

Using the *think aloud* for affective evaluations in the lab

A work-in-progress

Asimina Vasalou

Intelligent Systems and Networks Group
Imperial College London

Tanja Bänziger

Geneva Emotion Research Group, University of Geneva

Overview

The think aloud protocol originated from cognitive psychology (Ericsson and Simon, 1993). One of its many applications was to understand the processes underlying problem solving. Once trained in thinking aloud, users were 'left alone' with minimal, if no interference, to verbalise their actions. Thinking aloud has since been used for usability and is one of the most widely applied methods in the field. The reason for its success is that it gives insights on the 'why' behind a particular usability problem. Inevitably, the requirements of usability testing led to an extension of the think aloud from its original form (Ericsson and Simon, 1993). New procedures have been put in place for dealing with technical interventions for example (Boren and Ramey, 2000). Most importantly, the role of the facilitator and the user has been considered under a new framework which clearly defines the user as the speaker, giving information about the interface, and the facilitator as the listener, who is there to learn while partaking little in the verbal protocol.

Affective evaluation and the think aloud

With technologies that move beyond task performance and into our social territory evoking emotional experiences, the term usability has also grown to include user experience. Facing this challenge, usability experts have been called to consider whether old methods suffice for measuring and understanding this new kind of information. Our objective in WP9 was to find out whether the think aloud elicits more than just a cognitive reflection of a user's interface activities giving us information about the user's experience. We set out to answer a series of questions on this topic.

During the usability session, users are given the role of the expert, the one who knows best and speaks out. While using an application that has the ability to provoke an emotional experience, can thinking aloud provide a platform for expression? If it can, this expressive information may offer signals that allow us to tap into the user's experience. Nonetheless, people are known to often regulate their expressions to others. One can imagine that a user encountering a severe usability problem may suppress his or her expression in consideration of the facilitator present. Additionally, a positive expression may be ironic, signifying in fact a negative attitude. Given this concern, how reliable are expressions shown in the dyadic communication between the user and the facilitator and hence what kind of methodological considerations should usability experts keep in mind if using this method? Finally and most crucially, by using the think aloud in the lab, can we as usability experts, get data that goes beyond performance problems with the interface and that leads us to a deeper understanding of a user's experience? Many consider the lab to be an artificial setting that is most appropriate for uncovering performance issues only. Despite the strength of this argument, the lab is often the only solution available. Can it thus be used for evaluations that go beyond its a priori performance role?

The study

We set out to answer all of the questions posed above in a study conducted at the Geneva Emotion Research Group in July 2006. 20 fluent French speakers between the ages of 20 and 40 took part. The application evaluated was Yahoo! Avatars France. Yahoo! Avatars provides users with tools to personalise an avatar which is then broadcasted into their social networks. On arrival, users were first trained on how to do the think aloud. They were then given two sequential tasks to complete requiring them to use Yahoo! avatars. In the first task, they rehearsed an upcoming date they were looking forward to. In the second task, they used the application to send their mother a hypothetical gift i.e. a week long holiday in the Greek islands to celebrate her birthday. During the usability session, video and audio recordings of participants were made. Participants' activity onscreen was also recorded.

Analysis and preliminary insights

The data was replayed by us post to the experiment. During this processes, we were attentive to what was happening onscreen, users' verbalisations of the incident and users' facial and vocal non-verbal

expressions. Using these three guiding cues, two kinds of incidents were extracted for further analysis concluding to a total of 320 incidents: *usability problems* e.g. ‘Something happened onscreen that will not allow me to efficiently complete my task’ and *user experience* e.g. ‘None of the faces look like me... (sigh) but it will do’, e.g. ‘I selected this haircut (laugh out loud)... yes this haircut is funny’.

Although we are in the early stages of data analysis, we mention some of the preliminary conclusions we have drawn that address our three questions.

Are users expressive while thinking aloud?

Not all, but many users displayed non-verbal expressions while doing the think aloud. At the moment we are undergoing a rating study to ‘objectively’ verify whether these non-verbal expressions are perceivable by others who cannot draw the links and inferences derived from the context. During the sessions, we discovered that the most expressive users had altered the think aloud without instruction by including in the protocol their opinions and preferences. This alteration was unexpected but not surprising: an application such as Yahoo! Avatars which may have emotional meaning to its users may naturally prompt them to express the richness of their interaction which goes beyond satisfying their task, towards their personal preferences and experiences.

How reliable are these expressions?

Although in many instances, the user’s action onscreen explained his or her expression, we found that expression as an accurate signal of experience has to be approached with some caution. For example, users were ironic when expressing dislike. Even more, they joked about options they found in bad taste and usability problems which are expected to provoke irritation were at times fun. In this part of the analysis, we hope to conclude with a list of these methodological considerations that usability experts can draw on during their own analysis.

Can we retrieve user experience data with this method given its artificial setting?

Our experience with this method leads us to believe that usability experts can expect to obtain more than performance information. A few representative examples of our findings are listed here:

- Users used their own physical appearance as a guide for choosing their face. Others chose faces on the basis of their mood at the time.
- Users customised their avatars to send intimate messages; they had fun imagining others’ reactions when seeing their avatar.
- Users drew on personal values to make their customisation choices e.g. choosing places they would like to visit in real life; rejecting places they dislike in their offline life.

Boren, M.T. and Ramey, J. (2000) Thinking Aloud: Reconciling Theory and Practice. IEEE Trans. Prof. Comm., 43, pp 261-278

Ericsson K A and Simon H. A. (1993). Protocol Analysis: Verbal Reports as Data MIT Press, Cambridge, MA.