Human-Computer Interaction -- INTERACT'03
M. Rauterberg et al. (Eds.)
Published by IOS Press, (c) IFIP, 2003, pp. 757-760

Towards Guidelines for Touch Screen Design: Perception of Button Form and Extension

M. Fjeld, H. J. Zuberbühler, S. Guttormsen, F. Voorhorst, H. Krueger
IHA, ETH Zurich, Clausiusstr. 25, CH-8092 Zurich, Switzerland
morten@fjeld.ch

Abstract: An evaluation of design-relevant perceptual cues for the design of touch screens is presented. Button form (round and rectangular) and button extensions (height and size) were examined. For each of the two forms and with size and height as perceptual cues, we estimated perceptual threshold in terms of pixels. Using a within-group experimental design and four conditions, 12 subjects were presented with 16 varieties of a find-and-select task, 8 tasks with a lower time limit and 8 tasks with a higher time limit. Significant differences in perceptual thresholds between the conditions were observed. The main conclusion is that button height should be preferred over button size as perceptual cue, which is equally valid for round and rectangular buttons. If height is chosen as cue, round buttons should be preferred. If size is chosen as cue, rectangular buttons should be preferred.

Keywords: Touch screen, button, design, perceptual threshold, best-PEST

1 Introduction

Although the practice of touch screen design is widespread, the availability of established design guidelines for such applications is scarce. A major interactive element in touch screens is the button, which may take different forms and have various perceptual cues. From previous research into perception of button cues available, we have seen that colour and texture of buttons are of far less importance than button form, size, and height (Voorhorst et al., 2002). Therefore, we decided to carry out an experiment focussing on the perceived form and extension of buttons. The forms examined were (round, rectangular); the extensions examined were (height, size). For each of the two forms and using height and size as perceptual cues, we estimated perceptual threshold values in terms of pixels. An experiment based on a within-group design with four conditions involved 12 subjects repeatedly presented with a find and select task. Each subject solved a total of 16 tasks at a touch screen, 8 with a lower time limit and 8 with a higher time limit. There were 3x4 buttons, where one was different in size or height. The different one had to be found and selected within the time limit. The perceptual threshold is defined as the number of pixels that a button had to be different from the other ones, in terms of size or height, in order to be found and selected. A typical task start-up under the condition (rectangular, height) is shown in Fig. 1. The threshold estimate was adjusted continuously using the so-called best-PEST method, which is described below. For all four conditions we estimated perceptual thresholds, which were found to be significantly different. This was an exploratory study, and no hypotheses were tested. Rather, we sought estimates for perceptual cues to generate guidelines for design of touch screen button.

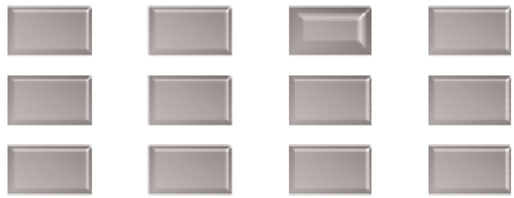

Figure 1: Task start-up in condition (rectangular, height).

2 Estimation Methods

For the measurement of threshold, psychophysics suggests different procedures. Some of them stem from basic research and yield maximal accuracy at the price of an enormous effort (i.e. hundreds of trials have to be run). Others stem from applied research and provide suitable accuracy with minimal effort. The requirement of minimal number of trials comes from the fact that after a long run of trials, experimental subjects tend to fatigue and to be bored, resulting in an apparent drift of their thresholds. In order to avoid such effects, so-called adaptive psychophysical procedures have been developed. The purpose of these methods is to minimise the number of trials. Some methods are:

- Forced-Choice Tracking (Zwislocki et al., 1957)
- Up-Down Transformed Response (UDTR) Method (Wetherill and Levitt, 1965; Levitt, 1971)
- Parameter Estimation by Sequential Testing (PEST) (Taylor and Creelman, 1967)
- Improved PEST (Findlay, 1978)
- best-PEST (Pentland, 1980; Lieberman and Pentland, 1982)

Pentland (1980) showed that best-PEST is the most efficient method in terms of accuracy and number of trials. The most striking difference between best-PEST and other adaptive methods is that best-PEST uses a statistical estimation of the observer's threshold, which is made from all the results obtained from task start-up. That is, all available information determines how much the intensity on the next trial will be increased or decreased. If the stimulus intensity of the prior trial is below the estimated threshold, it will be increased on the next trial, and vice versa. After each trial, a new estimation of the threshold is calculated and the stimulus intensity is adjusted accordingly. On the one hand, this is the reason for best-PEST's remarkable efficiency. On the other hand, the method becomes vulnerable to the subject's failures at the beginning of the experiment. Hence, in order to choose the adequate method, one has to trade off between rapidity and error vulnerability. Based on experimental data Rammsayer and Sachsee (1990) suggest UDTR, if the subjects - for whatever reason - tend to give false responses, and best-PEST, if minimal number of trials are required. For the purpose of rapid evaluation of thresholds, we considered the best-PEST method to fulfil the requirements best: Accuracy and rapidity at the price of some error vulnerability. In the following we describe the minimal background necessary to utilise the best-PEST calculator (Zuberbühler, 2002). For a detailed description of adaptive procedures see the overview of Treutwein (1995).

3 Empirical evaluation

Psychophysical procedures dispose of various testing paradigms. Two major ones are the yes-no mode and the forced-choice (nAFC: n-alternative-forced-choice) mode. With the yes-no mode, subjects are given a series of trials, in which they must judge the presence or absence of a stimulus. This is of little use in our case, since users have to find and select one of many. A testing mode better suited to our needs is the nAFC. Here, subjects are given a variety of n (here: n=3x4) alternatives, from which they have to choose the one containing the stimulus (here: different button height or size). The subjects know that exactly one alternative contains the stimulus and that the other ones have a zero stimulus. The dependent variable is the percentage of correct responses for a given stimulus intensity (here: number of pixels, 0-20), expected to take values between 8.33% and 100%. The lowest value is expected at zero stimulus, where subjects have a chance of 1/(3x4) to select the different button.

3.1 Method

The first of the two independents was condition, given by form (rectangular or round) and extension (height or size). The four conditions were defined as: (RectH, RectS, RoundH, RoundS). The second independent was time limit, being either 2 or 8 seconds. The first of the three dependents was perceptual threshold, measured in numbers of pixels. The second dependent was task average of trial time, from trial start-up to selection or time limit. The third dependent was task average of correct (+1) and wrong (-1) trials.

The experiment was run on a PC NT4 with an ELO Touch Screen. This touch screen has 1024 x 768 pixels. We computed the size of each pixel to be 0.3 mm. The screen was placed in front view and its distance to subjects was not controlled. There were 12 unpaid subjects, 5 females and 7 males, aged 26 to 47. To examine potential learning effects, we controlled repetition of conditions and task number. Potential effects of age and gender were also examined.

A touch screen with 3x4 buttons was presented. The task was to find and select the different button as fast as possible and within a time limit. The button was different either in height or in size. Each task consisted of ten trials. Selecting a wrong button or not selecting before the time limit were both counted as wrong. The experimental set-up, including stimulus presentation, best-PEST algorithm (Zuberbühler, 2002), and data acquisition was implemented using Macromedia Director's scripting language Lingo. Based on an unmodified rectangular or round button as zero stimulus, a set of increasing stimulus intensities was generated by increasing size, resp. height[1], stepwise from 1 to 20 pixels. At task start-up, the stimulus always differed by 10 pixels from zero stimulus. All tasks were variations of a 3x4 button matrix, all buttons being either rectangular or round. The different button was randomly positioned in each trial. At correct button selection, best-PEST decreased the estimate of the perceptual threshold in the following trial. At wrong or no selection, best-PEST increased the estimate of

[1] Using bevel width, which is a PhotoShop plug-in function.

the perceptual threshold in the following trial. Each subject solved 16 tasks. The presentation of different tasks was given in the following way: For one half of the subjects, the first 8 tasks were presented with the lower time limit, the last 8 tasks with the higher time limit. For the other half of the subjects, 8 tasks with the higher time limit were followed by 8 tasks with the lower time limit. The sets of 8 were pairs of the four conditions, ordered so that, each condition appeared the same number of times at each position.

3.2 Effects of Perceptual Cues

The centre column of Fig. 2 shows estimated thresholds; the left-hand columns shows zero stimulus buttons; the right column shows buttons at task start-up. Significant differences in estimated pixel thresholds were observed (Fig. 3; Table 1). ANOVA gave significant differences in perceptual threshold for (RectH, RectS, $p < 0.0001$), (RectH, RoundS, $p < 0.0001$), (RectS, RoundH, $p < 0.0001$), (RectS, RoundS, $p = 0.0028$), (RoundH, RoundS, $p < 0.0001$). ANOVA gave significant differences in correct trials for (RectH, RectS, $p=0.0189$), (RectS, RoundH, $p=0.0109$) (Fig. 4). No significant differences in trial time per condition were found.

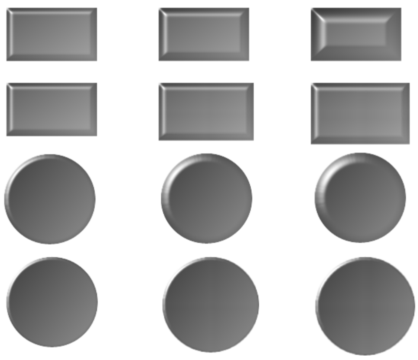

Figure 2: Conditions (rows) and stimulus (columns): unmodified; estimated threshold; task start-up.

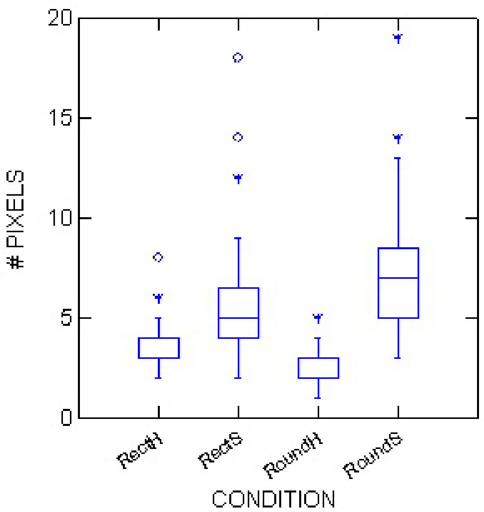

Figure 3: Perceptual threshold per condition.

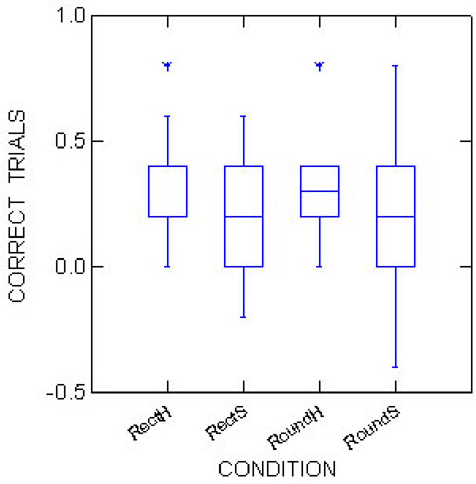

Figure 4: Correct trials per condition.

	RectH	RectS	RoundH	RoundS
Pixel, mean	3.46	5.88	2.73	7.31
Pixel, SD	1.25	2.95	0.94	3.24
Pixel, integer	3	6	3	7

Table 1: Mean, standard deviation, and integer values for perceptual thresholds of pixels per condition.

3.3 Further observations

Under the lower time limit, the perceptual threshold was slightly higher (Fig. 5). Further, under the lower time limit, there was less variation in, and lower values of trial time (Fig. 6). The subjects got no training prior to task solving. Still, no learning effect was observed from the first to the last task presented. There was no effect of age or gender.

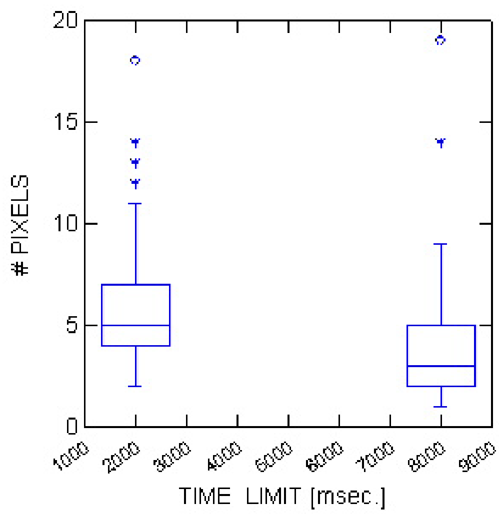

Figure 5: Perceptual threshold per time limit (2 or 8 sec.).

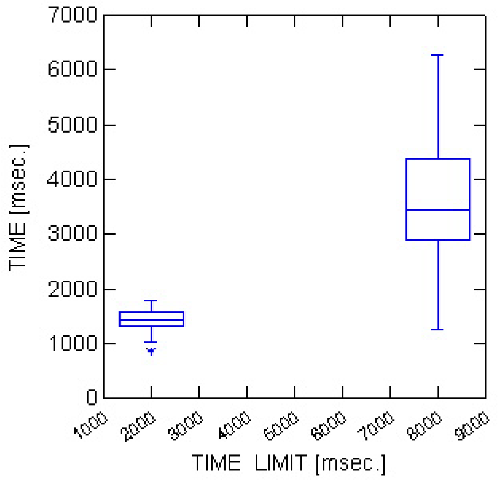

Figure 6: Trial time per time limit (2 or 8 sec.).

3.4 Discussion of Experimental Design

First, the screen was not a flat but a normal screen. Hence, the buttons at the edges of the screen might have appeared with different size than those in the centre due to distortion. A control of this effect could have been reached by logging the position of the buttons selected. Using a flat screen might have compensated the whole effect. Second, for right-handed subjects, the right-hand side of the screen was more in focus and easier to use, potentially introducing an unwanted skewness in the experiment. Third, button height was generated using bevel width. Therefore, it is worthy to ask whether one pixel height is comparable to one pixel size. Fourth, there was a sound feedback whenever the screen was touched. That is, there was also a sound feedback when an area without any button was touched, which might have been misleading. For future experiments, maybe there should only be sound feedback when a button is selected. Finally, a further feedback, informing about right or wrong button, may possibly enhance learning.

4 Conclusion

The experiment focussed on the perception of button size and button height. The perceptual thresholds for these different perceptual cues were examined. As a major result, we found that fewer pixels are required to make buttons stand out by modifying their height than by modifying their size. For touch screen designers, this means that when there is an equal choice between the use of height and size as perceptual cue, they should preferably use height. This recommendation applies to rectangular as well as to round buttons. Ranking button form and cue by number of pixels, gives (low to high number of pixels, see also Table 1):
(RoundH: 3, RectH: 3, RectS: 6, RoundS: 7).

For the design of touch screen buttons, the following guidelines may be suggested, where the first one has higher priority than the two following ones:

1. With equal choice among height and size as cue, height should be the preferred cue.
2. Given height as cue and equal choice among button forms, round buttons should be preferred.
3. Given size as cue and equal choice among button forms, rectangular buttons should be preferred.

References

Findlay, J. M. (1978): Estimates on probability functions: A more virulent PEST. *Perception & Psychophysics*, vol. 23, pp. 181-185.

Lieberman, H. R., & Pentland, A. (1982): Microcomputer-based estimation of psychophysical thresholds: The Best PEST. *Behaviour Research Methods & Instrumentation*, vol. 14, pp. 21-25.

Levitt, H. L. (1971): Transformed up-down methods in psychophysics. *J. Acoust. Soc. Am.*, vol. 49, pp. 467-477.

Pentland, A. (1980): Maximum likelihood estimation: The best PEST. *Perception & Psychophysics*, vol. 28, pp. 377-379.

Rammsayer, T., & Sachsee, S.(1990): Zur Anwendung adaptiver Verfahren in psychophysischen Experimenten. *Psychologische Beiträge*, vol. 32, pp. 361-379.

Taylor, M. M., & Creelman, C. D. (1967): PEST: Efficient estimates on probability functions. *J. Acoust. Soc. Am.*, vol. 41, pp. 782-787.

Treutwein, B. (1965): Adaptive Psychophysical Procedures. *Vision Research*, vol. 35, pp. 2503-2522.

Voorhorst, F., Fjeld, M., Leuthold, K., Guttormsen, S., & Stulzer, W. (2002): *The Touch-Ability of Touch-Screens*. Report KTI-Nr. / CTI-No. 4615.1, Bern.

Wetherill, G. B., & Levitt, H. (1965): Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology*, vol. 18, pp. 1-10.

Zuberbühler, H. J. (2002): Rapid Evaluation of Perceptual Thresholds - The Best-Pest Calculator: A web-based application for non-expert users. http://www.psychophysics.ethz.ch/tools

Zwislocki, J. J., Maire, F., Feldman, A. S., & Rubin, H. (1957): On the effect of practice and motivation on the threshold of audibility. *J. Acoust. Soc. Am.*, vol. 30, pp. 254-262.