Investigation of Indirect Oral Operation Method for Think Aloud Usability Testing

Masahiro Hori, Yasunori Kihara, and Takashi Kato

Graduate School of Informatics, Kansai University,
2-1-1 Ryozenji-cho, Takatsuki, Osaka 569-1095 Japan
{horim,tkato}@res.kutc.kansai-u.ac.jp

Abstract. Usability testing with prototypes is typically conducted with a concurrent think-aloud protocol. Due to the simultaneous process of prototype operation and verbalization, participants of the think-aloud testing sometimes say very little and are likely to become silent when they are required to think abstractly or complete complex tasks. In this paper, we propose a method of user operation with oral instruction, which facilitates thinking aloud because oral operation would help participants to keep a continuous flow of verbalization. To investigate the quantity and quality of utterances made during think aloud protocols, we conducted a comparative study between oral and conventional manual operation methods. The study was carried out with two test objects: an interactive prototype of a touch-screen digital camera and photo album software with standard mouse/keyboard user interface. Our results demonstrated that the oral operation method was more effective in drawing more utterances for explanation and observation that would be an important source of discovering usability problems although the effect was dependent on the user interface of test objects.

Keywords: Concurrent think-aloud protocols, keep talking, usability testing, prototype evaluation.

1 Introduction

Prototyping is widely used for the evaluation of design ideas, in particular, aspects of usability and user experiences [1]. Prototypes are becoming richer and more interactive [2][3], and that would be helpful for testing usability of emerging products enhanced their capabilities by input components such as touch screen and sensors. However, interactive prototypes may not necessarily achieve operational performance expected in the final products, and that would expose slow or inaccurate responses to test users. The limitation of an interactive prototype in its early development stage is that the users cannot distinguish if flaws they experienced are intrinsic problem of the artifact or come from insufficient operational performance of the prototype.

Usability testing with prototypes is typically conducted with a concurrent think-aloud protocol, which is the process of having the potential users verbalize what they are thinking while completing tasks with an artifact to be tested [4][5][6][7]. Think-aloud protocols have been widely used to collect qualitative data through the actual

use of an artifact [8][9]. The assumptions underlying the think-aloud method is to provide insights from the information resides in the short-term memory in parallel with the users' thought process [10]. To put it another way, the think-aloud method allows to reveal cognitive activities that may not be visible without users' verbal reports.

In usability testing with think-aloud protocols, users are instructed to continuously verbalize their thoughts. However, if users keep silent for a long time, the verbalization will not be usable because significant information in the short-term memory may not be tracked down [11]. The difficulty thus stems from the simultaneous process of verbalization and cognitive task to be made relying on the user's short-term memory. Actually, users say very little and are likely to become silent during concurrent think aloud protocols when they are required to think abstractly or complete complex tasks [12].

It is suggested in the original think-aloud method [13] that with careful instructions for the participants at the beginning of a test session, the session should be conducted with minimum interruption except for reminding the participants to keep talking when they fall silent. However, it is pointed out [8] that actual practice of think aloud protocols for usability testing diverges from the theoretical foundation [13] in the manner of reminding participants to think aloud.

In pursuit of the constant verbalization, there has been proposed a dialogue approach where a test session is conducted interactively between participant and facilitator [14]. Although the dialogue approach may be helpful when the participant is stuck or the prototype system crashes, it would be possible for the test facilitator to detect such situations and handle as extraneous intervention without relying on the dialog approach. Boren and Ramey [8] proposed an interactive think-aloud method based a speech communication framework, where thinking aloud is regarded as a shared task for participant and facilitator rather than just a participant task. In particular, it is argued in the method that careful use of acknowledged tokens (*e.g.*, "OK", "yeah", or "mm hm") will keep continuous flow of verbalization. However, such an interactive approach may reflect the facilitators' judgments on the participant thoughts, and cause validity problems in usability studies [11]. For instance, acknowledgements from the facilitator might be given earlier than the response of a test object, and hinder participant's opportunities to think about actual feedback from the test object.

In this paper, we propose a method of user operation with oral instruction for think aloud usability testing with interactive prototypes. The proposed operation method allows participants to concentrate on the evaluation of a test object regardless of the insufficient operational performance of prototypes. Moreover, the proposed method facilitates thinking aloud because oral operation would help participants to keep a continuous flow of verbalization, which is not necessarily easy in the conventional thinking aloud method [9].

The oral operation approach does not entail dialogue between the participant and facilitator, but include verbal interaction with a test object through the facilitator. The idea behind this approach is to involve the facilitator into one side of the interaction, namely, from the participant to a test object, and minimize intervention into another side of from the test object to the participant. The oral or verbalized operation does not interrupt the participant's flow of thinking process and contributes to giving more chances for the participant to think aloud because the oral instruction is the action the

participant is going to take to work with the test object with reference to information in the short-term memory.

In order to figure out situations that allow participants to concentrate on evaluation of a test object regardless of its insufficient operational performance of its prototype, this paper describes an experiment to compare conventional manual operation and indirect oral operation methods for think-aloud protocols. Following research questions will be addressed in the experiment presented in the remainder of this paper:

1. Do the two methods differ in terms of the easiness of operation?
2. Do the two methods differ in terms of the quantity/quality of utterances?

2   Method

2.1   Test Object

In pursuit of the above research questions, we conducted a study to compare results of usability testing for an existing touch-screen digital camera and its accompanying photo album software. In particular, we used two test objects: a working product of the photo album software, and an interactive prototype of a digital camera created with presentation software.

Although it is assumed in this study that operational performance of prototypes is insufficient, such insufficiency may not occur depending on type of test objects. For instance, if indirect oral operation was forced for use of ordinarily-familiar product such as desktop PCs with mouse/keyboard interface, it may hinder users to make spontaneous verbalization due to cumbersome interaction with the test object. Therefore, we adopted two test objects with different user interface. One is a prototype of a touch-screen digital camera created with presentation software on a notebook tablet PC (Figure 1). The prototype of a digital camera provides single-touch screen interface that allows operations such tapping for selection, flick for scrolling the screen, and drag-and-drop for moving an object. Another test object is a product version of photo album software operated with ordinary mouse/keyboard user interface.
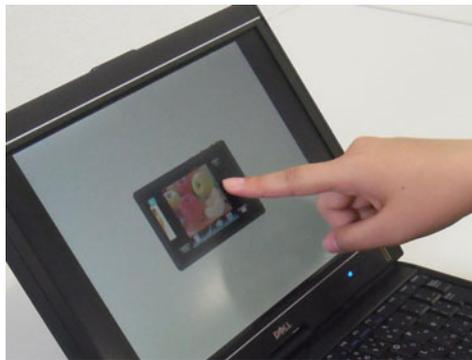

Fig. 1. Prototype of touch-screen digital camera on a tablet PC

2.2   Design

The experiment had a nested design with mixed within and between-participant factors. The main within-participant factor was operation method (direct manual vs. indirect oral), and the between-participant factor was user interface (single-touch screen vs. standard mouse/keyboard). Participants were thus divided into two test groups. Dependent variables were: task-completion time, quantity/quality of utterances, subjective measures from a questionnaire given to the participants.

2.3   Participants

The experiment was conducted with a sample of 32 participants, all of whom were third or fourth year students of Faculty of Informatics at the Kansai University. All the participants had no prior experience in think-aloud protocols, and were paid to participate in the experiment. The participants were recruited by means of e-mail and printed announcements, and selected on a first-come, first-serve basis.

2.4   Tasks

To consider differences between the manual and oral operation methods, four tasks listed in Table 1 were given to all the participants. Tasks 1 and 2 were designed for the prototype of a digital camera, and tasks 3 and 4 for the photo album software. Given on the right of each task description is the number of operation steps to complete the task. For example, the task 2 requires 7 steps, which consist of four tapping, two flick, and one drag-and-drop operations on the test object.

Table 1. List of tasks

Test object		Task description	Task steps
Digital camera	(1)	Delete a photo of strawberry	8
	(2)	Move a photo of muscat into the folder '1'	7
Photo album software	(3)	Put letters 'apple' on a photo of apple, and change the photo file name as 'apple.jpg'	9
	(4)	Change image quality of a banana photo to standard, and move the photo into the folder '1'	11

2.5   Questionnaires

Two forms of questionnaire were prepared to obtain participants' evaluations on a seven-point Likert scale (1 being very negative and 7 being very positive), and handed out to all the participants at the end of a session for each test object. Q1 asked participants rate the ease of finding target objects (*e.g.*, menu, button, and icon) to interact with, and Q2 to rate the ease of applying actions (*e.g.*, clicking, tapping, and flicking) to the objects found as targets.

2.6 Procedure

The experiment was conducted individually in a meeting room (7 x 8 m) at the Research Center of Advanced Informatics, Kansai University, and each session lasted about 30 minutes. One facilitator was present in the room throughout the sessions to observe participants. Before starting a session, participants were informed that all the sessions would be videotaped from behind. However, participants were not told that their verbalizations would be analyzed for comparing operation methods because that may affect participants' verbalization quantity and/or content. Instead, participants were told that the session was held for the usability evaluation of target objects.

When starting a session with each test object, participants were provided with a printed user manual. All the participants were asked to complete a set of tasks, and to constantly verbalize their thoughts while working on the tasks. In particular, one half of the participants were asked to operate prototypes directly by themselves [Figure 2 (a)] whereas another half was asked to operate indirectly, giving oral instructions to an operator [Figure 2 (b)].

After a practice task on making crane of paper folding (*Origami*) to rehearse oral or manual operation method as well as thinking aloud, participants attempted the four tasks. The two tasks were given for the prototype of a digital camera with touch screen interface, and the other two for the photo album software with a standard mouse/keyboard interface. In each group, the task order was counterbalanced across participants to prevent order effect. After completion of the two tasks for each test object, participants were asked to rate on a seven-point scale the ease of operation with the test objects.
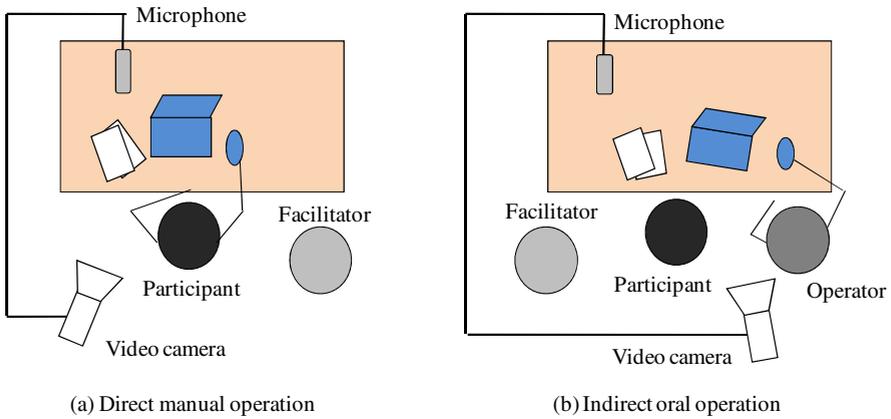

(a) Direct manual operation   (b) Indirect oral operation

Fig. 2. Experimental setting for participant's operation

3 Results and Discussion

3.1 Easiness of Operation

Since the tasks in this study consist of different number of steps (Figure 1), the task completion time was normalized and compared as operation time per task step. A

two-way ANOVA for the single-step operation time (Table 2) revealed that there was a signifiant interaction [$F(1,30)=12.76$, $p<0.005$] between the operation method and user interface. The simple main effect of operation method was significant when the touch-panel interface was used [$F(1,30)=23.05$, $p<0.0001$]. For the user interface, the simple main effect was significant when the oral operation method was used [$F(1,30)=38.43$, $p<0.0001$]. These results show that tasks with touch-panel interface took more time (46.24 sec. per task step) to complete when the oral operation method was used. In contrast, when the manual operation method was used, there was no significant difference between the two types of user interface.

The participants' rating for the easiness of operation is given in Table 3. A two-way ANOVA for the easiness of finding target objects (Q1 in the quesionnaire) revealed a significant interaction [$F(1,30)=4.95$, $p<0.05$] between the operation method and user interface. The simple main effect of user interface was significant when the oral operation method was used [$F(1,30)=21.67$, $p<0.001$]. In addition, for the easiness of applying actions to target objects (Q2 in the quesionnaire), there was a significant interaction [$F(1,30)=4.48$, $p<0.05$] between the operation method and user interface. The simple main effect of user interface was significant when the manual operation was used [$F(1,30)=7.09$, $p<0.05$], and was not significant under the condition of the oral operation.

Table 2. Operation time per task step (in seconds)

User interface	Manual (n=16)		Oral (n=16)	
	Mean	SD	Mean	SD
Touch screen	25.17	14.22	46.24	16.52
Mouse/keyboard	20.73	7.76	22.24	6.51

Table 3. Participants' rating for the easiness of operation. (Higher means indicate more positive assessment).

		Manual (n=16)		Oral (n=16)	
	User interface	Mean	SD	Mean	SD
Q1: Finding objects	Touch screen	3.88	1.05	3.31	1.67
	Mouse/keyboard	4.63	1.41	5.63	1.05
Q2: Applying actions	Touch screen	4.81	1.88	4.94	1.44
	Mouse/keyboard	6.31	1.04	4.75	1.56

The above results of the participants' rating particularly for the oral operation method indicate that finding target objects is easier for the software with mouse/keyboard (5.63) than the interactive prototype with a touch-panel screen (3.31). However, there is no difference in the perceived easiness of applying actions with the oral operation. When an action is being applied to an object, the target object is already identified on the screen by the user. Therefore, it is probable that the intrinsic difficulty of the oral operation will be in the process of identifying a target object.

3.2   Quantity/Quality of Utterances

In addition to the easiness of operation, the utterances made during the experiment were investigated with regard to the quantity and quality. Verbalization categories assumed in this experiment are given in Table 4, and overall numbers of utterances made during the four tasks are given in Figure 3 with classification into the categories.

Table 4. Category of transcribed unit of verbalization

| Verbalization category	Description
Explanation	Describe the reason for participant's behavior; make a prediction about a result of participant's activity
Procedure	Describe participant's activities
Observation	Describe the observation of what occurred as a result of participant's activity
Other	Utterances that do not fit into one of the other categories
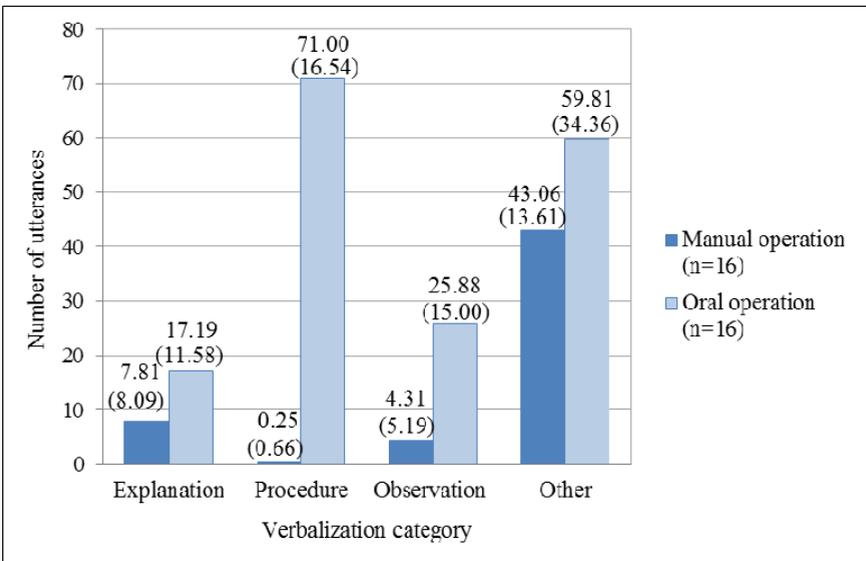
Fig. 3. Number of utterances made during the four tasks (SD in parentheses)

Since the participants in the group of manual operation performed tasks without verbalizing their operations, the average number of procedure utterances with manual operation 0.25 was much fewer than those with oral operation 71.00 as shown in Figure 3. More importantly, the numbers of explanation (17.19) and observation (25.88) utterances with the oral operation were both significantly higher than the corresponding utterances with manual operation ($p < 0.05$ and $p < 0.01$ respectively, Welch's t-test). In addition, for the number of 'other' category utterances, there was no

significant difference between the oral and manual operations ($p > 0.05$, Welch's t-test). These results indicate that the number of utterances for the explanation and observation increased substantially when the oral operation method was used. Note here that participants' utterances for explanation and observation are important source of discovering usability problems and crucial for the improvement of target artifact.

Figure 4 shows the sum of explanation and observation utterances in each experimental condition, which is normalized as a value per single task step. A two-way ANOVA for the utterances revealed a significant interaction [$F (1, 30) = 10.12$, $p < 0.005$] between the operation method and user interface. The simple main effect of the operation method was significant when the touch-panel interface was tested [$F (1, 30) = 27.58$, $p < 0.0001$]. For the user interface, the simple main effects were significant in both oral [$F (1, 30) = 45.24$, $p < 0.0001$] and manual [$F (1, 30) = 4.96$, $p < 0.05$] operation methods, and more utterances were made when the touch-panel was tested.

These results suggest that the oral operation method could contribute to the increase of utterances for explanation and observation, while the prototype of a touch-pane interface might have an influence on the facilitation of verbalization. The oral operation method avoids bringing users' attention overly to insufficient performance of an interactive prototype, and that will help users concentrate more on the other aspects of a target object to be tested. Consequently, the effects of the oral operation method in think aloud usability test still depends on types of interface. However, it would be worthy of consideration when it is crucial to draw more utterances in usability testing.
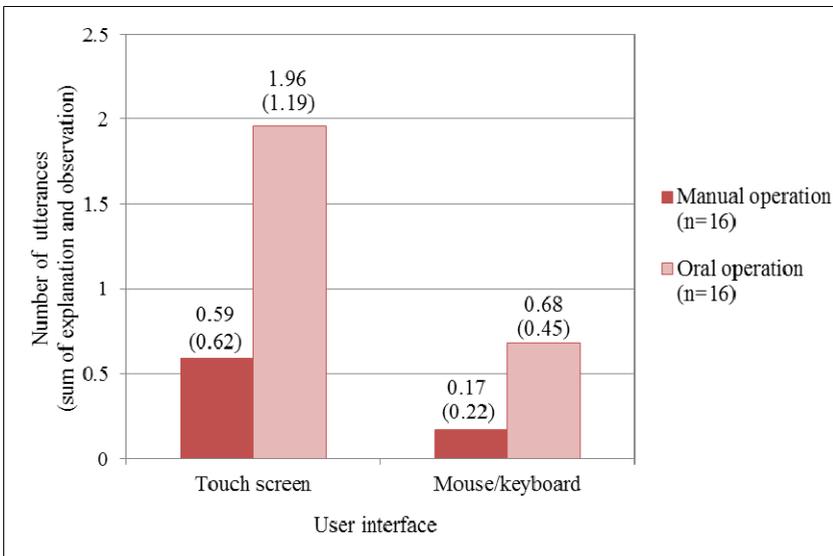

Fig. 4. Number of utterances for explanation and observation as values normalized per task step (SD in parentheses)

References

1. Vredenburg, K., Mao, J.-Y., Simith, P.W., Cary, T.: A survey of user-centered design practice. In: Proceedings of CHI 2002, Minneapolis, Minnesota, pp. 471–478 (2002)
2. Arnowitz, J., Arent, M., Berger, N.: Effective Prototyping for Software Makers. Morgan Kaufmann, San Francisco (2007)
3. Berger, N., Arent, M., Arnowitz, J., Sampson, F.: Effective Prototyping with Excel: A practical Handbook for Developers and Designers. Morgan Kaufmann, San Francisco (2009)
4. Hennipman, E.-J., Oppelaar, E.-J.R.G., van der Veer, G.C., Bongers, B.: Rapid and rich prototyping: Proof of concepts for experience. In: Proceedings of the 15th European Conference on Cognitive Ergonomics, Madeira, Portugal (2008)
5. Lim, Y.-K., Pangam, A., Periyasami, S., Aneja, S.: Comparative analysis of high- and low-fidelity prototypes for more valid usability evaluations of mobile devices. In: Proceedings of NordiCHI 2006, pp. 291–300 (2006)
6. Snyder, C.: Paper Prototyping: The Fast and Easy Way to Define and Refine User Interfaces. Morgan Kaufmann Publishers, San Francisco (2003)
7. Virzi, R.A., Sokolov, J.L., Karis, D.: Usability Problem Identification Using Both Low- and High-Fidelity Prototypes. In: Proceedings of CHI 1996, Vancouver, BC Canada, pp. 236–243 (1996)
8. Boren, M.T., Ramey, J.: Thinking aloud: Reconciling theory and practice. IEEE Trans. on Professional Communication 43(3), 261–278 (2000)
9. Van den Haak, M.J., de Jong, M.D.T., Schellens, P.J.: Retrospective versus concurrent think-aloud protocols: Testing the usability of an online library catalogue. Behaviour & Information Technology 22(5), 339–351 (2003)
10. Nielsen, J., Clemmensen, T., Yssing, C.: Getting access to what goes on in people's heads? Reflections on the think-aloud technique. In: Proceedings of NordiCHI, pp. 101–110 (2002)
11. Krahmer, E., Ummelen, N.: Thinking about thinking aloud: A comparison of two verbal protocols for usability testing. IEEE Trans. on Professional Communication 47(2), 105–117 (2004)
12. Cooke, L.: Assessing concurrent think-aloud protocol as a usability test method: A technical communication approach. IEEE Trans. on Professional Communication 53(3), 202–215 (2010)
13. Ericcson, K.A., Simon, H.A.: Protocol Analysis: Verbal Reports as Data, revised ed. MIT Press, Cambridge (1993)
14. Buur, J., Bagger, K.: Replacing usability testing with user dialogue - How a Danish manufacturing company enhanced its product design process by supporting user participation. Communications of the ACM 42(5), 63–66 (1999)