Evaluation of Pointing Efficiency on Small Screen Touch User Interfaces

Ryosuke Fujioka[1], Takayuki Akiba[2], and Hidehiko Okada[2]

[1] Kobe Sogo Sokki Co. Ltd.,
4-3-8 Kitanagase-dori, Chuo-ku, Kobe, Hyogo, 650-0012, Japan
ryo@sogosokki.co.jp
[2] Kyoto Sangyo University,
Kamigamo Motoyama, Kita-ku, Kyoto 603-8555, Japan
hidehiko@cc.kyoto-su.ac.jp

Abstract. Researchers have been investigating screen designs for small screen touch user interfaces (UIs), but further research is still required for smaller-screen devices including current smart phones. This paper reports on our evaluation of pointing efficiency on devices with touch-by-stylus small screen UIs. User performances were measured by experiments with three devices: a mobile phone, a PDA and a tablet PC. The size of pointing targets was designed so that the target index of difficulty (ID) by Fitts' law ranged in a consistent interval among the three devices. Users' pointing speed and accuracy were compared in terms of throughput and error rate respectively. It is found that the throughput and the error rate for the mobile phone were significantly smaller than those for the PDA and the tablet PC. It is also found that the error rate was not significantly larger in the case where users performed tasks with the mobile phone held by their hands than in the case where they did with the mobile phone put on desktop, although it was in the case of the PDA.

Keywords: usability, touch user interface, small screen, throughput, error rate, Fitts' law.

1 Introduction

Usability of mobile phones has been more important because more numbers and kinds of users use mobile phones. Smart phones with touch-by-stylus (or fingers) UIs have been available, providing new style of interactions. In designing UIs for such mobile phones, designers should be careful that screen sizes are smaller than other devices as PDAs and tablet PCs. Researchers have been investigating screen designs for small screen touch UIs [1-5], but further research will still be required for smaller-screen devices, some of which less than 3 inches.

The degree of difficulty for a user to point targets was formulated by Fitts, well-known as index of difficulty (ID) [6, 7]. ID is formulated as a function of the size and the distance of targets. Widgets (buttons, icons, menu items, etc.) can be designed for devices with various screen sizes so that ID values are theoretically consistent among

M.J. Smith and G. Salvendy (Eds.): Human Interface, Part II, HCII 2009, LNCS 5618, pp. 375–384, 2009.

the devices: larger/smaller sizes&distances for larger/smaller screens. If the ease of pointing targets under each of the widget design variations is consistent among the design variations, users' pointing performances are also consistent among them. The aim of our research is to investigate whether such scalability of pointing difficulties in screen sizes holds: if yes/no, widgets that are smaller-sized in accordance with the screen sizes are acceptable/not recommended for the smaller devices. The authors compare user' pointing performances by user experiments with three touch UI devices.

2 Evaluation Experiment

Methods and conditions of the experiment were designed as follows.

2.1 Test Tasks

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

2.2 Conditions

In this experiment, the following three conditions were employed. The "device" condition is necessary in this research, and the other two conditions were employed to investigate whether these conditions could affect user performance differences among devices: these conditions might help us discover some characteristic differences among devices and the differences may contribute to develop widget design guidelines.

Devices. This condition is for comparing user performances among devices. A mobile phone, a PDA and a tablet PC were used in the experiment. They were all commercial products with touch screens. Screen sizes of the devices were shown in Table 1. In the followings of this paper, the three devices are denoted as devices S, M and L. All the devices were set in portrait orientation. Each stylus attached to each device was used when a task with the device was performed.

Device Positions. The devices S&M can be used by holding them in a hand and pointing with a stylus by the other hand. Pointing performances might be affected by the device position, so this condition was employed in this experiment: pointing

Table 1. Device screen sizes

Devices	Sizes (inch)
(S) Mobile Phone	2.9
(M) PDA	3.6
(L) Tablet PC	10.2

performances were analyzed for the two device positions. Under both position conditions, each participant was seated and performed test tasks. The device L was excluded from the comparison under this condition because it was not easy for users to hold the device L *in a hand* as the devices S&M.

Errors. Pointing speed and accuracy are usually a tradeoff [8]. Users will point widgets carefully where mispointings cause critical problems (e.g., unintended and not-undoable deletion of important data). In such situations, pointing speed becomes much slower as the target widget is more difficult to point. The amount of decrease in speed may not be the same among the devices. Performances were analyzed for two error conditions: errors acceptable or not. In a test task where errors were acceptable, a participant could continue the task even if s/he made an error, 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 mispointing and the task was retried until the count of no-error attempts reached to a predefined number. The error condition was told to each participant before performing each task: s/he had to try a task more carefully in the errors-not-acceptable condition.

2.3 Design of Pointing Targets

The size of the target 1 was a constant with which the target could be pointed easily enough. The size of the target 2 was random within a predefined range. The size range was determined as follows. A small pointing target instance in the device S commercial screen design was a scroll bar of a web browser: the size was around 2.0mm. Based on this widget instance, the lower limit of the target 2 size was determined to 2.0mm. Besides, the device S doesn't have 10 keys and provides an on-screen keyboard for character inputs. The size of an alpha-numeric key on the software keyboard was 3.0mm. This size was employed as the upper limit of the target 2 size. Thus, on the device S screen the target 2 size was random within [2.0, 3.0]mm. The target 2 size ranges on the devices M&L were determined so that the range of "index of difficulty (ID)" in Fitts' law were consistent among the three devices: [2.0, 3.0]∗(3.6/2.9)mm for the device M, and [2.0, 3.0]∗(10.2/2.9)mm for the device L. The ID ranges in this experiment were shown in Table 2: mean and SD values of ID were around 3.40 bit and 0.72 bit, consistently among the three devices.

The width and the height of the target 2 were independently determined. The positions of the targets 1 and 2 were both random under the constraint that the two targets had no overlap. The size of the target 1 was constantly 6.0mm (= 3.0mm ∗ 2, where 3.0mm was the upper limit size of the target 2 on the device S). Participants seldom made errors for pointing the target 1 (a 6.0mm∗6.0mm rectangle). Fig. 1 shows a screenshot of targets 1 and 2 on the device S. The targets 1 and 2 are the

Table 2. Index of difficulty of the pointing targets

Device	Mean (bit)	SD (bit)	Min (bit)	Max (bit)
S	3.37	0.72	1.19	4.83
M	3.41	0.71	1.07	4.95
L	3.36	0.72	0.80	4.86

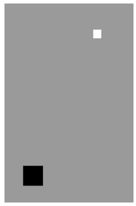

Fig. 1. Screenshot for target pointing tasks

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 participant 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.

2.4 Methods of Experiments

Table 3 shows 10 combinations of the conditions. Each participant performed test tasks seven times under each of the 10 condition combinations. The order of the 10 condition combinations was random. Under the condition where "position"="handheld", each participant was told not to put her/his hands on the desktop or her/his legs.

The first 2 out of the 7 trials of the test tasks were for training, so the log data of the 2 training trials were excluded from later analyses. The first 5 attempts in each trial of a task were also excluded because user performances in these attempts (beginning stage of a task) might not be stable enough. The number of successful pointing attempts in a task was set to 10.

Table 3. Condition combinations in the experiment

No.	Device	Position	Error
1	(S) Mobile Phone	Desktop	Acceptable
2			Not acceptable
3		Handheld	Acceptable
4			Not acceptable
5	(M) PDA	Desktop	Acceptable
6			Not acceptable
7		Handheld	Acceptable
8			Not acceptable
9	(L) Tablet PC	Desktop	Acceptable
10			Not acceptable

2.5 Participants

Eight subjects, the minimum recommended number for a user segment in CIF recommendations [9], participated in the experiment. They were university graduate or undergraduate students. They were all novices in using devices with touch-by-stylus UIs, but they had no trouble in performing test tasks after the two training trials for each condition combination.

2.6 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 land 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.

3 Data Analysis and Findings

3.1 Methods for Comparing Pointing Speed/Accuracy

Comparison of Pointing Speed. Pointing speed was measured by the throughput [10]. It is known as Fitts' law that target pointing time can be modeled as a function of target ID [6,7].

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

$$ID = \log_2(A/W+1). \quad (2)$$

In Eqs. (1-2), t is the pointing time, A is the distance between the target 1 and 2, W is the target 2 size and a, b are constant that depend on experiment conditions. 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 smaller of the target 2 width or height. This "smaller-of" method for W was reported as better in obtaining linear regression model (Eq.(1)) from users' pointing log data [11]. Throughput is defined as ID/t in Eqs. (1-2) [10]. Throughput is larger as a target with larger ID is pointed faster. (ID, t) could be observed for each attempt, so a value of throughput could also be obtained for each attempt. Mean and SD values of the throughput were calculated for comparing pointing speeds among different conditions. In addition, it was tested by t-test whether there was a significant difference between population mean values of throughput for two conditions.

It should be noted that error attempts were included in the data under the condition "error"="acceptable". Error attempts might be faster (of larger throughput values) than successful attempts. Thus, two ways are possible for analyzing throughput under the condition: with all the data including both successful and error attempts or with the data of successful attempts only. It was found by calculating throughput mean and SD values by both of the method that the differences of the mean/SD values were very small: <1% in the mean values and <5% in the SD values. In the following of this paper, the result by only the former method was reported.

Method for Comparing Pointing Accuracy. To measure pointing accuracy, error rate was defined.

Error rate = (#error attempts in a task trial)/(#total attempts in the trial) (3)

Mean and SD values of the error rate were calculated for comparing pointing accuracies among different conditions. In addition, it was tested by t-test whether there was a significant difference between population mean values of error rate for two conditions. Error rate could be calculated for only the condition "error"="acceptable" because the data under the condition "error"="not acceptable" didn't include any error attempt (if an error was occurred in a trial under the condition "error"="not acceptable", the trial was cancelled and retried[1]).

3.2 Mean & SD Values of Throughput and Error Rate

Table 4 shows mean & SD values of throughput and Table 5 shows those of error rate.

Table 4. Mean and SD values of throughput (bit/sec)

		Desktop		Handheld	
		Acceptable	Not Acceptable	Acceptable	Not Acceptable
S	mean	6.45	5.24	5.48	4.55
	SD	1.38	1.16	1.00	0.91
M	mean	6.68	5.94	6.16	5.42
	SD	1.25	1.06	1.19	0.92
L	mean	6.83	6.06	-	-
	SD	1.16	0.95	-	-

Table 5. Mean and SD values of error rate (%)

		Acceptable	
		Desktop	Handheld
S	mean	13.9	12.2
	SD	9.87	9.27
M	mean	3.56	8.82
	SD	6.46	9.40
L	mean	3.64	-
	SD	7.26	-

3.3 Comparisons among Devices

User performances are compared among devices in this section. Table 4 revealed that the smaller the screen size of a device, the smaller the mean value of throughput. Table 6 shows the result of t-test for testing whether there was a significant difference in population mean values of throughput among devices. In Table 6, for example,

[1] The number of "retries" could be an accuracy measure for the "not acceptable" condition, but the measure was not used in our analysis. This was because the number of retires could not be measured on the task basis but on the participant basis so that the number of data was small.

Table 6. T-test for throughput (comparison among devices)

		S vs. M	S vs. L	M vs. L
Desktop	Acceptable	t=-2.64**	t=-4.51**	t=-1.82
	Not acceptable	t=-8.90**	t=-11.0**	t=-1.74
Handheld	Acceptable	t=-9.29**	-	-
	Not acceptable	t=-13.4**	-	-

values in the column "S vs. M" are the t values for comparing the devices S&M (conditions other than the devices are the same). "**"-marked t-scores are those with p<0.01, and non-marked t-scores are those with p>0.05. Table 6 revealed that the throughput with the device S was significantly smaller than those with the devices M&L but no significant difference was observed between the throughputs with the devices M&L. This result indicates that users are likely to point targets slower on devices with smaller screens as the device S.

Table 5 revealed that the mean value of error rate with the device S was larger than those with the devices M&L, especially under the condition "position"="desktop". Table 7 shows the result of t-test for testing whether there was a significant difference between population mean values of error rate among devices. Table 7 revealed the followings.

- Under the condition "position"="desktop", the error rate with the device S was significantly larger than those with the devices M&L, but no significant difference was observed between the error rates with the devices M&L.
- Under the condition "position"="handheld", the difference of error rates between the devices S&M was not significant.
- Although the device M had a smaller screen than the device L, the estimated population mean of error rate was slightly smaller for the device M than for the device L. It should be noted that the difference between the devices M&L was much larger in mean throughput values (Table 6) than in mean error rate values (Table 7). It was likely that participants performed tasks much faster (and thus less accurately) on the device L so that the error rate with the device L was relatively larger.

Table 7. T-test for error rate (comparison among devices)

		S vs. M	S vs. L	M vs. L
Acceptable	Desktop	t=5.46**	t=5.22**	t=-0.05
	Handheld	t=1.60	-	-

It was found from the comparisons described above that pointing speed and accuracy were significantly smaller with the device S than those with the devices M&L (under the condition "error"="acceptable" and "position"="handheld", the accuracy was not but the speed was). Therefore, UI designers should pay more attention in designing widgets in a smaller-sized touch screen as the device S: widgets should be designed so that the ID values are smaller than those for larger screens as the devices M&L. To make the ID value smaller, the size/distance of widgets (W and A in Eq.(2)) are required to be larger/smaller. Further research is necessary to investigate which of the size or distance is more effective in assuring desirable pointing performance on smaller touch screen devices.

3.4 Comparisons among Device Positions

User performances are compared among device positions in this section. Table 4 revealed that the mean value of throughput was smaller under the condition "handheld" than that under the condition "desktop". This holds for both of the devices S&M. Table 8 shows the result of t-test for testing whether there was a significant difference in population mean values of throughput among device positions. Table 8 revealed that, for both of the devices S&M, the throughput for "handheld" was significantly smaller than that for "desktop". This result indicates that users are likely to point targets slower if they hold a device in their hands than if they put the device on a desktop, which will hold for both PDAs and smaller smart phones.

Table 8. T-test for throughput (comparison between device position conditions)

		Desktop vs. Handheld
S	Acceptable	t=12.2**
	Not acceptable	t=9.33**
M	Acceptable	t=6.20**
	Not acceptable	t=7.40**

Table 5 revealed that, the mean value of error rate with the device M was larger for "handheld" than for "desktop", but those with the device S was larger for "desktop" than for "handheld" (it should be noted that the difference was relatively large for the device M but small for the device S). Table 9 shows the result of t-test for testing whether there was a significant difference in population mean values of error rate among device positions. Table 9 revealed that, the error rate with the deice M was significantly smaller for "handheld" than for "desktop" but no significant difference was observed in population mean values of error rate with the device S for "handheld" and "desktop".

Table 9. T-test for error rate (comparison between device position conditions)

		Desktop vs. Handheld
Acceptable	S	t=0.77
	M	t=-2.88**

It was found from the comparisons described above that, with the device M the speed and the accuracy were both smaller for "handheld" than those for "desktop", but with the device S only the speed was smaller (the accuracy was not). This will be partially because the device S was smaller that the device M so that the device S was easier for users to hold in a hand. Although the participants were asked to perform tasks as fast&accurate as possible, they tended to point targets much slowly (to keep accuracy to a certain extent) in the case of the device S and the "handheld" condition.

This result indicates that designers of smaller screen touch UIs as in the device S should keep in mind users may not be able to point widgets more accurately even if they put the device on a desktop (more stable than the handheld situation). This finding will be useful for future mobile UI designs in which smart sensors enable them to design device-position-adaptive UIs. Further research is necessary to investigate user performances while users stand performing tasks.

3.5 Comparisons among Error Conditions

User performances are compared among error conditions in this section. As noted in 3.1 the error rate could be calculated for "acceptable" only, so the error rates cannot be compared among error conditions: only the throughputs are compared.

Table 4 revealed that the mean value of throughput was smaller for "not acceptable" than for "acceptable", which held regardless of the devices {S,M,L}. Table 10 shows the result of t-test for testing whether there was a significant difference in population mean values of throughput among error conditions. Table 10 revealed that, for all the devices {S,M,L}, the throughput for "not acceptable" was significantly smaller than that for "acceptable". This result indicates that there is no clear difference in the effect of the error conditions between the device S and the others (M, L): screen design of devices as the device S will require the same considerations as that of lager screen devices as the devices M&L.

Table 10. T-test for throughput (comparison between error conditions)

		Acceptable vs. Not acceptable
S	Desktop	t=14.0**
	Handheld	t=14.2**
M	Desktop	t=9.11**
	Handheld	t=10.2**
L	Desktop	t=10.4**

4 Conclusions

Users' pointing speed and accuracy on a mobile phone with a small screen touch UI were compared with those on a PDA and tablet PC with larger screens. Our findings in designing widgets on the UI are summarized as follows.

- On the mobile phone, the speeds were significantly smaller for any condition and the accuracies were also significantly smaller for most of the conditions, than on the PDA and the tablet PC. Thus, designers should keep in mind that relatively difficult-to-point widgets will be acceptable on device as the PDA and the tablet PC but should be avoided for usability on devices as the mobile phone (more specifically, widgets with ID 4.0-5.0[2]). It should also be kept in mind that, on small screen devices as the mobile phone, decreases in the widget distance (A in Eq.(2))

[2] See Table 2: in this experiment the ID values ranged in about [1.0, 5.0], so the difficult-to-point targets were those with IDs in [4.0, 5.0].

theoretically make the ID values smaller but may not contribute well to improve users' pointing performances: instead, increases in the widget size (W in Eq.(2)) will be necessary.

- On the PDA, the accuracy was significantly larger for the "desktop" condition than for the "handheld" condition. On the mobile phone, however, the accuracy was not. This means that, in a case of a small screen device as the mobile phone, designers should not expect users can point targets more accurately even if they put the devices on a desktop. Smart sensors will enable to design device-position-aware UIs which can adapt widget designs to the current device position, but widget sizes should not be changed smaller for the "put-on-a-desktop" situation.

The authors hope these findings will contribute to usable screen designs for small screen devices. Further research will still be necessary to, e.g., obtain more specific guidelines in widget sizes and distances according to screen sizes, and investigate ways to appropriately control IDs (investigate how theoretical changes in ID values by changing W and/or A affect users' actual pointing performances). Experiments with more devices, more participants, under other conditions such as performing tasks while standing, are also included in our future work.

References

1. McClintock, M., Hoiem, D.: Minimal Target Size in a Pen-based System. In: Abridged Proc. of 5th Int. Conf. on Human-Computer Interaction (HCI International 1993), p. 243 (1993)
2. Douglas, S.A., Mithal, A.K.: The Ergonomics of Computer Pointing Devices. Springer, Heidelberg (1997)
3. Sarah, A., Douglas, S.A., Kirkpatrick, A.E., MacKenzie, I.S.: Testing Pointing Device Performance and User Assessment with the ISO 9241, Part 9 Standard. In: Proc. of ACM conf. on Human Factors in Computing Systems (CHI 1999), pp. 215–222 (1999)
4. Soukoreff, R.W., MacKenzie, I.S.: Towards a Standard for Pointing Device Evaluation, Perspectives on 27 Years of Fitts' Law Research in HCI. Int. J. of Human-Computer Studies 61(6), 751–789 (2004)
5. Oehl, M., Sutter, C., Ziefle, M.: Considerations on Efficient Touch Interfaces - How Display Size Influences the Performance in an Applied Pointing Task. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007. LNCS, vol. 4557, pp. 136–143. Springer, Heidelberg (2007)
6. Fitts, P.M.: The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. Journal of Experimental Psychology 47(6), 381–391 (1954)
7. MacKenzie, I.S.: Fitts's Law as a Research and Design Tool in Human-Computer Interaction. Human-Computer Interaction 7, 91–139 (1992)
8. Plamondon, R., Alimi, A.M.: Speed/Accuracy Trade-offs in Target-Directed Movements. Behavioral and Brain Sciences 20(2), 279–349 (1997)
9. ANSI/NCITS 354-2001, Common Industry Format for Usability Test Reports (2001)
10. ISO 9241, Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) - Part 9: Requirements for Non-Keyboard Input Devices (2000)
11. MacKenzie, I.S., Buxton, W.: Extending Fitts' Law to Two-dimensional Tasks. In: Proc. of ACM Conf. on Human Factors in Computing Systems (CHI 1992), pp. 219–226 (1992)