International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:8, No:1, 2014

Evaluation of Fitts' Law Index of Difficulty Formulation for Screen Size Variations

Hidehiko Okada, Takayuki Akiba

Abstract—It is well-known as Fitts' law that the time for a user to point a target on a GUI screen can be modeled as a linear function of "index of difficulty (ID)." In this paper, the authors investigate whether the traditional ID formulation is appropriate independently of device screen sizes. Result of our experiment reveals that the ID formulation may not consistently capture actual difficulty: users' pointing performances are not consistent among pointing target variations of which index of difficulty are consistent. The term A/W may not be appropriate because the term causes the observed inconsistency. Based on this finding, the authors then evaluate the applicability of possible models other than Fitts' one. Multiple regression models are found to be able to appropriately represent the effects of target design variations. The authors next make an attempt to improve the definition of ID in Fitts' model. Our idea is to raise the size or the distance values depending on the screen size. The modified model is found to fit well to the users' pointing data, which supports the idea.

Keywords—Fitts' law, pointing device, small screen, touch user interface, usability.

I. Introduction

IT is well-known as Fitts' law that the time for a user to point a target can be modeled as a linear function of "index of difficulty (ID)," where ID is formulated as a function of the target size and distance [1,2].

$$t = a + b * ID \quad (1)$$

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

In (1) and (2), t is the pointing time, A is the amplitude (distance) to the target, W is the target size and a, b are constants that depend on experiment conditions. ID is larger as A is larger and/or W is smaller. Values of a and b in (1) are determined by sampling (A, W, t) data and applying the linear regression analysis to the data. Equation (2) shows that ID values are the same for (A, W) and (nA, nW) where $n > 0$.

This research is motivated by recent smart phones that employ touch user interfaces (UIs). Compared with other touch screen devices such as tablet PCs, mobile phones have smaller screens so that widgets on mobile phone screens are likely to be smaller. Widgets can be designed for devices with various screen sizes so that theoretical ID values in (2) are consistent on the same device: larger size & distance widgets and smaller size & distance widgets. If ID in (2) is an appropriate index of actual pointing difficulty independently of screen sizes, users' pointing performances on the same device are consistent among widget designs (A, W) and (nA, nW): note that a, b in (1) are constant (independent of ID) so that a, b must be the same for two data sets sampled with the two widget designs (A, W) and (nA, nW) on the same device. The aim of this research is to investigate whether the above is true: appropriateness of the ID formulation in (2) is evaluated from the viewpoint of dependency on screen sizes, by experiments with subjects.

Limitations of Fitts' law have been researched and extensions have been proposed. For example, an extension for 2D pointing tasks was proposed [3]. Our research aims at investigating possible limitations on screen size variations. A related research was previously reported [4]. They investigated how display size influenced pointing performances on a touch UI and reported that in large displays a fast and comparably accurate execution was chosen in contrast to a very inaccurate and time-consuming style in small displays. In their research the size of small screen was 6.5", and only a large screen touch UI device was utilized for user experiments: screen sizes were controlled by means of software program as virtual screens on the device display. In our research, the size of small screen is less than 3", and a commercial smaller-screen mobile device is utilized. Several researches have been reported on the usability of small-screen touch UIs, including, the comparison among devices with various screen sizes [5], the minimal target size in a pen-based system [6,7], effective pointing methods for small targets [7-10]. In these previous researches, the traditional ID formulation was utilized and its appropriateness was not in question.

II. Experiments

A. Test Tasks

Each subject were asked to point targets on a screen. A test task consisted of pointing two rectangle targets (target 1, 2) in a predefined order. An "attempt" was the two successive pointings of target 1 and 2, and a test task consisted of a predefined number set of the attempts. For each combination of experiment conditions, each subject was asked to perform a predefined set of the tasks. The pointing operations were logged for later analyses of pointing speed and accuracy.

B. Conditions

1) Devices: Three commercial devices were used in our experiment: two tablet PCs and a PDA which have a $10.2", 6.0", 2.8"$ touch screen respectively. The PDA was

The authors are with Kyoto Sangyo University, Kamigamo Motoyama, Kita Ku, Kyoto, 603-8555, Japan.
E-mail: hidehiko@cc.kyoto-su.ac.jp

TABLE I
TARGET SIZES AND DISTANCES

	Device S				Device M				Device L			
	Targets S		Targets L		Targets S		Targets L		Targets S		Targets L	
ID	W	A	W	A	W	A	W	A	W	A	W	A
2.00	4.00	12.00	12.00	36.00	8.53	25.60	25.60	76.80	14.61	43.82	43.82	131.45
2.15	3.80	13.07	11.40	39.20	8.11	27.87	24.32	83.62	13.88	47.71	41.63	143.12
2.30	3.60	14.13	10.80	42.39	7.68	30.14	23.04	90.42	13.15	51.59	39.44	154.77
2.45	3.40	15.18	10.20	45.53	7.25	32.38	21.76	97.14	12.41	55.42	37.24	166.26
2.60	3.20	16.20	9.60	48.60	6.83	34.56	20.48	103.69	11.68	59.16	35.05	177.47
2.75	3.00	17.18	9.00	51.54	6.40	36.65	19.20	109.96	10.95	62.74	32.86	188.21
2.90	2.80	18.10	8.40	54.30	5.97	38.61	17.92	115.84	10.22	66.09	30.67	198.27
3.05	2.60	18.93	7.80	56.80	5.55	40.39	16.64	121.17	9.49	69.13	28.48	207.40
3.20	2.40	19.66	7.20	58.97	5.12	41.93	15.36	125.79	8.76	71.77	26.29	215.31
3.35	2.20	20.23	6.60	60.70	4.69	43.16	14.08	129.49	8.03	73.88	24.10	221.63
3.50	2.00	20.63	6.00	61.88	4.27	44.01	12.80	132.02	7.30	75.32	21.91	225.96

(ID: bits, W&A: mm)

selected because several recent smart phones have such small touch screens (i.e., the PDA was used as a substitute for the recent smart phones). Screen sizes of the devices were relatively larger/middle/smaller. In this paper, these devices are denoted as devices L/M/S respectively. Subjects performed test tasks by using a stylus attached to each of the three devices.

2) Target sizes & distances: For each of the three devices, two sets of targets were designed so that ID values in (2) were consistent between the two sets. Targets in one of the two sets were designed with larger sizes and distances, and those in the other were designed with smaller ones. Specific designs of the two target sets are described in Subsection II.C. In this paper, these two target sets are denoted as targets L/S respectively.

3) Errors: Pointing speed and accuracy are usually a trade-off [11]. Subjects performed tasks under each of two error conditions: errors acceptable or not. In a test task where errors were acceptable, a subject could continue the task even if s/he made an error (mispointing), and the task was complete when the count of no-error attempts reached to a predefined number. In a condition where errors were not acceptable, a test task was cancelled by an error and the task was retried until the count of no-error attempts reached to a predefined number. The error condition was told to each subject before performing each task: s/he had to try a task more carefully in the "errors not acceptable" condition.

C. Pointing Target Designs

Table I shows the design of target sizes and distances. Values for the device M, L were determined as [values for the device S] $*$ [the ratio of screen sizes, i.e., $6.0/2.8$ for the device M and $10.2/2.8$ for the device L]. ID values were designed to range in $[2.00, 3.50]$ consistently among the devices S, M, L and the targets S, L. The size of target 1 was fixed to 6.0 mm, empirically found to be easy enough to point

Fig. 1. Screenshot for target pointing tasks

first, for all conditions. Positions of targets 1 and 2 were randomly determined for each attempt under the following two constraints.

- All areas of both targets were inside the device screen.
- Distance between center points of the two targets was a predefined value.

Fig. 1 shows a screenshot of targets 1 and 2 for the device M and the targets L. The targets 1 and 2 are the black and white rectangles respectively (the target colors were consistent for all the devices). The two targets were shown at the same time, and each subject was asked to find both targets before s/he pointed the target 1. This was because visual search time should not be included in the pointing time interval. After an attempt of pointing targets 1 and 2, new targets were shown for the next attempt.

D. Method of Experiment

Condition combinations were 12 in total: [the devices S, M, L] $*$ [the targets S, L] $*$ [errors "acceptable," "not acceptable"]. Each subject was asked to perform four trials of a task under each of the 12 condition combinations.

The number of attempts in a task trial was 11 (of which $ID = \{2.00, 2.15, ..., 3.50\}$ shown in Table I) for the "errors not acceptable" condition: none of the 11 attempts had to be an error. For the "errors acceptable" condition, a task trial included 11 successful attempts for the 11 IDs respectively in Table I and $0+$ error attempts.

Each subject first performed a training task trial under each of the 12 condition combinations (thus, 12 training trials), and then performed tasks in a random order of the 12 condition combinations. The order of the 11 IDs in a trial was also randomized for each trial.

E. Subjects

Twelve university graduate or undergraduate students participated in the experiment, but 3 of the 12 subjects could for the devices S and L only due to the experiment schedule. Thus, users' pointing log data set (A, W, t) were collected with 12 subjects for the devices S and L and 9 subjects for the device M. They were all novices in using devices with touch-by-stylus UIs, but they had no trouble in performing test tasks after the 12 training trials.

F. Logging Pointing Operations

The following data was recorded for each pointing (each tap by a stylus) into log files.

- Target: 1 or 2
- Target position: (x, y) values
- Target width and height: pixels
- Tapped position: (x, y) values
- Tap time: msec
- Error: Yes or No

The tapped position and the tap time were logged when the stylus was landed on the screen, and the pointing was judged as an error or not based on the tapped position. No attempt was observed for which the stylus was landed on the target 1, moved into the target 2 and left off.

III. Data Analyses and Findings

Pointing speed and accuracy were measured by throughput [12] and error rate respectively. In this research, t is the interval from the target 1 tap time to the target 2 tap time, A is the Euclid distance between the tapped points for targets 1 and 2, and W is the target width (= height). Throughput is defined as ID/t in (1) and (2). (ID, t) could be observed for each attempt, so a throughput value could also be obtained for each attempt. To measure pointing accuracy, error rate was defined.

$$Error\ rate = \frac{\#error\ attempts\ in\ a\ task\ trial}{\#total\ attempts\ in\ the\ trial} \quad (3)$$

Error rate could be calculated for only the condition "errors acceptable" because the data under the condition "errors not acceptable" didn't include any error attempt (if an error was occurred in a trial under the condition "errors not acceptable," the trial was cancelled and retried).

Mean and standard deviation (SD) values of the throughput and the error rate were calculated to compare user performances on targets S to those on targets L, for each device. Mean and SD values of the throughput were calculated

TABLE II
Mean and SD Values of Throughput (Bit/Sec)

		Device S		Device M		Device L	
		Targets S	Targets L	Targets S	Targets L	Targets S	Targets L
Acceptable	Mean	5.73	5.73	5.86	5.76	5.52	4.76
	SD	1.37	1.14	1.30	1.80	1.34	0.87
Not acceptable	Mean	5.15	5.57	5.69	5.63	5.32	4.60
	SD	1.20	1.21	1.30	1.78	1.23	0.97

TABLE III
Mean and SD Values of Error Rate (%)

		Device S		Device M		Device L	
		Targets S	Targets L	Targets S	Targets L	Targets S	Targets L
Acceptable	Mean	11.23	0.52	0.93	0.69	1.56	0.52
	SD	10.35	2.04	2.66	2.34	3.29	2.04

from the data of $\{tp(s, k, a)\}$ for all of the subjects, the task trials and the attempts in a task: $tp(s, k, a)$ denotes the throughput value for the s-th subject, k-th task and the a-th attempt in the k-th task by the s-th subject. Mean and SD values of the error rate were calculated from the data of $\{er(s, k)\}$ for all of the subject and the task trials: $er(s, k)$ denotes the error rate value in the k-th task by the s-th subject.

In addition, it was tested by t-test whether there was a significant difference between population mean values of throughput and error rate for the conditions of targets S&L.

It should be noted that error attempts were included in the data under the condition "errors acceptable." Error attempts might be faster (of larger throughput values) than successful attempts. In the following of this chapter, throughput values were calculated with both of successful and error attempt data.

Table II shows mean and SD values of the throughput, and Table III shows those of the error rate.

Tables IV&V show t-test results for throughput and error rate respectively. In Tables IV&V, **-marked t-scores are those with $p < 0.01$, and non-marked t-scores are those with $p > 0.05$.

These tables revealed the followings.

- On the device L, subjects could point targets S significantly faster than targets L, but on the devices S&M they couldn't. Instead, on the device S, they could point targets L significantly faster than targets S under the "errors not acceptable" condition. This result indicates that, even though ID values by (2) are designed consistently among targets S&L, users' pointing speeds will not be consistent: faster for larger/smaller size&distance widgets on smaller/larger screen devices, respectively.
- On the devices M&L, no significant difference was observed in the pointing accuracy among targets S&L, but on the device S subjects could point targets L significantly more accurately than targets S. This result indicates that, even though ID values by (2) are designed consistently among targets S&L, users' pointing accuracies will not be consistent too: more accurate for larger size&distance widgets on smaller screen devices.

TABLE IV
T-TEST FOR THROUGHPUT

	Device S		Device M		Device L	
	Acceptable	Not acceptable	Acceptable	Not acceptable	Acceptable	Not acceptable
Targets S/L	$t=3.65\times10^{-3}$	t=-5.74**	t=0.875	t=0.514	t=11.04**	t=10.66**

TABLE V
T-TEST FOR THROUGHPUT

	Acceptable		
	Device S	Device M	Device L
Targets S/L	t=7.03**	t=0.393	t=1.87

Thus, it is found that the ID definition in (2) may not consistently capture actual pointing difficulty among target designs. The result of our experiment shows that, on a smaller/larger screen, targets with smaller/larger sizes&distances are actually more difficult to point than those with larger/smaller ones. A/W in (2) is not appropriate in terms of screen size variations because the term caused the observed inconsistency.

In the following two sections, the authors investigate better formulation of ID. Based on the finding in this section, the authors first evaluate the applicability of possible models other than the Fitts' one, and then make an attempt to improve the definition of ID in the Fitts' model, i.e., (2).

IV. EVALUATION OF MULTIPLE REGRESSION MODEL

Our finding in Section III implies that a model in which A and W independently affect the pointing time t may capture the effect of device screen size more appropriately: such a model may be able to represent that A (W) affects more than W (A) where device screen sizes are larger (smaller). For example, a power function model was previously proposed [13].

$$t = a * A^b * W^c \tag{4}$$

$$log_2 t = a + b * log_2 A + c * log_2 W \tag{5}$$

The following model has also been investigated [2].

$$t = a + b * log_2 A + c * log_2 W \tag{6}$$

Based on these previous researches, the authors evaluate multiple regression models in (6) and (5) by applying the models to the data collected by user experiments in our research.

By normalizing the data of t, $log_2 t$, $log_2 A$ and $log_2 W$ in (6) and (5) respectively, a becomes 0 and the value of b can be directly compared with the value of c.

$$t' = b * (log_2 A)' + c * (log_2 W)' \tag{7}$$

$$log_2 t' = b * (log_2 A)' + c * (log_2 W)' \tag{8}$$

TABLE VI
VALUES OF B AND C IN (7)

(i) Errors acceptable

	Device L		Device S	
	Targets L	Targets S	Targets L	Targets S
b	0.13	0.23	0.10	0.04
c	-0.33	-0.18	-0.39	-0.37

(ii) Errors not acceptable

	Device L		Device S	
	Targets L	Targets S	Targets L	Targets S
b	0.004	0.21	0.16	0.14
c	-0.42	-0.23	-0.35	-0.32

TABLE VII
VALUES OF B AND C IN (8)

(i) Errors acceptable

	Device L		Device S	
	Targets L	Targets S	Targets L	Targets S
b	0.16	0.26	0.19	0.14
c	-0.30	-0.15	-0.33	-0.32

(ii) Errors not acceptable

	Device L		Device S	
	Targets L	Targets S	Targets L	Targets S
b	0.07	0.27	0.22	0.25
c	-0.36	-0.16	-0.28	-0.25

In (7) and (8), t', $(log_2 t)'$, $(log_2 A)'$ and $(log_2 W)'$ are normalized ones (i.e., x' denotes the normalized values of x). Each of t', $(log_2 t)'$, $(log_2 A)'$ and $(log_2 W)'$ follows $N(0, 1)$.

Table VI shows values of b and c for the model in (7) obtained by applying the multiple regression analysis to the data of $(t', (log_2 A)', (log_2 W)')$.

Table VII shows values of b and c for the model in (8) obtained by applying the multiple regression analysis to the data of $((log_2 t)', (log_2 A)', (log_2 W)')$.

Tables VI and VII revealed the followings.

- Values of b are all positive, and values of c are all negative. Thus, the models by (7) and (8) appropriately represent that the pointing time becomes larger (smaller) as the target distance A (the target size W) becomes larger.
- For the device S, $|b| \leq |c|$ in all of the tables so that the target size W affects the pointing time more than the target distance A. This is consistent with the result reported in Section III.
- For the device L, $|b| > |c|$ for three condition combinations (targets S in Table VI(i), targets S in Table VII(i), and targets S in Table VII(ii)) so that the target distance A affects the pointing time more than the target size W under these conditions. This is also consistent with the result reported in Section III. However, $|b| < |c|$ for the other five condition combinations (e.g., targets L in Table VI(i)), which is not consistent with the result. This inconsistency should be further investigated in our future work.

TABLE VIII
THROUGHPUT VALUES (ID BY (9))

		Device L	
		Targets S	Targets L
Errors acceptable (α=1.62)	Mean	12.5	12.5
	SD	2.89	1.91
Errors not acceptable (α=1.61)	Mean	11.9	11.9
	SD	2.60	2.25

TABLE IX
T-TEST FOR THROUGHPUT (ID BY (9))

	Device L	
	Errors acceptable	Errors not acceptable
Targets S vs. L	$t=-5.75\times10^{-11}$ (α=1.62)	$t=4.17\times10^{-12}$ (α=1.61)

TABLE X
THROUGHPUT VALUES (ID BY (10))

		Device S	
		Targets S	Targets L
Errors acceptable (β=1.00)	Mean	5.73	5.73
	SD	1.37	1.14
Errors not acceptable (β=1.15)	Mean	4.78	4.78
	SD	1.15	1.09

TABLE XI
T-TEST FOR THROUGHPUT (ID BY (10))

	Device S	
	Errors acceptable	Errors not acceptable
Targets S vs. L	$t=-2.68\times10^{-11}$ (β=1.00)	$t=-3.60\times10^{-11}$ (β=1.15)

This result shows the multiple regression models by (7) and (8) represent the effects of target sizes and distances on the pointing time well for the small screen device and partially for the large screen device.

V. IMPROVEMENT IN FITTS' LAW ID FORMULATION

The authors next investigate an improvement to the definition of ID in the Fitts' model. The advantage of multiple regression models was shown in Section IV, but a drawback of the models is that users' pointing throughput values cannot be calculated. This is because a single index of difficulty is not defined in the case of the multiple regression models.

Our idea for the improvement is to raise A or W depending on the screen size as shown in (9) and (10).

$$ID = log_2(\frac{A^\alpha}{W} + 1), \alpha > 1 \tag{9}$$

$$ID = log_2(\frac{A}{W^\beta} + 1), \beta > 1 \tag{10}$$

Equation (9) is employed for larger screen devices and (10) for smaller ones.

The modified model is applied to the collected data. Appropriate values of α and β in (9) and (10) are explored by the bisection method so that there is no significant difference between population mean values of throughputs for the targets S&L (i.e., the throughputs are consistent between the two target sets) on the same device.

It is found that the modified model well fits to the data where $\alpha = 1.61, 1.62$ for the device L and $\beta = 1.00, 1.15$ for the device S (Tables VIII-XI): under these values of α and β, no significant difference is observed between population mean values of throughputs for the targets S&L. Thus, the modified indexes of difficulties by (9) and (10) well represent actual pointing difficulties for users so that users' pointing throughputs become consistent on the same device among target design variations (c.f., was inconsistent in the case of traditional ID, (2)).

This result shows that our idea of ID improvement is effective: the modified ID formulations can capture users' actual pointing difficulties better than the traditional ID. Further evaluations with additional experiments will be our future work.

VI. CONCLUSIONS

Index of difficulty formulation in Fitts' law was evaluated from the viewpoint of consistency in widget size&distance design variations. It was found that ID in (2) may not appropriately capture actual difficulty: user performances on the same device were not consistent among target designs (A, W) and (nA, nW).

Based on this finding, two multiple regression models were evaluated. These models were $t = F(A, W)$ (c.f., $t = F(\frac{A}{W})$ in the Fitts' model) which predicted the time t to point a target with the distance A and the size W. The models were found to be able to appropriately represent that W affected the index of difficulty more than A in the case of the small screen touch UI device. The models however did not work so well in the case of the large screen device, which remained to be investigated in our future work.

The authors next tried to improve the Fitts' law ID formulation. Our idea was to raise A or W depending on the screen size. The modified model could derive consistent results between target designs (A, W) and (nA, nW), which supports our idea.

REFERENCES

[1] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," Journal of Experimental Psychology, vol.47, vo.6, pp.381-391, 1954.
[2] I. S. MacKenzie, "Fitts's law as a research and design tool in human-computer interaction," Human-Computer Interaction, vol.7, pp.91-139, 1992.
[3] I. S. MacKenzie, and W. Buxton, "Extending Fitts' law to two-dimensional tasks," Proc. CHI'92, pp.219-226, 1992.
[4] M. Oehl, C. Sutter, and M. Ziefle, "Considerations on efficient touch interfaces - how display size influences the performance in an applied pointing task," in M. J. Smith, and G. Salvendy (eds.), Human Interface, Part I, HCII 2007, LNCS 4557, pp.136-143, 2007.
[5] R. Fujioka, T. Akiba, and H. Okada, "Evaluation of pointing efficiency on small screen touch user interfaces," in M. J. Smith, and G. Salvendy (eds.), Human Interface and the Management of Information, Part II, HCII 2009, LNCS 5618, pp.375-384, 2009.

[6] M. McClintock, and D. Hoiem, “Minimal target size in a pen-based system,” Abridged Proc. HCI International’ 93, p.243, 1993.
[7] X. Ren, and S. Moriya, “Selection strategies for small targets and the smallest maximum target size on pen-based systems,” IEICE trans. on information and systems, vol.E81-D, no.8, pp.822-828, 1998.
[8] R. L. Potter, L. J. Weldon, and B. Shneiderman, “Improving the accuracy of touch screens: an experimental evaluation of three strategies,” Proc. CHI’88, pp.27-32, 1988.
[9] A. Sears, “High precision touchscreens: design strategies and comparisons with a mouse,” International Journal of Man-Machine Studies, vol.34, no.4, pp.593-613, 1991.
[10] X. Ren, and S. Mizobuchi, “Investigating the usability of the stylus pen on handheld devices,” Proc. 4th Annual Workshop on HCI Research in MIS, pp.30-34, 2005.
[11] R. Plamondon, and A. M. Alimi, “Speed/accuracy trade-offs in target-directed movements,” Behavioral and Brain Sciences, vol.20, no.2, pp.279-349, 1997.
[12] ISO 9241, “Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: requirements for non-keyboard input devices,” 2000.
[13] T. Kvalseth, “An alternative to Fitts’ law,” Bulletin of the Psychonomic Society, vol.5, pp.371-373, 1980.

Hidehiko Okada is currently a Professor with the Department of Intelligent Systems, Faculty of Computer Science and Engineering at Kyoto Sangyo University. He received the B.E. degree in industrial engineering and the Ph.D. degree in engineering from Osaka Prefecture University in 1992 and 2003, respectively. He had been a researcher with NEC Corporation from 1992 to 2003, and since 2004 he has been with the university. His current research interests include computational intelligence and human-computer interaction. He is a member of Information Processing Society of Japan, Institute of Electronics, Information and Communication Engineers, Society of Instrument and Control Engineers, Japanese Society for Artificial Intelligence, Japan Society for Fuzzy Theory and Intelligent Informatics and Human Interface Society. He received the best paper award in the 1st International Conference on Industrial Application Engineering 2013.

Takayuki Akiba received the B.E. and M.E. degrees in information communication engineering from Kyoto Sangyo University in 2007 and 2009 respectively. While he was a student in the university, he had been researching on usability of small-screen touch user interfaces.