big. When people use pen to tap the target with randomly changing colors, the performance time is a little longer than they tap the target without color or without changing color. The reason might be that when people use the mouse, there is friction, so the speed is not very fast and it is easier for subjects to adjust the performance power when the mouse approaches the target, no matter the target is white, colorful or even with randomly changing colors. However, using the pen, there is no friction to limit the speed of the pen, and sometimes it is difficult to change the accelerating power to adjust the

6.7 Conclusions

route of the pen. If the targets color is changed without previous warning, it is difficult for the subjects to make change according to their visual feedback. However, these phenomena were only apparent for the novice subjects. For the experienced subjects, the difference will be reduced.

These conclusions imply that even though there is no big difference for different color targets in the usual occasions, for some special situations, for instance, novice user, or tablet pen, the designers need to deliberatively consider the application of colors and the color demonstrating or changing styles.

This is a blank page.

Chapter 7

General Conclusions

This chapter addresses the main contributions of this dissertation and summarizes the research that was carried out.

7.1 Overview of Contributions

Models for pointing tasks are crucial for devices evaluation. The human subjective factors should never be neglected in this area. In the traditional model studies for pointing task, some psychological effects of human beings on the performances have not been emphasized enough. Therefore, there were problems in applying the traditional models. We have to restrict the subjects to perform with the "standard error rate", or ignore the difference in movement time incurred by different performance accuracy. These methods that we have utilized during the past to deal with the inaccuracy of the traditional models were either unrealistic, or unfair for device evaluation.

The two layers of speed-accuracy tradeoffs (the task layer and the subjective layer) in Fitts' law tasks had been studied in detail in Chapter 2. Then a thorough comparison between the two forms of Fitts' law (the task form (Equation 1.2) and the behavior form (Equation 1.4)) has been achieved. We defined an index of target utilization, $I_u = \log_2(4.133\sigma/W)$, to quantify the discrepancy between the nominal task precision and the actual behavior precision caused by the subjective layer of speed-accuracy tradeoff. A series of special controlled experiment were carried out for the observation of the effects of the two layers of speed-accuracy tradeoffs. The experimental results show that I_u is never constant in an experiment, even with the same instruction, except when an enforcement method is applied, as in Experiment SAT4. The overall I_u level can be influenced by the experimental instruction in the laboratory, or by performers' preference and task strategy in real world tasks. Through the study introduced in Chapter 2, theoretically, we know that the two layers of speed-accuracy tradeoffs, have different impact on task performance. Practically, the findings in Chapter 2 suggest that in order to accurately measure Fitts' law parameters, I_u should be kept as close to zero as possible and its variance should be kept as low as possible. This study also provides an empirical foundation for the application of W_e or its more aggressive and more

complete version, W_m, to adjust for I_u changes, in case I_u is highly varied. Although these adjustments consistently yield more logical Fitts' law parameter estimates, we should still be aware of the limitations and side effects of W_e or W_m, including reduced correlation between pointing time and index of difficulty within each operating strategy and their incomplete compensation for the subjective layer of speed-accuracy tradeoff.

The study on the speed-accuracy tradeoff problem and the comparison between the two forms of Fitts' law intrigued us to develope other researches about the Fitts' law models. We studied and compared two methods for calculating W_e. The results show that the Combined-coordinate-system Method is better than the Separate-coordinate-system Method, i.e., it is better to map all the abscissa data into one combined coordinate system to do the calculation, rather than divide the data into two separate groups according to the corresponding target positions. Thus the data shown by Chapter 3 affords a detailed and reliable comparison of the two methods of W_e calculation based on the information derived from the input hits with different target sizes. Since no research has been reported on this problem, the results of this study will be of great help for the applications of the effective target width in modeling for pointing tasks.

We utilized the results of Experiment SAT3 to analyze the information processing procedure of pointing tasks. By analyzing and comparing the estimates of the regression coefficients, we explored deep into the background theory of the coefficients of the models. This study mainly discussed the variation of b in Fitts' law models and related it to the information processing or transmission rate. The analytical results show that the ID model can not describe the information transmission or processing procedure of pointing tasks completely with different speed-accuracy tradeoffs. On the other hand, the ID_e model can describe the information transmission procedure during pointing tasks better. This study discovers a new factor which is essential for models evaluation, and lends support to the use of the ID_e model of Fitts' law to model the pointing performance in varied conditions.

Since we have concluded that it is impossible to completely reconcile the two layers of speed-accuracy tradeoffs by Fitts' law models, we thought about proposing an alternative model for solving the problems of the traditional Fitts' law models. Therefore, in Chapter 5, we proposed a new model named as SH-Model and demonstrated that the SH-Model is better than the traditional models based on AIC analysis. This is the first time that the human factors have been demonstrated separately from the system factors on pointing task performance. This model can also help us to escape from the restriction on the normal distribution of the input hits. Then we also testified the feasibility of using the SH-Model to evaluate and compare different input devices. The research results introduced in Chapter 5 will contribute to user interface design by affording reliable guidance.

Finally, considering the importance of different colors on performances and the lack of study on this topic, we thoroughly compared the effects of whether using colors in the pointing tasks through Fitts' law and ANOVA. The data afforded by Chapter 6 imply that even though there is no big difference for different color targets in the usual occasions, for some special situations, for instance, novice user, or tablet pen, the designers need to deliberatively consider the application of colors and the color demonstrating or changing styles. These conclusions will also be instructive for the future UI designs.

Comprehensively, in this thesis, first, we examined the feasibility of the traditional models for pointing tasks, and revealed the main features of the pointing task performance, especially the effects of the human subjective factors. Based on the detailed analysis of traditional models, we proposed a new model to demonstrate the human subjective factors in pointing tasks. This new model helps us to have a more accurate and comprehensive observation of the performance of pointing tasks with different devices, and hence is more reliable in devices evaluation. The results are important for human computer interface design and evaluation.

7.2   Future Directions

We aim to establish a model which can accurately include the human physiological and psychological information into a mathematical function. Such a model will be really reliable and applicable for human computer interaction input devices evaluation. It will also help to predict human performance. All these works can assist us to know whether the existing devices are appropriate for the performance, and if they are not suitable, what kind of modification in computer interfaces design should be adopted to enhance peoples performance efficiency without much labor.

For the present situations, Fitts' law is mostly applied in time-minimizing tasks as the pointing tasks described in this thesis. We also need to model the behavior of space-minimizing tasks, such as navigating hierarchical menus and tracing the outline of a shape. We had better also accomplish the work to unify different motor behaviors in one theoretical framework [37].

Since the technologies of interfaces between human and computer have been developed significantly, it is also necessary to carry out the model related research on the performance models application for new input techniques. Since currently most human potential has not been explored completely, the study on human performance models will be helpful to develop new devices or interfaces utilizing those unexplored body parts and give fair evaluation of those lately developed hardware or software.

7.3   Final Summary

Pointing task, together with some other basic performances, is of great importance in HCI studies. Developed from the area of motor control or motor behavior, modeling for pointing tasks helps us understand the performance. The understanding will be instructive for not only devices comparison, but also design guidance.

The studies introduced in this dissertation will contribute to the modeling work mainly from the aspect of considering human's factors or subjective factors in models' studies. These works will motivate much more explorations in human's factors, or subjective factors in modeling the pointing performance. The knowledge will be instructive for UI design comprehensively.

Acknowledgement

This is a blank page.

References

[1] Accot, J., Zhai, S., 1997. Beyond Fitts' law: models for trajectory-based HCI tasks. *Proceedings of CHI 1997: ACM Conference on Human Factors in Computing Systems. ACM*, 295-302.

[2] Accot, J., Zhai, S., 1999. Performance evaluation of input devices in trajectory-based tasks: an application of the steering law. *Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI '99*, 466-472.

[3] Accot, J., Zhai, S., 2002. More than dotting the I's – Foundations for crossing-based interfaces. *CHI Letters*, 4(1), 73-80.

[4] Accot, J., Zhai, S., 2003. Refining Fitts' law models for bivariate pointing. *Proceedings of CHI 2003, ACM Conference on Human Factors in Computing Systems, CHI Letters 5(1). ACM, Fort Lauderdale, Florida*, 193-200.

[5] Adam, J. J., 1992. The effects of objectives and constraints on motor control strategy in reciprocal aiming movements. *Journal of Motor Behavior 24*, 173-185.

[6] Akaike, H., 1974. A new look at the statistical model identification, *IEEE Trans. Auto. Control*, AC-19, 716-723.

[7] Akaike, H., 1981. Likelihood of a model and information criteria, *Journal of Econometrics*, Vol.16, 3-14.

[8] Burnham, K.P., Anderson, D.R., 2002 *Model selection and multimodel inference–a practical information-theoretic approach, second edition.* Springer-Verlag New York, Inc.

[9] Card, S. K., English, W. K., Burr, B. J., 1978. Evaluation of mouse, rate controlled isometric joystick, step keys and text keys for text selection on a CRT. *Ergonomics 21*, 601-613.

[10] Chan, R. B., and Childress, D. S., 1990. On a unifying noise-velocity relationship and information transmission in human-machine systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(5), 1125-1135.

[11] Crossman, E. R. F. W., 1956. The measurement of perceptual load in manual operations. *Unpublished doctoral dissertation*, Birmingham University.

[12] Crossman, E.R.F.W., 1960. The information capacity of the human motor system in pursuit tracking. *Quarterly Journal of Experimental Psychology*, 12, 1-16.

[13] Crossman, E. R. F. W., Goodeve, P. J., 1983. Feedback control of hand-movement and Fitts' law. Quarterly Journal of Experimental Psychology 35A, 251-278.

[14] Douglas, S.A., Mithal, A.K., 1997. *The ergonomics of computer pointing devices.* Springer-Verlag, New York.

[15] Douglas, S.A., Kirkpatrick, A.E., Mackenzie, I.S., 1999. Testing Pointing Device Performance and User Assessment with the ISO 9241, Part 9 Standard. *Proceedings of the CHI '99 Conference on Human Factors in Computing Systems.* New York: ACM, 215-222.

[16] Epps, B.W., 1986. Comparison of six cursor control devices based on Fitts' law models. *Proceedings of the human factors society-30th annual meeting.* 327-331.

[17] Everitt, B.S., 1998. *The Cambridge Dictionary of Statistics*, Cambridge University Press, London.

[18] Fitts, P.M., 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47, 381-391.

[19] Fitts, P. M., Peterson, J. R., 1964. Information capacity of discrete motor responses. *Journal of Experimental Psychology 67*, 103-112.

[20] Fitts, P.M. and Radford, B.K., 1966. Information capacity of discrete motor responses under different cognitive sets. *Journal of Experimental Psychology*, 71(4), 475-482.

[21] Guiard, Y., 1987. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of Motor Behavior*, 19, 486-517.

[22] Guiard, Y., Bourgeois, F., Mottet, D., Beaudouin-Lafon, M., 2001. Beyond the 10-bit barrier: Fitts' law in multi-scale electronic worlds. *Proceedings of People and Computers XV: Joint Conference of IHM 2001 and HCI 2001.* Springer, 573-587.

[23] Guiard, Y., Ferrand, T., 2002. Does the index of difficulty measure difficulty? A note on Fitts' aimed-movement paradigm. *Unpublished manuscript.*

[24] Isokoski, P., Raisamo, R., 2002. Speed-accuracy measures in a population of six mice. *Proceedings of APCHI2002: 5th Asia Pacific Conference on Computer Human Interaction.* Science Press, Beijing, China, 2002, 765-777.

[25] ISO9241-9: Ergonomic design for office work with visual display terminals (VDTs) – Part 9: Requirements for non-keyboard input devices. 2000. *International Standardization Organization.*

[26] Kabbash, P., Buxton, W., 1995. The "Prince" technique: Fitts' law and selection using area cursor. *Proceedings of CHI'95: ACM Conference on Human Factors in Computing Systems. ACM, Dever, Colorado*, 273-279.

[27] Keele, S. W., 1968. Movement control in skilled motor performance. *Psychological Bulletin 70*, 387-403.

[28] Kitagawa, G. and Gersch, W., 1996.*Smoothness Priors Analysis of Time Series*, Springer-Verlag, New York.

[29] Kong, J., Ren, X., Jiang, X., Takeda, F., 2004. Comparison of Four Input Devices for Pointing Task by Applying SH-Model, *Proceedings of WEC2004*. 168-174.

[30] Lewis, J. R., Kennedy, P. J., LaLomia, M. J., 1992. Improved typing-key layouts

References

for single-finger or stylus input. *IBM Technical Report TR 54.692.*

[31] MacKenzie, I. S., 1989. A note on the information-theoretic basis for Fitts' law. *Journal of Motor Behavior 21*, 323-330.

[32] MacKenzie, I. S., 1991. Fitts' law as a performance model in human-computer interaction. *Unpublished doctoral dissertation*, University of Toronto, Toronto, Canada.

[33] Mackenzie, I. S., Sellen, A. and Buxton, W., 1991. A comparison of input devices in elemental pointing and dragging tasks. *Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI '91*, 161-166.

[34] Mackenzie, I.S., 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 91-139.

[35] Mackenzie, I.S., Oniszczak, A., 1998. A comparison of three selection techniques for touchpads. *Human Factors in Computing Systems, CHI'98 Conference Proceedings.* ACM Press, New York, 336-343.

[36] MacKenzie, I. S., Zhang, S. X., 1999. The design and evaluation of a high-performance soft keyboard. *Proceedings of CHI'99: ACM Conference on Human Factors in Computing Systems.* 25-31.

[37] Mackenzie, I.S., 2003. Motor behaviour models for human-computer interaction. In *J.M.Carroll (Ed.) Toward a multidisciplinary science of human-computer interaction*, 27-54.

[38] McGuffin, M., Balakrishnan, R., 2002. Acquisition of expanding targets. *Proceedings of CHI 2002: ACM Conference on Human Factors in Computing Systems, CHI Letters 4(1). ACM*, 57-64.

[39] Meyer, E. D., Abrams, R. A., Kornblum, S., Wright, C. E., Smith, J. E. K., 1988. Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological Review 95*, 340-370.

[40] Newell, A., 1990. *Unified theories of cognition.* Cambridge, Massachusetts: Harvard University Press.

[41] Pew, R. W., 1966. Acquisition of hierarchical control over the temporal organization of a skill. *Journal of Experimental Psychology 71*, 764-771.

[42] Press, S.J., 1989. *Bayesian Statistics – Principles, Models, and Applications*, John Wiley & Sons, New York, 40.

[43] Press, W. H., Teukolsku, S. A., Vetterling, W. T., Flannery, B. P., 1992. *Numerical recipes in C — the art of scientific computing (2ed), Chapter 15 — Modeling of data.* Cambridge University Press.

[44] Ren, X., Kong, J. and Jiang, X., 2005. SH-Model: A Model Based on both System and Human Effects for Pointing Task Evaluation. *IPSJ Journal.* Vol. 46, No. 5, 1343-1353.

[45] Sakamoto, T., Ishiguro, M., and Kitagawa, G., 1986. *Akaike Information Criterion*

Statistics, D. Reidel, Dordrecht.

[46] Schwarz, G., 1978. Estimating the dimension of a model, *Annals of Statistics*, 6(2), 461-464.

[47] Shannon, C.E., 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423, 623-656.

[48] Surakka, V., Marko, I. and Isokoski, P., 2004. Gazing and frowning as a new human-computer interaction technique. *ACM Trans. Applied Perceptions Vol. 1, No. 1*, 40-56.

[49] Vinod, H.D. and Ullah, A., 1981. *Recent Advances in Regression Methods*, Marcel Dekker, New York.

[50] Welford, A. T., 1968. *Fundamentals of skill*. London.

[51] Woodworth, R. S., 1899. The accuracy of voluntary movement. *The Psychological Review, Series of Monograph Supplements 3*, 1-114.

[52] Zhai, S., Morimoto, C., Ihde, S., 1999. Manual and gaze input cascaded (MAGIC) pointing. *Proceedings of CHI'99: ACM Conference on Human Factors in Computing Systems. ACM Press*, 246-253.

[53] Zhai, S., Hunter, M., Smith, B. A., 2002. Performance optimization of virtual keyboards. *Human-Computer Interaction 17*, 89-129.

[54] Zhai, S., Conversy, S., Beaudouin-Lafon, M., and Guiard, Y., 2003. Human on-line response to target expansion. *Human Factors in Computing Systems, Proceedings of CHI'2003*, 177-184.

[55] Zhai, S., 2004. Characterizing computer input with Fitts' law parameters–the information and non-information aspects of pointing. *International Journal of Human-Computer Studies*, 61, 791-809.

[56] Zhai, S., Kong, J. and Ren, X., 2004. Speed-accuracy trade-off in Fitts' law tasks on the equivalency of actual and nominal pointing precision. *Int. Journal of Human Computer Studies* 61, 823-856.

Appendix A

Related Publications

A.1 Articles in or submitted to refereed journals

1. Kong, J. and Ren, X. (2005). The Optimal Calculation Method to Determine the Effective Target Width for the Application of Fitts' law. Submitted to *IEICE Trans. on Information and Systems*.

2. Kong, J. and Ren, X. (2005). The Information Processing Rate Analysis and Its Application in Fitts' law Models Comparison. Submitted to *IEICE Trans. on Information and Systems*.

3. Kong, J. and Ren, X. (2006). Effective Target Width Calculation and the Effects in Pointing Task. To appear in *Journal of Information Processing Society of Japan (IPSJ) (technical note)*, Vol. 47 No.5.

4. Kong, J., Ren, X. and Kyo, K. (2006). Comparison of Input Devices in Pointing Tasks through the Observation of the Human Effects—An Application of the SH-Model. Accepted by *Journal of Human Interface Society (HIS)*.

5. Kong, J., Ren, X. and Shinomori K. (2006). Influence of Colors on Pointing Tasks in Human Computer Interfaces (the title is subjected to revision). Conditionally accepted by *IEICE (the Institute of Electronics, Information and Communication Engineers) Trans. on Information & Systems*.

6. Ren, X., Kong, J. and Jiang, Q. (2005). SH-Model: A Model Based on both System and Human Effects for Pointing Task Evaluation, *Journal of Information Processing Society of Japan (IPSJ)*, Vol.46 No.5. pp 1343-1353.

7. Zhai, S., Kong, J. and Ren. X. (2004). Speed-accuracy trade-offs in Fitts' law tasks - on the equivalency of actual and nominal pointing precision, *Int. J. Human-Computer Studies*, Vol.61, No.6, pp 823-856.

8. Ren, X., Kong, J., (2004) A study of the Optimal Size for Pen-Input Character Boxes (in Chinese), *Information*, Vol. 7, No. 6, pp 747-754.

9. Kong, J., Ren, X., and Jiang, X.(2004) A new model describing the interaction between system and human for pointing tasks (in Chinese), *Information*, Vol. 7, No. 6, pp 737-746.

A.2 Articles in full paper refereed international conference proceedings

10. Kong, J., Ren, X., Shinomori, K. (2005). Influence of colors on pointing tasks in human computer interfaces, Proceedings of *The IASTED International Conference on HUMAN-COMPUTER INTERACTION, November 14-16, 2005 Phoenix, USA*, pp 7-12.

11. Kong, J., Ren, X. (2005). Comparison of Effective Target Width Calculation Methods for Pointing Task, *Proceedings of CIT2005 (The 5th International Conference on Computer and Information Technology, Shanghai, China, 21-23 September 2005).* IEEE Computer Society, pp 530-534.

12. Kong, J., Ren, X., Jiang, X., Takeda, F. (2004). Comparison of Four Input Devices for Pointing Task by Applying SH-Model, *Proceedings of WEC2004 (the World Engineers Convention, Shanghai, China, November 2-6, 2004). pp 168-174.* (This paper is awarded by the Foundation of C & C Promotion.)

13. Ren, X., Kong, J., Kato, T. (2004). A study of the size of pen-input character boxes for PDAs, *Proceedings of CIT2004 (The 4th International Conference on Computer and Information Technology, Wuhan, China, 14-16 September 2004)*, IEEE Computer Society, pp 115-122.

14. Ren, X., Kong, J., Jiang, Q., Liu, Z. (2004). A new model for different speed and accuracy requirements in pointing tasks, *Proceedings of CIT2004 (The 4th International Conference on Computer and Information Technology, Wuhan, China, 14-16 September 2004)*, IEEE Computer Society, pp 86-93.

15. Ren, X., Tamura, K., Kong, J. and Zhai, S. (2003). Candidate Display Styles in Japanese Input, *Proceedings of INTERACT 2003 - Bringing the Bits together (Ninth IFIP TC13 International Conference on Human-Computer Interaction, September 1-5, 2003 - Zurich, Switzerland)*, pp 868-871.

A.3 Articles in abstract refereed international conference proceedings

16. Kong, J., Ren, X. (2005). Comparing models by the information transmission capability expressed by the coefficient of the difficulty index in Fitts' law, *Proceedings of HCII2005 (11th International Conference on Human-Computer Interaction, Las Vegas, Nevada USA, July 22-27, 2005).*

17. Kong, J., and Ren, X. (2005). Information processing rate analysis in pointing tasks. *Proceedings of NEINE'05 (the International Conference on Next Era Information Networking, Shanghai, China, 4-5 September 2005)*, pp 442-446.

18. Kong, J., and Ren, X. (2004). Effective target width calculation and the effects on the speed and accuracy interaction in pointing task, *Proceedings of NEINE'04 (the International Conference on Next Era Information Networking, Kochi, Japan, 26-27 September 2004)*, pp 172-179.

19. Ren, X., Kong, J., and Kato, T. (2004). A study of the size of pen-input character Boxes for PDAs, in *Abstract Book of ICP2004 (the 28th International Congress of Psychology, August 8-13, 2004, in Beijing, China)*, pp 1249.

20. Kong, J., Ren, X., and Jiang, Q. (2004). SH-Model: Considering both systematic and human factors, *Abstract Book of ICP2004 (the 28th International Congress of Psychology, August 8-13, 2004, in Beijing, China)*, pp 154.

21. Tamura, K., Kong, J, and Ren, X. (2003). Japanese Input with Conversion Candidate Display Methods, In the Abridged *Proceedings of the 10th International Conference on Human-Computer Interaction (HCI International 2003, June 22-27, 2003, Crete, Greece)*,pp 105-106.

A.4   Articles in local conference proceedings

22. Kong, J., and Ren, X. (2005). Considering human factors in performance evaluation models in human computer interaction. *Proceedings of SJCIEE2005 (Tokushima, Japan, September 28, 2005)*.

23. Kong, J. and Ren, X. (2004). Modeling Human Computer Interaction for Pointing Task, *Proceedings of SJCIEE2004 (Tokushima, Japan, September 25, 2004)*.

24. Kong, J., Ren, X., and Jiang, X.(2003), SH-Model: Considering both Systematic and Human Factors, *Proceedings of International Academic Symposium — Fusion and Development on Science & Technology in the Twenty-First Century (December 20-21, Tokyo)*.