Differences in Task Descriptions in the Think Aloud Test

Lene Nielsen[1], Sameer Chavan[2]

[1]Center for Applied ICT, Copenhagen Business School, Howitzvej 60, 2000 Frederiksberg, Denmark. [2]MSC Software, Pune, India
ln.caict@cbs.dk Sameerhere@yahoo.com

Abstract. This paper analyzes and discusses the ways tasks are described and perceived in a remote Think Aloud (TA) usability tests session. The paper includes reports from a study and the problems encountered during a session of remote TA tests. The sessions were performed as synchronous tests, where the facilitator and observers received data and managed the evaluation in real time with a remote participant. It was done using a system with audio conferencing and remote application sharing. The analysis and discussion include both a task description perspective and a cultural difference perspective and hereby adds to existing knowledge of usability testing.

Keywords: Usability, Remote Think Aloud Test, Cultural Usability.

1 Introduction

Think Aloud (TA) is one of the protocols for usability testing where the participant is asked to think aloud or vocalize his or her thoughts, feelings, and opinions while interacting with a product. TA helps the test leaded to understand what problems the participant is facing and to question him depending on his think aloud. There has been criticism that TA does not simulate normal tasks as in "real" life, the users do not annotate each action with thinking aloud while they are doing tasks. Another critique is that in TA, the participants tend to forget the actual task and just verbalize the screen text rather then speaking what they are thinking. Another technique is Retrospective Think Aloud (RTA). With RTA, users do the tasks silently and then talk about what they did afterwards by watching a videotape of their own actions.

   Remote TA tests differ in many ways from the more traditional usability tests, but the main difference is the lack of contextual presence between the test person and the test leader, whether the test is performed synchronous or asynchronous (Dray & Siegel, 2004). The focus in this paper is exclusively on task description and perception in both traditional and remote usability tests. For both test settings, the task descriptions are written or oral instructions, but when remote usability tests are performed, the task is perceived by the test person in solitude and for the test leader there are – usually and with existing technologies - no possibilities of eye contact and perception of e.g. insecurity expressed through body language during the test session.

The literature that deals with the practical implication of the think aloud test describes the test set-up, but has limited descriptions of how the tests are to be implemented in detail. Looking through literature that introduces the traditional TA test methods to students ([2]; [4]; [6]; [7]; [9]; [10]; [11]; [12]; [13]; [16]; [17] it has only been possible to find four who include task descriptions [9], [4] [12]. In the literature of user testing there seems to be at least two different suggestions for task descriptions: One type favors descriptions of tasks and the second favors identification with a user in a specific situation. The variations of procedures is in this paper coined *the task focused procedure* and *the scenario focused procedure*.

2 Remote TA tests

Remote usability tests are gaining importance in recent days with the advancement of technology in desktop sharing. Remote tests provide a number of advantages such as:

- Remote testing substantially reduces costs, as it eliminates costs for travel and logistic.
- Testing can be done with a diverse pool of participants who are spread across the globe and not just local users.
- Specialists, who are otherwise not available as they have to travel to a test site, are now available as they can perform the test from their own location.
- Remote TA tests the participants in their real environment rather then a artificially created lab setup. The latter can result in anxiety and does not reflect real world working conditions.
- During Remote TA everyday disturbances can be encountered like a phone call or a colleague coming for a talk.
- There have been many studies showing that remote test can find more issues than lab tests and that the task completion rate in Remote TA can be higher than lab tests as the participants are not in stressed condition.

But there are also disadvantages with the remote TA test.

- The main disadvantage is the lack of contextual presence between the test participant and the test leader.
- The test leader does not get any clues of how many people are present together with the test participants, either just observing or actively assisting the participant.
- There can be many interruptions in the test. Often the participant gets phone calls, or some of his friends interrupt, there can be emails, and background noise. In our study an Internet connection went off and the test leader and the participant were disconnected.
- Communication with the participant by the test leader is another challenge as the test leader cannot observe body language. The test leader cannot gauge if the participant is tired, frustrated, or confused. Remote testing relies totally on what the participants say and their mouse movements on the screen. Only a skilled test leader can find the above issues by observing the screen movements.
- Not all participants are vocal.

- Performance is yet another factor. If the internet speed is slow then the response time is slow. Also, the desktop sharing tool requires some installations. And the test leader cannot remotely solve desktop problems at the participant end.
- Scheduling remote test also involves a lot of communication and mails and takes time to organize. If there are differences in time zones then it may result in non favorable times for the test.
- In a lab test you can easily video tape the screen movements, voice, and facial expressions. On Remote tests it depend on the tool whether it is possible to record all.
- Security is off concern too. The test leader cannot access if the screens are being captured. Some participants ask for the material or test links in advance. This will create security issue and also imbalance the test.
- If participants are to give feedback the participants prefer to send it back through email. These responses may change if the participants do fill in the questions immediately but take days to send it back. Unlike in a lab test where participants complete this immediately after the test.

3 The study

This paper report from a study of remote TA, the tests were conducted with participants from France, Germany, Japan, UK, and US. Participants had to call a toll-free number to join the test and login to an internet desktop sharing application. The participants were shown the task on a digital format similar to a power point presentation (PPT), where they were able to switch back and forth between the test application and the tasks. They were allowed to make a copy of the tasks in the PPT and paste them into the test application.

Since the test involved adding new features in the test application, 20 minutes of training was given. This did not involve showing how things were done in the application, but introduced the user to how data was arranged in the application and what things were possible.

After a few initial tests, there was a need to rearrange the task description in order to facilitate the users to more easily understand the tasks and spend time on the application rather then wasting time understanding the task. This resulted in that the participants did not receive the same task descriptions. The participants from US, Germany and UK received task focused descriptions while the participant from Japan and France received task descriptions closer to the scenario focused task descriptions, with an included small introduction.

Example of the task focused description:
- Open the program (name of program)
Under
- Work area (name and number of work area)
- Which is in application area (name and number of application area)
- Under domain (name of domain)
- Open and view properties for source code (name) under this program

Example of the scenario focused description:
We have started a trial for a treatment of an illness (described in length). Create a report and definition of the clinical trial report.

Five tests, one from each country, were investigated in length and observations were made on the type of questions the participants asked and the way they interpreted the tasks. In the analysis, these are compared to the way the tasks were described. Finally a brief introduction to culture is presented in order to explain some of the differences in the participants' performances.

4 Analyses

The differences in the task descriptions seemed to generate distinctively differences in procedure. When the task focused procedure is followed the participants follows the task description rigorously, but seems to have difficulties in understanding the overall ideas behind the system. This is especially noticeable when more of the participants show that it is not clear to them that they have performed the task.

The participants from US and Germany had no problems following the procedures. They did not question the tasks or demanded to understand the task. The US participant found it difficult to assess whether or not a task was completed. The German participant did clearly not know what she was doing.

The English participant did not know when the task is done. In the final interview she is asked what she found confusing in the system and answers: "I'm not really sure what is meant by (name of menu 1) and (name of menu 2). Maybe (name of menu 1) means… well I'm not sure what it means". The names of the menus are core to understand the whole system and shows how capable she is to follow procedures without understand the overall ideas behind the system.

When the scenario focused procedure is followed the participant finds it difficult to remember the task, but can more easily understand the idea behind the system. In the abovementioned example of the scenario focused task description, is included the reason behind the task "to create a report" and a subset of tasks as well as a row of other tasks follow, this seem to make the participants forget the overall goal of "creating a report". A consequence of the description is that in later tasks the participants get confused about what they are doing.

The Japanese and the German participants understood the tasks, but in later tasks, the test leader had to explain that the idea behind the tasks was to create a report. The participants forgot that, as the subsequent tasks did not include scenarios, but built on the scenario created for the first task.

The scenario focused procedure proved difficult for the participants as they had to remember more and also find the task in a lengthy text, but it gave the participants a better understanding of the overall idea behind the system. In the task focused procedure the test leader could see that the participant followed the tasks procedures, but it made the test leader loose track of whether the participant understood what he

was doing. This proves an even bigger problem, due to the constraints of the remote TA, where communication is restricted and tacit observations not possible.

As mentioned earlier the remote TA involved a short introduction/training of new features. This created complexity, as users started to compare the training information and expected the same behavior in the application.

The participants also forgot to think aloud after some time and read only the text on the screen rather then speaking out their thinking process.

There was a challenge for the participants if it involved a series of sub tasks, e.g. if the task description was big with multiple sub tasks, they would forget where to start.

5 TA tests and culture

The issue of cultural differences has for some time been discussed within the area of HCI. Most studies of cultural differences falls within design of user interfaces [3], [19], [1]. [14] report from a literature study of cultural issues in UI design and presents different studies that incorporates or studies cultural differences in the attitude to interfaces and to usability problems. Their study reflects upon the differences in attitude, but do not reflect upon, whether or not the methods used for testing the UI favors one culture over the other. Other reports of differences in cultural aspects in usability studies contain implications of effects on the results between similarity in culture of interviewer and test person and of no similarity [18], [15], [5].

An approach outside HCI is the psychologist Nisbett, who reports from several studies that the easterners view the world as holistic while Westerners see the world in an atomistic view. These differences in viewpoint might have an impact on the way tasks should be presented, with easterners in want of a more holistic view of a task and for westerners to reject holistic descriptions, to "get to the task". [8] (pp. 109). Returning to the above mentioned instructional literature of how to perform tests, the literature does not consider cultural differences in the test set-up or the task descriptions. Another reason why the literature does not discuss cultural differences is the underlying assumption, that a test result is only valid if task descriptions are identical for all participants.

In our study we did not encounter cultural differences in the understanding of the tasks and the sample is too small to make observations of this kind. We did encounter differences between the participants that can be explained with differences in culture. The European participants were more inclined to question the task and the system and to comment on them. They did not withhold their opinions and questioned the system immediately. The German participant found that the system she was used to worked better and encouraged the test leader to take a look at it. The French participant questioned the tasks and had an overall urge to discuss as can be seen in the following: "Test leader: What is happening? Participant: I don't know, it is your system." This was quite different for the US and Japanese participants who seemed more accepting of the situation. This, Nisbett explains with differences attitude towards either dialectical approach (Easterners) or logic reasoning (Westerners) [8] p. 37. It can be argued that the European culture favors an active debate and

individualistic attitude in an egalitarian culture, the latter might explain why the US participant did not have the same attitude as the Europeans.

An observation was made with some participants not present in this study. It was the same tasks, but the participants were Indian pilot participants. These participants succeed in solving the tasks more than other country participants. The reason might be that the Indian nature encourages investigation and the finding of answers. The participants did not give up and did not declare that they could not solve the task. They do not complain either. There was a tendency to learn the system immediately, rather then complaining on about a bad design.

6 Conclusion

In the task focused procedure the participants tend to complete the task as following a user manual. They do not think of real goals and report grammatical and UI standard based issues. E.g. in our study, one of the users was use to desktop systems and was trying to do a right click action on a web item.

The scenario focused procedure seems to apply a better understanding of the context to the participants and they are innovative in finding the solution. In our study these users reported interaction issues e.g. "This set of items should be in different tab".

If the task description is at length or more technical in nature, the TA creates a distraction in the minds of test participant. This creates a need to both a break down of length and a break down into simple language too.

The Remote TA creates both advantages and difficulties given the time and remoteness of participants, the TA also involves continuous reminders from the test leader to the participants. And finally the test leader also has to make his judgments by observing the participants' screen movements.

In summary, usability test results are dependent on how the task is written, the length of the task descriptions with sub tasks and of the type of cultural mix of the participants used.

References

1. Barber, W., Badre, A.: Culturability: The Merging of Culture and Usability. 4th Conference on Human Factors & the Web. Basking Ridge. (1998)
2. Jordan, P.W.: An Introduction to Usability. Taylor & Francis. London. (1999)
3. Marcus, A., Goul, E.W.: Crosscurrents: cultural dimensions and global Web user-interface design. Interactions, Vol. 7, 4 (2000) 32 - 46.
4. Molich, R.: Brugervenligt webdesign.: Ingeniøren Bøger. København. (2001)
5. Murphy, J., Howard, S., Kjeldskov, J.: Playing away from home - usability testing in a global world. CSI Communications. Vol. 29, 3 (2005) 18-24.
6. Nielsen, J.: Designing Web Usability. New Riders, Indianapolis. (2000)

7. Nielsen, J., Mack, R.L.: Usability Inspection Methods. John Wiley and Sons, New York. (1994)
8. Nisbett, R.E.: The Geography of Thought. Nicholas Brealey Publishing, London. (2005)
9. Preece, J., Rogers, Y., Sharp, H.: Interaction Design. John Wiley and Sons. (2002).
10. Preece, J., et al.: Human-Computer Interaction. Addison-Wesley, Harlow (1994)
11. Rose, K. Sørensen, N.: Brugervenlighed i praksis - en håndbog. Frydenlund. København (2004)
12. Rosson, M.B., Carroll, J.M.: Usability Engineering. Morgan Kaufmann Publishers, San Fransisco (2002)
13. Rubin, J.: Handbook of Usability Testing. John Wiley and Sons, New York (1994)
14. Shen, S.-T., Wooley, M., Prior, S.: Towards culture-centred design. Interacting with Computers. Vol. 20. (2006) 1-33.
15. Shi, Q., Clemmensen T.: Cultural Usability -The Effects of Culture on Usability Test. The 6th Danish Human-Computer Interaction Research Symposium. Denmark, Århus (2006)
16. Shneiderman, B., Plaisant, C.: Designing the User Interface. Pearson, Harlow (2005)
17. Snitker, T.: Breaking Through to the Other Side. Nyt Teknisk Forlag (2004)
18. Vatrapu, R., Pérez-Quiñones, M.A.: Culture and International Usability Testing: The Effects of Culture in Structured Interviews. Journal of Usability Studies. Vol. 1, 4. (2006) 156-70.
19. Yeo, A.: Cultural user interfaces: a silver lining in cultural diversity. ACM SIGCHI Bulletin. Vol. 28, 3 (1996) 4-7