Working Paper

What Kind of Information does an HCI expert want?
- on concurrent usability testing

By

Janni Nielsen
Carsten Yssing

No. 15-2004

Institut for Informatik

Handelshøjskolen
i København

Howitzvej 60
2000 Frederiksberg

Tlf.: 3815 2400
Fax: 3815 2401
http://www.inf.cbs.dk

Department of Informatics

Copenhagen
Business School

Howitzvej 60
DK-2000 Frederiksberg
Denmark

Tel.: +45 3815 2400
Fax: +45 3815 2401
http://www.inf.cbs.dk

What Kind of Information does an HCI expert want?
- on concurrent usability testing

Janni Nielsen
HCI Research Group
Department of Informatics
Copenhagen Business School
janni.nielsen@cbs.dk

Carsten Yssing
HCI Research Group
Department of Informatics
Copenhagen Business School
cy.inf@cbs.dk

ABSTRACT
Working with usability techniques, with focus on how a given technique enables data capture made us ask the question: What kind of information is it an HCI expert want from the user when conducting a usability test? We answer the question by discussing two techniques both relying on concurrent data. Think Aloud is one of the most frequently used techniques and almost an institution in itself. Eye-tracking is new in usability testing and still at an experimental level in HCI. We reflect critically upon the two obtrusive techniques. We discuss the usability of concurrent data capture, suggesting participatory analysis and retrospective verbalisation as a possible step in usability testing.

Keywords
Concurrent data, usability test, Think Aloud, Eye-tracking, mind, participatory analysis and retrospective verbalisation

INTRODUCTION
It has been pointed out that most methods often are taken "off the shelves" rather than chosen on the basis of pre-analysis and grounded decisions (Bødker and Sejer Iversen, 2002). In a survey on user centred design (Mao et al, 2001) the claim is that the most frequently applied methods/techniques are: simple prototyping, heuristic evaluation and usability test. Usability testing focus on user's interaction with the computer, and our interest is the way the user's eye travels around the graphic dynamic interfaces, the actions of the user (the navigation) and the way the user experience the interaction: What does the user see, why does the user do what she does and what does she think? In our search for techniques we have come to wonder about what information a usability expert is looking for when conducting iterative user testing. Among usability test it is especially the concurrent techniques which are of interest, and there are two techniques that stand out, but at each end of the usability line: Think Aloud (TA) and Eye-tracking. Think Aloud (TA) is old in the HCI business, it is one of the most popular techniques, used equally in industry and academia, and it is taken for granted that it gives access to user's thoughts. Eye-tracking is new to HCI, it is being applied rather exploratory but with great expectations and promises access to the user's visual interaction.

The short paper is organised around these two techniques. We describe the use of TA which promises access to cognitive processes, and introduce Ericsson and Simons classical distinction between Talk Aloud, Think Aloud and Retrospective Verbalisation. We discuss the data capture and point out that the interface is visual, not verbal. Any usability technique should be able to capture the visual interaction, and we turn to Eye-tracking (ET) as it gets beyond the verbalisation requirement and promises access to visual data capture. We describe ET and point out that logging of cursor, fixation of cursor, paths of cursor, saccades of cursor and the different interfaces capture do not tell us what the users feel, think experience. We raise the question: What kind of data does one gets access to when conducting concurrent testing? We suggest that other approaches are considered, and discuss the data capture that retrospective verbalisation enables. We propose a next step in our work which is cursor tracking and interface capture followed by participatory analysis.

THINKING ALOUD
Think Aloud (TA) is the most popular usability test, often referred to as the usability method. It is used equally by industry and academia (Boren and Ramey,2000, Clemmensen and Leisner 2002). TA is cheap, does not require heavy investments in technology, can even be conducted by non-usability experts and only requires 5-8 users. It seen as a straightforward technique, ready to use with proper handling (Hackos and Redish 1998, Molich, 1994), and has been given the credit of simplicity (Dix, Finlay, Abowd and Beale, 1997). Especially Jakob Nielsen has been a tireless promoter (1994, http://www.useit.com).

The understanding embedded in most studies is that the techniques allows us access to the cognitive processes, to mental behaviour and gives us insight into thinking.
By recording the verbal protocol, you will be able to "…detect cognitive activities that may not be visible at all" (Hackos & Redish 1998, p. 259). But caution has

Nielsen Janni and Carsten Yssing (2004) What Kind of Information does an HCI Expert Want? – on concurrent usability testing, WP. HCI Research Group, submitted to NORDICHI 2004, Tampere, Finland

been voiced because the technique puts a cognitive load on the user, which may interfere with the cognitive requirements of the interaction hence disturb the task. It has been argued that concurrent verbalisation is problematic because TA adds strain and cognitive load on the users ((Branch 1999)that users have difficulties in speaking and to speak aloud feels awkward (Preece, Rogers and Sharp 2002). Silence is the likely outcome of the situation, hence the need for the investigator to encourage the user to "keep talking".

The basis for TA in usability testing is the classic text on protocol analysis from 1984 by Ericsson and Simon. They discussed the use of introspective data in the study of task directed cognitive behaviour. It is the work of Ericsson and Simon which have reinstalled verbal data as a valid resource for understanding human cognitive processes. They did this by introducing the technique of Think Aloud. Their understanding was that most performance measures rely on responses that are psychologically indistinguishable from a verbal report, because some kind of verbal reporting is usually necessary to understand people's actions, even in very simple tasks.

Ericsson and Simon distinguished between three kinds of cognitive processes in their model;

- *Talk Aloud* is direct oral expressions of thoughts which already exist in verbal form
- *Think Aloud* is verbalisation of a sequence of thoughts, that are held in memory in some other form, e.g. visually
- *Retrospective verbalisations* of retrospective reports or thoughts not held in short term memory, i.e. explanations and descriptions.

They argue that a sentence is the verbal realisation of thought. The assumption is that everything we know has, at some point, gone through our short-term memory (STM), and we have been conscious of it. We can verbalise what we are learning while in the process of learning, and we can verbalise what we know if questioned shortly after the process of learning has taken place. This is because it is still retained in our short-term memory (talk aloud and think aloud). However, if there is a time span between learning and being requested to recall, we will produce descriptions and explanations (retrospective verbalisations) - not a report of our immediate thought, because the information from STM is lost. Ericsson and Simon were only interested in Talk aloud and Think Aloud, whereas they considered retrospective verbalisations too error prone due to the time lap and the reliance on users recall.

Critical Issues in TA

But what is it we get access to when asking users to think aloud? Does it really give us access to what goes on in people mind? Teaching graduate computer science students the TA test techniques, and requiring them to reflect on their experience raised a number of issues. Students complain that 1) they think faster than they can speak, 2) thought processes are much more complex than can be verbalised, 3) having to think aloud interferes with their interaction with the interfaces and the task and 4) thinking aloud does not come naturally. (Nielsen, Clemmensen and Yssing 2002a).

The assumption that performance measure have to rely on some kind of verbal data, and that a sentence is the verbal realisation of the thought has been questioned (Nielsen, Clemmensen and Yssing 2002b). The sentence that the user speak is a verbal realisation of thought, but there is not a 1:1 relationship between thoughts, actions and the spoken words. Using TA requires the user to shift focus in attention from giving sense to that which is perceived and does not exist in verbal form - to that of constructing sentences or words, and then expressing them aloud. TA requires perception and actions to be transformed to talk. Even if the speech is immediate and runs concurrently with the thoughts – user's attention has to shift focus from understanding to verbalisation (Nielsen and Yssing, 2003). As a consequence the process of understanding is interrupted, because attention keeps changing object. TA may result in verbal overshadowing (Schooler, Uhlsson and Brooks 1993) and we do not get access to mind.

Besides the user is interacting with net based multi modal interfaces. Colours, layout, forms, animations, video clip and endless jumps through links all interact and it is visualisations which is the main feature. Hence the interaction is mental and based in visual perception, and thoughts are not mainly verbal and directly accessible in oral speech, but percepts which are, to a large extent, tacit (Polanyi 1967).

EYE-TRACKING

In our search for techniques we took a closer look at Eye-tracking. It seems to get around the verbalisation problem, yet it captures concurrent data and the visual interaction. Eye-tracking builds on a mind-eye hypothesis and expectations are high. However, the technique is still at an experimental level within the HCI field. Where it has been used by to investigate cognitive workload and scanning behaviour (Salvucci 1999), interface and screen design (Ellis and Candrea 1998), to determine the position of visual feedback on the screen. (Rauterberg and Cachin 1993), and visual search on pull down menus (Byrne et al 1999).

There are different ways of conducting eye-tracking. One is the well known head mounted system with cameras that the user has to wear. One camera shows the scene that the user is looking at, the other camera is tracking the eye movements. A less obtrusive system is a remote eye-tracking system where the user works directly on the computer. A camera lens, mounted besides or under the

computer, is focused upon the user's eye tracking the pupil, most often with an infrared source.

Eye tracking makes it possible to follow the visual fixations and scan-paths of the user. It captures a user's visual focus of attention on a visual display unit – through special hardware and software. It registers x/y location and pupil size/border line information. It is especially the jerky movements of the eyes (saccades) which are followed by fixations (x,y location) and combined into scan-paths, and the smooth pursuit of eyes tracking slowly moving objects which are analysed (Goldberg and Wichansky 2003, Ellis and Candrea, 1998, Heyhoe, Shrivastava, Mruzcek and Pelz 2003).

Critical Issues in eye-tracking

The cost of acquiring, learning to operate and the maintenance of an eye tracking system have to be considered. Prices range from just around 18.000 Euro to 100.000 euro for a complete eye tracking system. Besides the need for calibrations are high, even with a remote tracking system. Hence the user is instructed to maintain a relatively stable head position because movements during tests require recalibration. But users have difficulties in keeping their head still, and the recommendation is to conduct "recalibration every few minutes" (Goldberg and Wichansky 2003). There are problems with pupil/borderline registration because the pupils contract and expand in response to light, there are large individual differences in eye tracking results, and users with glasses and contact lenses are often excluded. Even the analysis done by the computer software creates problems and "the investigator is strongly encouraged to review the (software) created fixations against images of viewed displays to ensure that the fixations are valid". Besides, the raw data has to be " aggregated off-line by the investigator into meaningful behavioural units of fixations and saccades", and the amount of data is enormous and the task is extremely time consuming.

But what do we get access to with eye-tracking? Ellis and Candrea (1998) used eye tracking to test a website with a two column lay-out with both text blocks and images. Links were embedded in the text blocks, but also images could be links to video sequences. They redesigned the web page in three variation: One version had many links, in the second version they replaced graphics with text blocks and in the third version they made the page look like a book page. They name it "dense-text". Their analysis of eye movement tracking showed that the dense text version scored highest on their usability test. But they conclude, somewhat surprised, " despite it's potentially superior usability, dense-text was the lowest rated of all the lay-outs by our testers". This data was not registered by the eye-tracking system, but was collected after the test when the users were asked to rate the different interfaces.

The surprise that Ellis and Candrea voice points to a serious problem with eye-tracking. It does not give access to mind. Interestingly, the title of their paper is "Windows to the soul"? qualifying it with "eyes reveal a great deal about a person's feeling and behaviour". But registration of eye movement do not tell us anything about the soul. Eye-tracking only register movement – not what goes on inside the human being. It does not give us access to mind, and the capturing does not tell us anything about users intentions. However, Ellis' and Candrea's data show that we need to go beyond the mere concurrent testing and follow up with further investigations. Though the hypothesis lying behind eye-tracking is the eye-mind assumption – we only have logging of cursor, fixation of cursor, paths of cursor, saccades of cursor and the different interfaces capture. We do not have access to mind and we do not know what the users feel, think experience.

FROM CONCURRENT TO RETROSPECTIVE USABILITY TESTING

Our initial attempt to solve the problem with the obtrusive concurrent techniques was to look for an unobtrusive data capture technique which would also allow us to deal with the multimodal interfaces. The solution was a software tool which enables concurrent data capture with cursor tracking an interface capture. In this way we can look at what the use looked at, and we can see how the user interacted. However, cursor motion does not necessarily track where user's visual attention is, s/he may forget to move the cursor because something on the multimodal interface disturbs or pleases her, or even be lost in daydreaming.

This brought our attention to the third level in Ericsson and Simons model: retrospective reporting which are thoughts not held in short term memory, i.e. explanations and descriptions. Ericsson's and Simon's argument is that if there is a time span between learning and being requested to recall, the user will produce retrospective verbalisations - not reports of their immediate thought, because the information from STM is lost. And retrospective reporting is more error prone because it relies on user's subjective recall – not on "hard facts", and subjective verbal data are not considered valid.

Retrospective verbalisation and participatory analysis

User's mind cannot be observed or registered. The only way to get access to user's experience is by probing the user: What does s/he see, why does s/he do what s/he does and what does s/he think? Our suggestion is to combine the capture of cursor tracking and interface with retrospective reporting in a participatory analysis. This technique will get us beyond the "total subjective recall". Because one of the unique advantages with cursor tracking and interface capture is that there is a recording of the actual actions of the user. It can be replayed and shown to the user. What the user sees is what s/he saw

while working on the test, only the screen capture includes user's movements with cursor. The recording of the screen and mouse can be stopped/resumed at any given point. During a stops a recording of the user's comments and reflections unfolding as a consequence of the probing will be captured on top of the frozen image. Thus, in the final analysis, the investigator is in possession of concurrent data in the form of 1) an uninterrupted recording of user's interaction with the interface during the test, and a retrospective reporting in the form of 2)a recording of the user's interaction with the interface with interruptions, and with a voice over.

Closing comments

The capture of interface with cursor tracking combined with participatory analysis seems promising because the processes of insight that runs associatively while the user interacts with the computer application may become partly explicit, and not be a total subjective recall. We call this technique Mindtape (Nielsen and Christiansen) and the replay triggers a running commentary at the same time as the events take place on the capture. These images may enhance the user's access to, and help recall, the thought processes that took place. The verbalization flows easy with the actual sequence of events structuring – not the users memory. This is important, a Mindtape is structured by the actual user-computer inter-actions as they unfolded during the test session. It is not the users memory which controls the recall, but the actual events.

Naming it eye/cursor movements, or eye-mouse correlation escape the fact that it is the hand the user has to move – represented on the interface through the cursor, hence coordination with the eye.

REFERENCES

Boren M. Ted and Judith Ramey(2000) Thinking Aloud: Reconciling Theory and Practice, |*IEEE Transactions on Professional Communication,* vol. 43, no. 3, September , p. 261-278

Branch, J.L.(1999) Investigating the information-seeking processes for adolescents, *Library & Information Science Research,* vol.22, 4, 371-392

Byrne Michael D., Anderson John R., Douglass Scott and Michael Matessa (1999) Eye Tracking the Visual Search on click-Down Menus, Proceedings of HI99, Pittsburgh, 402-409

Clemmensen, T and P. Leisner(2002) Community knowledge in an emerging online professional community, Proceedings of IRIS25, Denmark,

Dix A., Finlay J., Abowd G. And R. Beale(1997) Human-Computer Interaction, Prentice Hall, 1997

Ellis Steve and Ron Candrea(1998) Windows to the soul? What eye movements tell us about Software Usability, Proceedings of the 7th Annual conference of the Usability Professional Association, Washington DC, 151-156

Ericsson, K.A. and Simon, H.A. s984) Protocol Analysis. Verbal reports as data, Cambridge, Massachusetts

Goldberg, Joseph H. And Anna M. Wichansky(2003) Eye Tracking in usability Evaluation: A Practitioner's Guide, in Hyöna J., Racdach R. And Deubel H.(eds.) *The mind's eye: Cognitive and Applied aspects of eye movement research*, Elsevier Science BV, chapter 23, 493 - 516

Hackos, J.T. and Redish, J.C.(1998) User and Task Analysis for Interface Design, Wiley, USA

Hayhoe Mary M., Shrivastava Anurag, Mruczek Ryan and Jeff B. Pelz(2003) Visual memory and motor planning in a natural task, *Journal of Vision* (2993), 3, 49-63

Molich Rolf (1994) Brugervenlige edb-systemer, Nyt Teknisk Forlag, København

Nielsen, Jakob: http://www.useit.com/

Nielsen, Jakob(1994) Estimating the Number of subjects Needed for a Thinking aloud Test, *International Journal of Human-Computer Studies*, 41, 3, 385-397

Nielsen Janni and Nina Christiansen(2000) Mind tape: A Tool for Reflection in Participatory Design, in Proceedings, Conference Participatory Design, New York 2000

Nielsen Janni, Clemmensen Torkil and Carsten Yssing(2002a) Getting access to what goes on in people's Heads? NordiCHI proceedings, Denmark, 101-110

Nielsen Janni, Clemmensen Torkil and Carsten Yssing (2002b) People's Heads – People's Minds, Proceedings of the APCHI 2002 (5th Asia Pacific Conference on Computer Human Interaction): User Interaction technology in the 21st Century, 2002, China, Science Press, Beijing (897-906)

Nielsen Janni and Carsten Yssing (2003) Getting beyond the disruptive effect of Think Aloud, in Hertzum Morten and Simon Heilesen(eds.) *Proceedings of the third Danish Human-Computer Interaction Research symposium.* Datalogiske Skrifter, Roskilde University, 67-71.

Preece J., Rogers Y. And Sharp H.(2002) Interaction Design: beyond Human-Computer Interaction, John Wiley & Son

Polanyi, Michael(1967) Logic and Psychology, *American Psychologist*, no.23, 27-43

Rauterberg Matthias and Christian Cachin(1993) Locating the primary attention focus of the user, in Grechenig T. And M. Tscheligi(eds.) *Human Computer Interaction* (Lecture Notes in Computer Science 733, 129—140) Berlin, Springer

Salvucci Dario d.(1999) Inferring Intent in Eye-based Interfaces: Tracing Eye Movements with Process Models, Proceedings of CHI 99, Pittsburgh, USA, 254-261

Schooler Jonathan W. Uhlsson Stellan and Kevin Brooks(1993) Thoughts beyond Words, *Journal of Experimental Psychology,* vol. 122, 2, 166-183

ACKNOWLEDGEMENT

Thank you to the HCI research group at Copenhagen Business School

This work was made possible through the Danish Research Council grant ref. 9900403.