SH-Model and Its Application in Human Interface Design

Xiangshi Ren*, Keizo Shinomori and Yoshimasa Kimura

Department of Information Systems Engineering
Kochi University of Technology
185 Miyanokuchi, Tosayamada-town, Kami-city, Kochi 782-8502

E-mail of corresponding author: *ren.xiangshi@kochi-tech.ac.jp

Abstract : We proposed a new model (SH-Model) to observe the human effects in pointing tasks. The SH-Model is approved as a reliable evaluation model. We also show how the SH-Model can be applied as an evaluation tool to observe the human effects. 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 study 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.

1. Introduction

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.

1.1 Fitts' law

One of the most famous models used for pointing task evaluation and prediction is Fitts' law [6], which was proposed by Fitts in 1954. One widely accepted version of this law is presented in the following form[10]:

$$MT = a + bID \qquad (1)$$

where MT is the movement time in which the subject moves a pointing device from one target's center to another target's center. a and b are empirically determined constants. ***ID*** is the index of difficulty for the task, which can be expressed as:

$$ID = \log_2(\frac{A}{W} + 1) \quad (2)$$

Here W is the target width and A is the distance between the centers of the two targets (Fig. 1). We call Eq(1) the ***ID*** model.

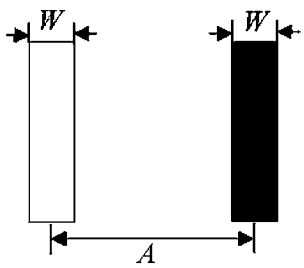


Figure 1: The two targets used in the pointing task experiment.

Nevertheless, Fitts' law formulations (Eq(1) and Eq(2)) are based on the analogy to the information capacity formulation. The use of the analogy without mathematical verification is questionable. Moreover, the performance accuracy is not considered. This is not reasonable for scientific device evaluation [17]. Therefore, some researchers prefer to use the "effective target width" (W_e) rather than the target width [4][16]. In this paper we call this kind of derivational version of Fitts' law the ID_e model.

The equation for the ID_e model is:

$$MT = a + b\log_2(\frac{A}{W_e} + 1) \quad (3)$$

The corresponding effective index of difficulty can be defined as:

$$ID_e = \log_2(\frac{A}{W_e} + 1) \quad (4)$$

However, the reliability of the ID_e model is also doubted by others [17]. Actually, the ***ID*** model and ID_e model have their respective advantages and problems in different application situations [18]. Thus, a model named the SH-Model was developed based on time series analysis and not limited to spatial constraint any more [14].

1.2 SH-Model

We established a new model, the SH-Model, based on time series analysis, while the traditional Fitts' law models are based on the concept of Shannon's capacity of channel and limited by the spatial distribution of the input hits.

In pointing tasks, the effects can be divided into two parts, the system effects and the human effects. Assuming that the system effects can be expressed by the variables of a pointing task (the amplitude between two targets and the target width), and the human effects can be indicated by the pointing success rate, Ren and colleagues established a model named the SH-Model [14]:

$$\ln(MT) = a + b\ln(SI_s) + c\ln(SI_h) \quad (5)$$

Another form of Eq(5) for computing the predictive value of MT is:

$$MT = e^a SI_s^b SI_h^c \quad (6)$$

In this model SI_s shows the effects of the system, and SI_h shows the effects of the human.

In Eq(5) and Eq(6),

$$SI_s = \log_2(\frac{1}{P_s}) = \log_2(\frac{A}{W} + \lambda + 1) \quad (7)$$

Here SI_s indicates the self-Information of the system. A and W indicate the target amplitude and target width respectively. In Eq(7), λ is not a parameter of the task, but a parameter determined by the experimenter. During the analysis, the experimenter can test different values of λ, and finally choose a comparatively better one by using the minimum AIC estimation method [14].

In Eq(5) and Eq(6),

$$SI_h = \log_2(\frac{1}{P_h}) \quad (8)$$

P_h indicates the probability that hits will fall into the target width. Thus, Eq(8) can be regarded as self-information depending on the probability of success, reflecting the effects of human performance.

The minimum AIC method [9][15] has proved that the SH-Model is better than the traditional models for pointing task evaluation [14][1].

2. 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 [7].

2.1 Subjects

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

2.2 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.

2.3 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, 9.6mm)
- 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.

2.4 AIC values of the three models from the data of Experiment 1

The results of the calculation are shown in Table 1.

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.

3. 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.

3.1 Subjects

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

3.2 Apparatus

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

3.3. Design

The experiment was a 3 x 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.

Figure 2: Experimental program

Because the actual time slot of the first trial was zero, the total number of data that we processed is 3(repeating times) x 3(targets amplitudes) x 3(target widths) x 11(trials) x 12(subjects) = 3564.

3.4 AIC values of the three models from the data of Experimental 2

The *ID* model, the ID_e model, and SH-Model are applied with the experimental data and their AIC values are shown in Table 2.

From the above computation, we can conclude that with the data of Experiment 2,

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

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

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

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

the SH-Model still has the lowest AIC (46077) (see Table 2). The experimental outcome gave powerful support to our previous conclusion.

4. Experiment 3: Devices Comparison

We executed Experiment 3 according to the Fitts' law paradigm experiment with four different devices to produce the data for device comparison.

4.1 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.

4.2 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.2), 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.3). The experimental program utilized the full-screen mode as shown in Fig.2.

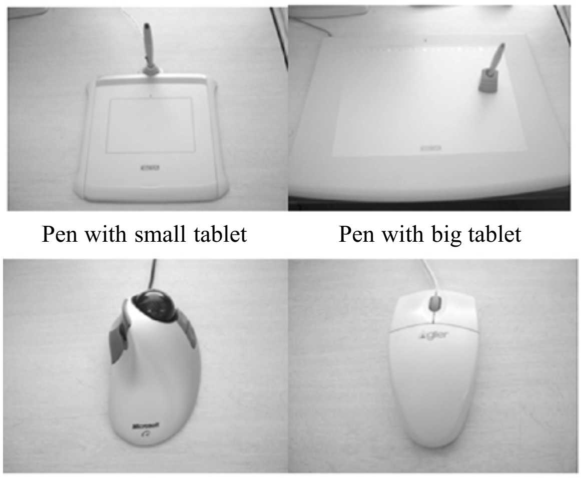
4.3 Design

The combinations of different width (W) and amplitude (A) are same as in Experiment 2. 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. Subjects performed the task with all the four devices.

4.4 Results

The AIC values of different models and different devices are listed in Table 3. It is obvious that the SH-Model can obtain the smallest AIC values for each of the four devices.

Fig.4 helps us to see what would happen if we use the ID model (Eq(1)). However, since Eq(1) cannot depict the complex interactions



Pen with small tablet　　Pen with big tablet

Trackball　　Mouse

Figure 3: Experimental input devices

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). 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 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 4. Note that although the estimation of *c* in the SH-Model is comparatively small as

Table 3: 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

an absolute value, the modifying quantity is still significant for the non-linear formulation (Eq(6)).

Fig.6[2] 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. 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. Discussions

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 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

Table 4: Coefficients in the SH-Model estimated by the least square method

Coeffcient	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

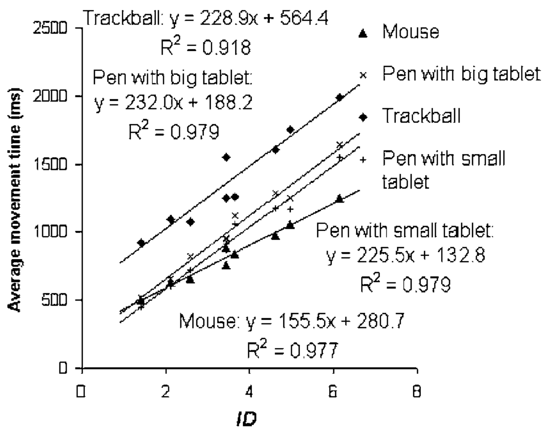


Figure 4: Regression lines for the four tasks using the *ID* model

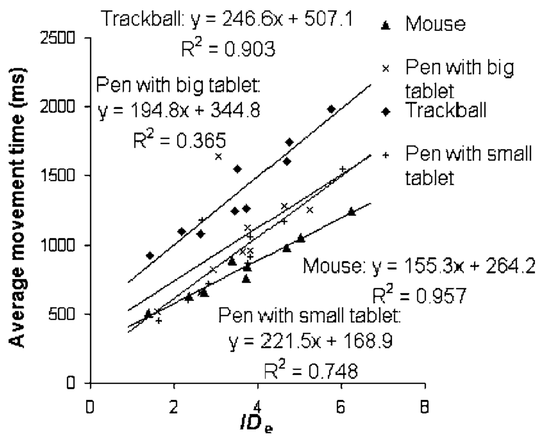


Figure 5: Regression lines for the four tasks using the ID_e model

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 3 and 4, the AIC results show that the new model is much better than the traditional ones.

Fig.4 and Fig.5 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, 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.6 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.6). 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.6 (a) shows that the mouse is the most suitable device for the pointing task because

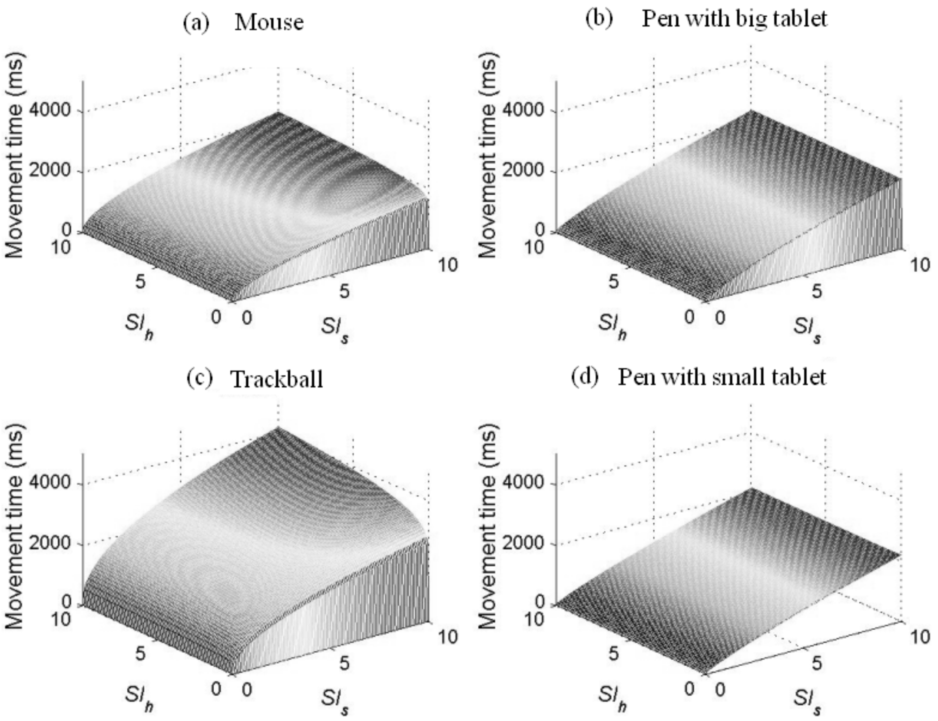


Figure 6: Regression curving surfaces with the SH-Model of the four tasks

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.

Comprehensively speaking, through applying the SH-Model (Table 4 and Fig.6), 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).

6. Conclusions

This study 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.

Acknowledgements

This research project was partially supported by a special grant-in-aid for graduate school enhancement (文部省大学院高度化推進特別経費) in Japan (No.755103411).

References

[1] Accot, J., Zhai, S. Performanceevaluation 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*, pp.466-472(1999).

[2] Akaike, H. A new look at the statistical model identification. *IEEE Trans. Auto. Control*, AC19, pp.716-723(1974).

[3] Card, S.K., English, W.K., and Burr, B. J. Evaluation of mouse, rate-controlled isometric joy-stick, step keys and text keys for text selection on a CRT. *Ergonomics*, 21(8), pp.601-613(1978).

[4] Crossman, E. R. F. W., Goodeve, P. J. Feedback control of hand-movement and Fitts' law (Original Report 1963). *Quarterly Journal of Experimental Psychology 35A*, pp.251-278(1983).

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

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

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

[8] Isokoski, P and Raisamo, R. Speed-accuracy measures in a population of six mice. *Proc. APCHI 2002*. Bejing, China: Science Press, pp.765-777(2002).

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

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

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

[12] Mackenzie, I.S. Fitts' law as a research and design tool in human-computer

interaction. ***Human-Computer Interaction* 7**, pp.91-139(1992).

[13] Douglas, S. A., Kirkpatrick, A. E. and Mackenzie, I. S. Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. ***Proceedings of the ACM Conference in Human Factors in Computing Systems -CHI '99***, pp.215-222(1999).

[14] Ren, X., Kong, J. and Jiang, X. SH-Model: A Model Based on both System and Human Effects for Pointing Task Evaluation. ***IPSJ Journal***. Vol. 46, No. 5, pp.1343-1353(2005). Online version: IPSJ Digital Courier, Vol.1, 193-203, http://www.jstage.jst.go.jp/article/ipsjdc/1/0/1_193/_article

[15] Sakamoto, T., Ishiguro, M., and Kitagawa, G. ***Akaike Information Criterion Statistics***, D. Reidel, Dordrecht(1986).

[16] Welford, A. T. ***Fundamentals of skill***. London(1968).

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

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

[1]AIC (Akaike's Information Criterion) is a model selection criterion which can be used for non-linear model evaluation[2]. It is defined on the basis of the maximum log-likelihood and the number of parameters to be estimated by the maximum likelihood method:

$$AIC = -2M + 2N \tag{9}$$

M is maximum log likelihood of the model, ***N*** is number of estimated parameters in the model. By using the least squares method, we can estimate the parameters in models, and get the AIC value of different models easily and then compare the effects of different models. The AIC method itself can reimburse the deviation brought by the parameters before it gives out the final results. Overall, the model with the smallest AIC value can be regarded as the best one.

[2]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.