Regular Paper

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

Xiangshi Ren,† Jing Kong† and Xing-Qi Jiang††

As a well-known human performance model, Fitts' law, a tool for evaluating pointing devices, has been accepted and applied for a long time in the human computer interaction field. However, doubts now exist regarding its validity. One challenging job for those who are doing research on human performance models is to resolve the problem relating to input hits' distribution (i.e., spatial constraint). We developed a new 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.

1. Introduction

These days, the need for a reliable model for evaluating computer input tasks and for testing the feasibility and efficiency of pointing devices is greater than ever before. For example, the ubiquitous mouse may not be the best choice for some kinds of interfaces [1]. In some situations certain devices are not convenient. Before we decide the optimal input device or interface design for a system, we must first evaluate the performance of users. Fitts' law, which was proposed by Paul Fitts (1954) [9], is a powerful model which is used to evaluate the performance of rapid and aimed movements. Although many research projects have been carried out applying Fitts' law or its derivation versions to evaluate input devices [6),7),11),14),15)], its adequacy remains debatable.

One common form of Fitts' law is found in the following equation:

$$MT = a + b\log_2\left(\frac{2A}{W}\right) \tag{1}$$

where A is the amplitude, which indicates the distance between the centers of the two targets, W is the target width, which indicates the width of the different targets, MT is the movement time of task completion, a and b are regression coefficients (or constants).

One fatal shortcoming in Eq. 1 is that it may produce a negative index of difficulty in the situation where a small target amplitude to target width ratio exists in the interface. This is clearly illogical. A powerful variation of Fitts' law has been developed from the approximation of Shannon's theorem 17 ☆ to the direct analogy of it:

$$MT = a + b\log_2\left(\frac{A}{W} + 1\right) \tag{2}$$

In this paper, we call this variation the Shannon style model. Here the index of difficulty is defined as:

$$ID = \log_2\left(\frac{A}{W} + 1\right) \tag{3}$$

In Eqs. 2 and 3, movement amplitudes are analogous to "signals" and target widths are analogous to "noise". Still the analogy contains some doubtful points. One problem here is that Shannon's theory is established from a great deal of complex and strictly mathematical deduction based on the assumption that the signal is disturbed by the AWGN (Additive White Gaussian Noise) [19]. Thus the analogous requirement in motor tasks is a normal distribution of hits [9),14)]. Mackenzie noted that to keep target width analogous to noise, 96% of the hits must fall inside the target, i.e., 4% of the hits can be allowed to missed the target [14]. However, the hits falling into the target width may not follow the normal distribution so accurately and subjects cannot be expected to maintain a 96% accuracy rate while working "as fast and accurately as possible", a condition which must be fulfilled if Fitts' law is to remain valid.

Thus, a lot of work has been developed to

† Kochi University of Technology
†† Asahikawa University

☆ The formulation of Fitts' law was derived from the Shannon Theorem 17: $C = B\log_2(S/N + 1)$, C indicates the capacity of the channel, B indicates the bandwidth, S/N indicates the ratio of the average power of signals to noise [19].

modify Fitts' law without any changes or modifications to the experimental situation.

One of them is the method of using W_e (effective target width)[5] in the formulation which appeared soon after Fitts' law's first publication, as shown in Eq. 4.

$$MT = a + bID_e \tag{4}$$

In this paper, we refer to it as W_e model. ID_e is the effective index of difficulty which is defined as:

$$ID_e = \log_2\left(\frac{A}{W_e} + 1\right) \tag{5}$$

W_e is computed from the observed distribution of hit coordinates in users' trials: $W_e = 4.133SD$, where SD is the standard deviation of the hit coordinates.

However this method is doubted by others. Zhai (2002) is one example[21]. Indeed, there are still other researchers who prefer the Shannon style model, although the W_e method has become one of the ISO standards 9241-9[12]. One problem is that the method used in the W_e model to convert errors (hits outside the target width) into time may unfairly represent some specific applications, because for different applications, error is not important to the same degree. Weighing error against time requires concrete consideration for different situations. Moreover, the relationship between target width and error rate is still unclear, more study should be launched to see if it follows the normal distribution curve. Even if the error rate is modified to be 4% using the W_e model as a post hoc method, it is still not certain that the experimental data will follow the normal distribution.

Both the add hoc and the post hoc 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%. Thus, we established a new model based on the concept of temporal distribution which is not limited by spatial constraints.

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☆. The system effect

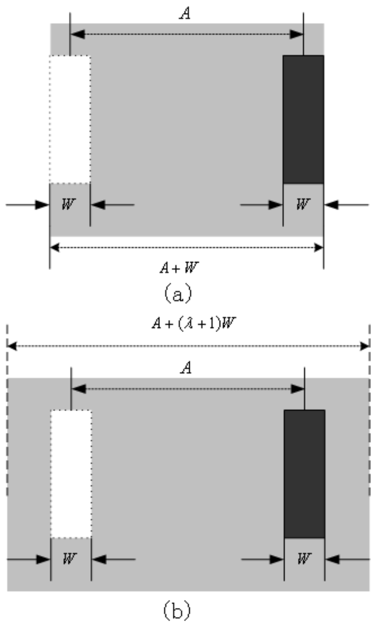

Fig. 1 The black target randomly appears within the gray area.

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, **Fig. 1** (a) shows a one-dimensional pointing input interface. Two rectangular targets are in the gray area. A is the amplitude of the centers of the two targets, and W is the target width. Assuming the black target randomly appears elsewhere within the gray area (i.e., the interval $A+W$), then the probability of the target falling into the gray area, P_s, can be represented as

$$P_s = \frac{W}{A + W} \tag{6}$$

Indeed, the target is not always located in the gray area of the interface. Assuming the target appears $A+(\lambda+1)W$ as indicated in Fig. 1 (b), where λ is a parameter, then the probability should be redefined as:

$$P_s = \frac{W}{A + (\lambda + 1)W} \tag{7}$$

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

☆ we use "effect" here rather than "factor" because "human factor" has been used with wider meaning.

$$SI_s = \log_2\left(\frac{1}{P_s}\right) = \log_2\left(\frac{A}{W} + \lambda + 1\right) \tag{8}$$

Here SI_s means *Self-Information*. The value of the parameter λ can be estimated by the minimum AIC method (described in Section 3). 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, Eq. 9 can be regarded as self-information depending on the probability of success reflecting the effects of human performance.

$$SI_h = \log_2\left(\frac{1}{P_h}\right) \tag{9}$$

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:

$$MT = a + bSI_s + cSI_h \tag{10}$$

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

In our view, in pointing tasks, experimental data do not always follow normal distribution. For experimental data that do not follow normal distribution, it may tend to incur biased estimation[20]. From observations of the movement time data to be analyzed in Sections 4 and 5, we can see that distribution of the data is close to lognormal distribution, that is, the data under logarithmic-transformation follow normal distribution[8]. 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. At the same time, to keep each part of the formulation identical, we took a logarithm of every part in Eq. 10.

Thus, we established the following new model:

$$\ln(MT) = a' + b'\ln(SI_s) + c'\ln(SI_h) \tag{11}$$

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 paper 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☆. Because Eq. 11 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 Eq. 11. 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.

$$\ln(MT_{real}) = a' + b'\ln(SI_s) + c'\ln(SI_h) + \varepsilon \tag{12}$$

Equation 12 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, Eqs. 11 and 12 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:

$$P_h = \frac{n}{m} \tag{13}$$

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

If we use Eq. 13 to calculate P_h, either of two

☆ Here error refers to the difference between the observation of movement time and the estimation of that calculated by corresponding equations.

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 11 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 [16]. The following equation gives the posterior mean of P_h.

$$P_h = \frac{n+1}{m+2} \qquad (14)$$

Omitting the error term ε, another form of Eq. 11 for computing the predictive value of MT is:

$$MT = e^{a'} SI_s^{b'} SI_h^{c'} \qquad (15)$$

3. 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 so that the predictive ability will also 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. Moreover, because the SH-Model is a non-linear model, we cannot apply this method to do model evaluation. Therefore, we have to find another model evaluation tool.

Another approach to model evaluation is to use information criteria (ICs), such as AIC and BIC [18]. AIC (Akaike's Information Criterion) is a criterion for model selection [2]. When a number of models are available, we have to select one as *the best* among the alternative models. Akaike's minimum AIC method [2),13),17], is developed for statistical model selection. This method can be interpreted from a maximization of the expected entropy of the predictive distribution approach [3]. It can be applied to comparisons for not only linear but also nonlinear models [4]. It is a better choice for us to compare the new model (SH-Model) with the traditional models (Shannon style 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:

$$AIC = -2M + 2N \qquad (16)$$

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 [13]. Overall, the model with the smallest AIC value can be regarded as the best one [2].

According to the characteristics of AIC, we used the original data for each combination of amplitude and target width rather than the average of the original data. For linear models, it makes no difference whether we use the original data or the average. However, for non-linear models such as Eqs. 11 and 15, original data and the average values give different results, i.e., the corresponding estimates may be different. According to statistical theory, estimation is more reliable if it is obtained from a larger sample. So it can be concluded that the estimated results are more reliable when we use the original data to establish the model. Another reason to use the original data is that the general formula of AIC is established based on the law of large numbers [8]. Therefore, we decided to use the minimum AIC method to evaluate the models.

4. Experiment 1: 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 [12].

4.1 Subjects

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

4.2 Apparatus

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

4.3 Design

The experiment was a 3 × 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) × 3 (target widths) × 29 (trials) × 12 (subjects) = 3,132.

4.4 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.

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 3,132 (complete data number) − 118 (accidental data number) = 3,014.

4.5 AIC Values of the Three Models from the Data of Experiment 1

To test the feasibility of our new model (Eqs. 11, 15), we applied the experimental data to the Shannon style model (Eq. 2) and the W_e model (Eq. 4) to see whether there was any difference in the effects of different models.

The results of the calculation are shown in **Table 1**. The model with the lowest AIC value will be regarded as the best one (see Section 3). P_h was calculated by each combination of A and W.

The corresponding AIC value of the Shannon style model (Eq. 2) is 38,927. The regression line is shown in **Fig. 2**.

The AIC value of the W_e model is 39,078, which is larger than that of the Shannon style model. The regression line is shown in **Fig. 3**.

To compare the effects of the new model with the traditional models, we set the parameter $\lambda = 0$ in the SH-Model (Eq. 15), then the model is determined as:

$$MT = e^{5.27} \left\{ \log_2 \left(\frac{A}{W} + 1 \right) \right\}^{0.64} \left\{ \log_2 \left(\frac{1}{P_h} \right) \right\}^{-0.03} \tag{17}$$

Table 1 AIC values of the three models with the Experiment 1 data.

Model	Formulation	AIC
Shannon Style Model	$MT = 197.39 + 75.3\log_2\left(\frac{A}{W}+1\right)$	38,927
We Model	$MT = -5.05 + 165.3\log_2\left(\frac{A}{We}+1\right)$	39,078
SH-Model	$MT = e^{5.27}\left\{\log_2\left(\frac{A}{W}+1\right)\right\}^{0.64}\left\{\log_2\left(\frac{1}{P_h}\right)\right\}^{-0.03}$	37,696

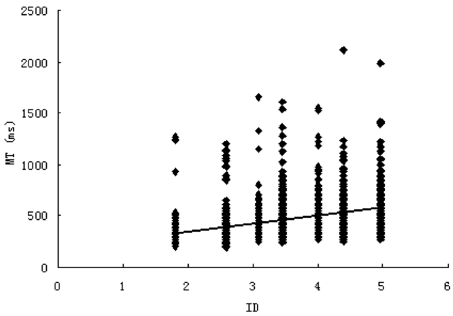

Fig. 2 Regression line of the Shannon style model.

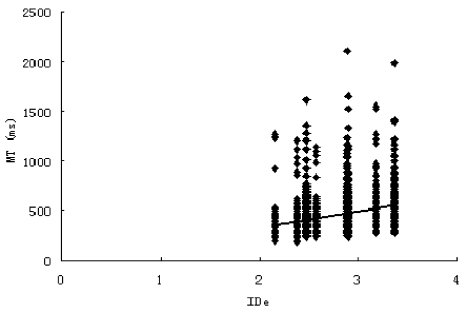

Fig. 3 Regression line of the W_e model.

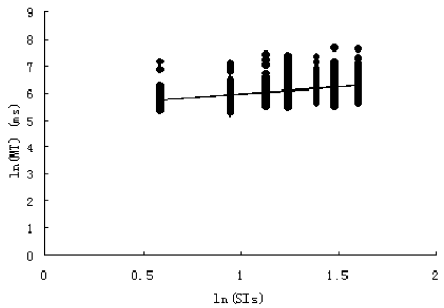

Fig. 4 Regression curved surface projection of the SH-Model ($\lambda = 0$).

The corresponding AIC value is 37,696 ☆.

In **Fig. 4**, 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 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 (37,696). 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 3.

5. Experiment 2: On Tablet PC

To make sure our models have universality and are not limited to PDA experimental data, we conducted an experiment which was the same as Fitts' reciprocal tapping paradigm to obtain the paradigm Fitts' law experimental data to see if it did indeed support our conclusions. Thus we performed an experiment on a Tablet PC.

5.1 Subjects

Twelve subjects (9 male and 3 female, aged from 21 to 38, mean = 26, all right handed) were tested in the experiment.

5.2 Apparatus

The tablet PC used in the experiment was a FUJITSU FMV STYLISTIC running Windows XP. The screen size was 21 cm × 15.6 cm, 1 pixel = 0.2055 mm. Experimental software was developed with Java.

☆ 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 [13].

Table 2 AIC values of the three models with the Experiment 2 data.

Model	Formulation	AIC
Shannon Style Model	$MT = 136.46 + 119.99\log_2\left(\frac{A}{W}+1\right)$	47,465
We Model	$MT = 53.52 + 153.05\log_2\left(\frac{A}{We}+1\right)$	47,859
SH-Model	$MT = e^{5.40}\left\{\log_2\left(\frac{A}{W}+1\right)\right\}^{0.71}\left\{\log_2\left(\frac{1}{P_h}\right)\right\}^{-0.00012}$	46,077

5.3 Design

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

- Target widths: 12, 36, 72 pixels (2.5, 7.4, 14.8 mm)
- Amplitudes, or distances between the center of targets: 120, 360, 840 pixels (24.7, 74.0, 172.6 mm)

Each subject performed the task in 12 trials in each of nine conditions. Each subject was instructed to repeat the experiment three times with different conditions, i.e., to tap the targets "as accurately as possible", "as accurately and fast as possible", and "as fast as possible". The goal was to make the subjects operate at a wide range of error levels☆. They were asked to take a rest before the next condition task. Targets were presented in random order to the subjects.

Because the actual time slot of the first trial was zero, the total number of data that we processed is 3 (repeating times) × 3 (targets amplitudes) × 3 (target widths) × 11 (trials) × 12 (subjects) = 3,564.

5.4 Procedure

In the experiment, the tablet PC was placed on a desk. Two rectangles (the height was 760 pixels) were shown on the screen. One was filled and the other was unfilled. Subjects sat down and tapped on the unfilled one. Upon contact, the rectangles would switch places and the subjects would again attempt to point to the unfilled rectangle.

Before the test, all subjects performed the same number of warm-up trials. During the experiment subjects were asked to point to the unfilled rectangle with a stylus pen.

Furthermore, in this experiment there were fewer accidental trials (28) than in Experiment 1 (118), and the total number of valid hits was 3,564 − 28 = 3,536.

☆ Fitts and Radford (1966) also manipulated three subjects' operational bias towards "accuracy, neutral, and speed" [10].

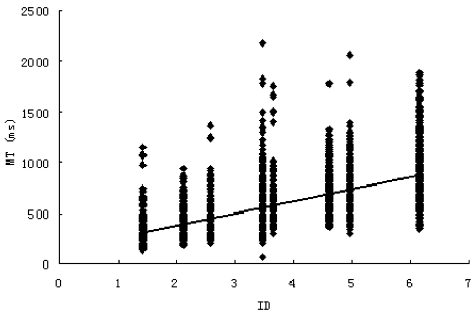

Fig. 5 Regression line of the Shannon style model.

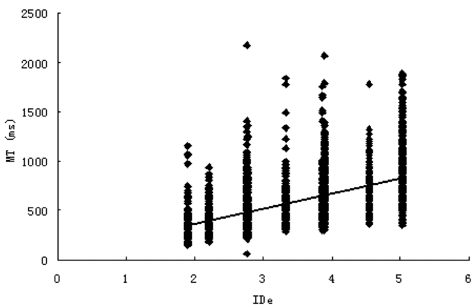

Fig. 6 Regression line of the We model.

The error rates of the three different conditions were 3.2%, 10% and 19.4% for the condition of "as fast as possible", "as accurately as possible" and "as fast as possible" respectively.

5.5 AIC Values of the Three Models from the Data of Experiment 2

The Shannon style model (Eq. 2), the W_e model (Eq. 4), and SH-Model (Eq. 11 or 15) applied from the experimental data and their AIC values are shown in **Table 2**.

The regression curves of the three models are shown in **Figs. 5** and **6**. **Figure 7** 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 Eq. 11 effectively.

From the above computation, we can conclude that with the data of Experiment 2, the SH-Model still has the lowest AIC (46,077) (see

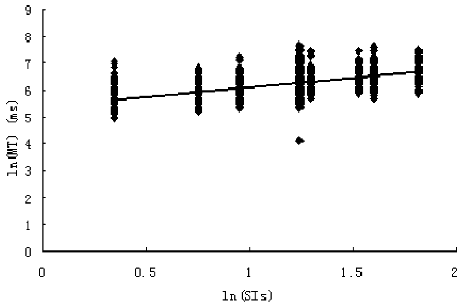

Fig. 7 Regression curved surface projection of the SH-Model ($\lambda = 0$).

Table 3 AIC values of SH-Model of different λs.

Experiment	$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
Experiment1	37,696	37,689	37,691	37,694
Experiment2	46,077	46,037	46,032	46,039

Table 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 3**.

6. Discussions

This study proposed an alternative model, SH-Model, for the development of the solution for Fitts' law's problems. Using the Shannon style 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 experimental data may not follow the normal distribution and/or keep the error rate of 4%. Using the W_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 Table 1 and Table 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 W_e model were even greater than the Shannon style model. One reason is that Experiment 1 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 2, the AIC value of the Shannon style model was still better than that of the W_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 W_e model cannot model the performance better than the Shannon style model. We calculated the error rate for Experiment 1 (26.11%) and Experiment 2 (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 AIC results showed that the Shannon style model was better than the W_e model even though this study does not focus on the comparison of the Shannon style model and W_e model. We note that while the results of comparing models might be different because of the different mathematical tools that we used to do the analysis. The SH-Model is a non-linear model, so we could not use the traditional methods for linear model evaluation (e.g., R^2) to do the analysis.

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 Experiment 1 data, $\lambda = 0$, the AIC value of the SH-Model is 37,696, and that of the Shannon style model is 38,927. There is only one more parameter in the SH-Model so to double the sum of 1 parameter plus one is 4, i.e., $2 \times$ (1 parameter $+ 1) = 4$. Then $38{,}927 - 37{,}696 = 1{,}231$ 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 Eq. 8 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 Eq. 9 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 W_e model (including the behavioral effects or accuracy into movement time). However, because the W_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: $\lambda = 0, 1, 2, 3$. The comparison of the results (see Table 3) shows that, for Experiment 1, $\lambda = 1$ produced the smallest AIC, for Experiment 2, $\lambda = 2$ 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 **Fig. 8**. We can explain this phenomenon in this way: when the subjects fulfilled the task, they would concentrate on the two targets as instructed. Simultaneously, they must perform the pointing task as quickly and accurately as possible. So

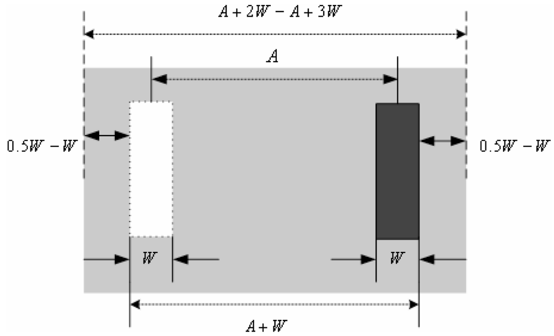

Fig. 8 The range of input hits when $\lambda = 1$ to 2.

they would certainly make mistakes, but the deviation of the hits' coordinate would not depart from the two targets too far.

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 Fig. 8. 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.

We analyzed the experimental data through the $ID - MT$ figures (shown in Figs. 2, 3 and 5, 6). 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 Fig. 4 and 7). 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 accu-

Table 4 Error distribution percentage.

	SH-Model		Shannon Style Model		*We* Model	
	Negative error percentage (%)	Positive error percentage (%)	Negative error percentage (%)	Positive error percentage (%)	Negative error percentage (%)	Positive error percentage (%)
Experiment 1	47.8	52.2	54.9	45.1	55.1	44.9
Experiment 2	52.8	47.2	58.9	41.1	59.7	40.3

rately and with the new model we can evaluate performance better [8),17)]. The estimated MT values can be observed from the trend lines in Figs. 2 to 7[☆].

7. Conclusions

Our goal in this study is to provide an alternative model for solving the problems of the traditional Fitts' 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.

This study has broader implications than just the solution of traditional modeling problems.

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. We used the minimum AIC value method to determine λ values on both a PDA and a Tablet PC. This is an innovative method for designing user interfaces, e.g., to determine the appropriate space between two objects displayed on a screen. Therefore, the significance of this research goes well beyond the new model proposed in this paper. More work on model evaluation can be developed through similar analysis.

In the experiments, we used the data derived from the use of a stylus pen to test the feasibility of different models. Our future work includes investigations of various pointing tasks and more pointing devices in order to clarifying the SH-Model's range of application.

☆ 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 Fig. 4 and Fig. 7 is different from Figs. 2, 3, 5 and 6, the simple comparison of the values of the Y axis will be meaningless.

References

1) Accot, J. and Zhai, S.: More than dotting the I's — Foundations for crossing-based interfaces, *CHI Letters*, Vol.4, No.1, pp.73–80 (2002).
2) Akaike, H.: A new look at the statistical model identification, *IEEE Trans. Auto. Control*, AC-19, pp.716–723 (1974).
3) Akaike, H.: Likelihood of a model and information criteria, *Journal of Econometrics*, Vol.16, pp.3–14 (1981).
4) Burnham, K.P. and Anderson, D.R.: *Model selection and multimodel inference — A practical information-theoretic approach*, second eidtion, Springer-Verlag New York, Inc. (2002).
5) Crossman, E.R.F.W.: The information capacity of the human motor system in pursuit tracking, *Quarterly Journal of Experimental Psychology*, Vol.12, pp.1–16 (1960).
6) Douglas, S.A. and Mithal, A.K.: *The ergonomics of computer pointing devices*, Springer-Verlag, New York (1997).
7) Douglas, S.A., Kirkpatrick, A.E. and Mackenzie, I.S.: Testing Pointing Device Performance and User Assessmentwith the ISO 9241, Part 9 Standard, *Proc. CHI '99 Conference on Human Factors in Computing Systems*, New York, ACM, pp.215–222 (1999).
8) Everitt, B.S.: *The Cambridge Dictionary of Statistics*, Cambridge University Press, London (1998).
9) Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement, *Journal of Experimental Psychology*, Vol.47, pp.381–391 (1954).
10) Fitts, P.M. and Radford, B.K.: Information capacity of discrete motor responses under different cognitive sets, *Journal of Experimental Psychology*, Vol.71, No.4, pp.475–482 (1966).
11) Isokoski, P. and Raisamo, R.: Speed-accuracy measures in a population of six mice, *Proc. APCHI2002: 5th Asia Pacific Conference on Computer Human Interaction*, Science Press, Beijing, China, pp.765–777 (2002).
12) ISO9241-9: Ergonomic design for office work

with visual display terminals (VDTs) — Part 9: Requirements for non-keyboard input devices, International Standardization Organization (2000).
13) Kitagawa, G. and Gersch, W.: *Smoothness Priors Analysis of Time Series*, Springer-Verlag, New York (1996).
14) Mackenzie, I.S.: Fitts' law as a research and design tool in human-computer interaction, *Human-Computer Interaction*, pp.91–139 (1992).
15) Mackenzie, I.S. and Oniszczak, A.: A comparison of three selection techniques for touchpads, *Human Factors in Computing Systems, CHI'98 Conference Proceedings*, ACM Press, New York, pp.336–343.
16) Press, S.J.: *Bayesian Statistics — Principles, Models, and Applications*, John Wiley & Sons, New York, p.40 (1989).
17) Sakamoto, T., Ishiguro, M. and Kitagawa, G.: *Akaike Information Criterion Statistics*, D. Reidel, Dordrecht (1986).
18) Schwarz, G.: Estimating the dimension of a model, *Annals of Statistics*, Vol.6, No.2, pp.461–464 (1978).
19) Shannon, C.E.: A Mathematical Theory of Communication, *The Bell System Technical Journal*, 27, pp.397–423, pp.623–656 (1948).
20) Vinod, H.D. and Ullah,: *A. Recent Advances in Regression Methods*, Marcel Dekker, New York (1981).
21) Zhai, S.: On the Validity of Throughput as a Characteristic of Computer Input, *IBM Research Report, RJ 10253*(*A0208-026*) (Aug. 21, 2002).

(Received April 1, 2004)
(Accepted February 1, 2005)
(Released April 20, 2005)

Xiangshi REN is an Associate Professor in the Department of Information Systems Engineering at Kochi University of Technology. He received a B.E. degree in electrical and communication engineering, and M.E. and Ph.D. degrees in information and communication engineering from Tokyo Denki University, Japan, in 1991, 1993, and 1996, respectively. After working for Tokyo Denki University, he has been at Kochi University of Technology since 2000. His research interests include all aspects of human-computer interaction, in particular, multimodal interactions, and user interface design and evaluation. He is a member of the IPSJ, the IEICE, and the Human Interface Society (all in Japan), the ACM, the ACM SIGCHI, the IEEE Computer Society, and the British HCI Group.

Jing KONG is a Ph.D. candidate in the Department of Information Systems Engineering in Kochi University of Technology, Japan. She received her bachelor's degree in communication engineering and master's degree in information and communication systems from Harbin Engineering University, China, in 1998 and 2003 respectively. Her current research interest includes human computer interaction modeling. She is a student member of ACM.

Xing-Qi JIANG is a Professor in the Department of Economics at Asahikawa University. He received his B.E. degree and M.E. degree in management engineering from Northeastern University, China, in 1982 and 1984 respectively, and Ph.D. degree in statistical science from the Graduate University for Advanced Studies, Japan, in 1993. His current research interests include Bayesian statistical modeling, estimation and identification of stochastic systems. He is a member of the Japan Statistical Society, the Institute of Systems, Control and Information Engineers, and ISBA.