A vertical strip on the left side of the page shows a close-up of a human eye. The eye is looking slightly to the right. The iris is a light brown color, and the pupil is dark. The eyelashes are visible at the top and bottom. The background of the eye is a warm, orange-yellow color.

Guidelines for Using the Retrospective Think Aloud Protocol with Eye Tracking

September, 2009

Short paper by Tobii Technology

Not sure of how to design your eye tracking study? This document aims to provide you with some basic guidelines and advice regarding what to consider when designing a usability study involving retrospective think aloud and eye tracking.

1 Table of Contents

- 2 What is Retrospective Think Aloud and Why is it Suitable for Eye Tracking Studies? 3
- 3 What Should I Consider When Designing My Eye Tracking Usability Study?..... 3
 - 3.1 How Do I Select the Participants? 3
 - 3.2 How Do I Create Tasks? 4
 - 3.3 How Do I Design The Study So I Get the Kind of Data I Want? 4
 - 3.4 How Can I Collect The Participants Thoughts and Opinions? 4
 - 3.4.1 Pre- and Posttest Questions 4
 - 3.4.2 Interview Protocol 5
 - 3.5 How Can I Make Sure That the Testing Will Run According To Plan? 5
 - 3.6 What Else Should I Consider When Designing the Study? 5
- 4 What Special Considerations Do I Need to Have When Designing a Cued RTA Study? 6
 - 4.1 When Should the Retrospective Session be Positioned Within the Test? 6
 - 4.2 Which Cue Should Be Used? 6
 - 4.2.1 Using a Gaze Plot as a Cue in an RTA Interview 7
 - 4.2.2 Eye Gaze Video as a Cue 7
 - 4.2.3 Other Cues 8
 - 4.3 How Can a RTA Session be Recorded? 8
- 5 What Should I Do Once the Testing Is Completed? 8
 - 5.1 How Can the Data Collected be Used? 8
 - 5.2 How Can the Findings be Presented? 9
- 6 Recommended reading 10

2 What is Retrospective Think Aloud and Why is it Suitable for Eye Tracking Studies?

In website usability testing, participants' eye movement patterns and fixations can be used to gain an understanding about how they complete a task and how difficult or easy it is to complete. An eye tracker produces objective numerical data as well as visual results that can provide additional insight into the participants' behavior and the interface tested. Ideally, when evaluating usability, eye tracking data should be combined with other methods and data because eye movements cannot always be interpreted correctly without the participant providing context to the data. For example, longer fixations can mean a participant found a particular area interesting, but it can also mean that he or she found the area difficult to interpret. Hence, it is important to supplement eye tracking data with additional information gained from the participants about their experiences.

Think aloud methods are often used when trying to detect usability problems. The most successful form of think aloud to use together with an eye tracker is a retrospective think aloud methodology (RTA), which means that participants verbalize their thoughts after completing a task or a set of tasks. RTA allows the participant to complete a task on their own and in silence. The other most common form of think aloud is the concurrent think aloud method (CTA) where participants verbalize their thoughts during task completion. Both methodologies have their own sets of limitations and problems, but using CTA in combination with eye tracking has proven to be less suitable as participants produce different eye movements compared with the RTA method. Differences in eye movements are for example caused by users looking away from the screen when describing something to the researcher in the CTA method or by focusing on certain areas of the screen while describing their thought processes regarding that area. Therefore, the RTA method is the recommended method when conducting usability tests where also objective eye movement data will be analyzed.

However, RTA is not a problem-free method; as the participant is asked to remember their experiences rather than provide their immediate response when doing the task, important things might be forgotten or imperfect memories become the source of fabricated usability problems. In order to aid the participant's memory, and to minimize the risk of fabrication, cued RTA methods have emerged. In a cued RTA, the participant is presented during the interview with a replay of their user journey while completing the task or tasks. The replay is then used to help cue their think aloud. Examples of commonly used cues are screen shots of the task they completed or a video playback of the participant's user journey or eye movements made during the same user journey.

3 What Should I Consider When Designing My Eye Tracking Usability Study?

How you design your test will largely depend on the goal of the study. It is important that this is thoroughly thought through before deciding on tasks or choices of methodology. For example, are you trying to get an idea of what people's first impression of a website or software are? Do you have specific problem areas you want to test? Or, are you looking to see what kind of problems users have while completing a task? The purpose of the test should be kept in mind while designing the study to avoid spending time and resources on things that turn out to be unimportant or that do not provide any valuable insights. Designing your study involves selecting tasks, determining your participant criteria, creating an interview protocol, and designing the RTA.

3.1 How Do I Select the Participants?

Although opinions vary about the optimal number of participants in a qualitative usability study, usually 5-15 participants (depending on the test) are needed. Since you will probably try to keep the number

of participants as low as possible, it is important to ensure you only recruit participants that fall into your specific user group or groups. If you are testing a website aimed at women between 15 and 25 it might be important to make sure that your participants fall into that category instead of having eight participants with completely varied demographic background. You might however occasionally want to verify how different user groups perceive the same site or software. In those cases it is important to recruit enough participants to represent each user group. This all depends on what you are testing and, again, the goal of the study. However, selecting the right user group(s) and ensuring that there are enough participants to provide valid data is very important as selecting too few or the wrong user group will jeopardize the validity of the findings generated by the study.

3.2 How Do I Create Tasks?

Select tasks that represent the problems you are attempting to solve, or issues you want to learn more about. For example, if you want to test how a registration process on a website works, have the participants go through the entire registration process. Or, if you want to figure out the users' general impressions about a website or software, have them complete common tasks for which most users use the website or program. Try to formulate the tasks in a way that allows the participants to relate to them on a personal level, i.e. as if they were actually doing the task for themselves and not for the sole benefit of the study. That way they will be more likely to point out things that they would notice and comment upon in a real use situation instead of only going through the task as an exercise. Using the example given above where you want to investigate a registration process, the task could be "You have heard about '*music service A*' and are curious about what it is. To find out more you decide to sign up for it. Use '*site a*' to do so." Print the tasks on a sheet of paper that you give to the participants so they do not need to memorize the tasks. As much as possible should be done to minimize the participants' cognitive load not caused by performing the actual task on the site or software.

3.3 How Do I Design The Study So I Get the Kind of Data I Want?

Decide how you want to analyze and use the eye tracking data from the recordings when designing the study and tasks. Depending on what you will use the eye tracking data for in your analysis or report, you might have to design your test and tasks accordingly. Eye tracking data can be used on many different levels, for example only as a cue to help participants remember their behavior, to create visualizations describing user behavior, or for getting objective quantitative data describing behavior. All of the above can be combined with RTA since the RTA method enables you to get both objective eye movement data as well as qualitative interview data. Before you start your test you should always test the webpages or software you intend to use in the test on the computer and on the eye tracking equipment you will use. Also produce visualizations and statistics to make sure you can get and produce the kind of results you need for your report and analysis.

3.4 How Can I Collect The Participants' Thoughts and Opinions?

In a usability study you generally ask questions either during the testing or before and/or after the testing is over. Here follows some advice about how to structure the questions you want to ask.

3.4.1 Pre- and Posttest Questions

It is not always necessary to have a pre- and/or a posttest questionnaire. As with the study as a whole, the purpose of these two questionnaires should be considered before initiating the design. Some data might already have been collected as part of the recruitment, but it might be useful to verify that the information collected during recruitment was accurate. The information collected in pre- and posttest questionnaires will need to be analyzed which requires time and resources. Hence, ensure they are relevant to the current study and that the data is useful in the overall findings analysis.

Example of things you might want to ask prior to testing include typically demographic details, such as age and gender, questions regarding the participant's IT proficiency, and use of the product/service or similar

products/services that the study aims to evaluate. If the participant is not informed about for whom the study is being conducted before the interview begins, the pretest questionnaire is the last possibility to gain non-biased information from them on their opinions about the brand or service you are testing and its competitors before the software or website (which usually reveals the company doing the testing) is introduced.

After the test is completed it is common practice to give the user a posttest questionnaire. The system usability scale (SUS) is a well-known set of questions used in such a document, but several other options are available, e.g. the after scenario questionnaire (ASQ). Both examples mentioned are very general and can be used for almost any kind of website or software. However, they do not provide questions about specific aspects of sites or software, e.g. certain functionality or specific pages. If you cannot find a pre-made questionnaire that fits your study, you can simply ask the questions you think are important. If you are conducting an RTA interview you may not need a very extensive posttest questionnaire and can instead include the questions during the interview. However, it might be useful to ask some key open-ended questions in a posttest questionnaire as this allows the participants to reflect and respond using their own words. Even though much information can be collected using RTA, some insights might come to the participant when they, in writing, are trying to formulate an answer to a question that was previously discussed during the RTA.

3.4.2 Interview Protocol

It is impossible to prepare for all the unexpected that might occur during usability testing. However, expect the unexpected and try to plan ahead whenever possible. The best way to do this when doing usability testing is to write an interview script. This allows the interviewer to keep track of the entire procedure and control that what is said to each participant is roughly the same for all so that no one is influenced by different instructions. Having a set interview protocol will also help you keep a timeline. While it is impossible to know how much people will talk, an interview script will allow you to keep track of how much is left to be covered during the remaining interview time and pace the interview accordingly.

Depending on the purpose of the study, the level of detail may vary when writing the questions to ask the participants. If you are doing a strict study to collect mainly quantitative data, the demand for a detailed specification of what is to be asked might be very high. In comparison, if you are doing a qualitative usability study, the script might only include topics that need to be covered, allowing the interviewer to investigate these in detail using open-ended questions and to follow up things said by the participant that are of particular interest. In addition, participant observations during task completion might compel the interviewer to ask questions about issues that appeared to be problematic to the participant when doing the task but that the participant did not bring up spontaneously during the following post task discussion. However, independent of what kind of a study you are conducting, specifying in the script what is to be asked or discussed during the RTA interview is important to ensure that similar topics are covered with all participants.

3.5 How Can I Make Sure That the Testing Will Run According To Plan?

Prior to starting the actual study with “real” participants, you may want to run one or two pilot tests to make sure your script, tasks and cues function properly. This way you can sort out any unanticipated problems or misunderstandings before the real testing begins. Ideally the participants in the pilot tests have the same characteristics as the participants used in the actual testing although this is not always necessary to test equipment, structure and timing. A pilot test is also useful for the interviewer to practice using the equipment, fine-tuning the tasks, questions and procedures.

3.6 What Else Should I Consider When Designing the Study?

Some other considerations include the role of the interviewer or experimenter as well as relevant regulations. Some testing might be done while the participant is left alone in a room to work in peace, other times you might have to be right next to them. In addition, a screen could be set up to provide a live feed of

the participants' eye movements allowing the interviewer (or clients viewing the test) to watch how the participant interacts with the interface in real time. Such a set up could provide the interviewer with information on which he or she could base questions to ask the participant after the task which might otherwise be missed. Another consideration is whether you need the participants to sign consent forms. If you are audio- and video recording the participants and plan to present it to members of a third party you may need to get the participants' signed consent. In some locations, health and safety regulations might require you to inform the participants about the risks of participating in a study before the interview begins. These are just some extra matters to consider before conducting a usability test.

4 What Special Considerations Do I Need to Have When Designing a Cued RTA Study?

4.1 When Should the Retrospective Session be Positioned Within the Test?

The RTA interview can be conducted in two different ways, either directly after the participant has completed a task or after the participant has completed all or a few tasks. Which method you choose depends on, among other things, how much time is usually required to complete the tasks, if the tasks are a part of a larger process and if you want to evaluate the entire process before interviewing. Conducting the interview after every task might also make the participant more aware of being eye tracked and tested which in some cases can have an impact on how the tasks are completed. This is generally a good thing to try out during pilot testing or to think through before starting the real test.

4.2 Which Cue Should Be Used?

Depending on the purpose of the testing and what kind of stimuli is to be used, different kinds of RTA methods should be considered. The quickest and easiest method is to use RTA without any cue, i.e. to just interview the participant after completing the tasks. However, this method does not encourage the participant to talk and remember problems encountered while completing the tasks. Using a screen recording video as a cue in the RTA interview will encourage the participant to talk more than using no cue at all. Research reveals that when using participants' own eye movements as a cue, participants tend to talk more and remember more usability problems than when they do not see their own eye movements. The table below is a comparison of the outcome of a study made by Tobii Technology (available as a whitepaper on tobii.com) comparing the outcome of different RTA methods.

No cued RTA	Video cued RTA	Gaze plot cued RTA	Gaze video cued RTA
Produced significantly less data (comments and words) than any of the cued RTA methods.	Stimulated participants to produce ' manipulative ' and ' visual ' comments Stimulated participants to comment on usability problems regarding ' layout ' and ' navigation '. Produced less data (comments and words) than eye movement cued RTA methods.	Stimulated participants to produce ' visual ' and ' cognitive ' comments. Stimulated participants to comment on usability problems regarding ' terminology ' and ' comprehension '. The second best performing RTA method; produced the second highest number of words and comments.	Stimulated participants to produce ' visual ', ' cognitive ' and ' manipulative ' comments. Stimulated participants to comment on usability problems regarding ' layout ' and ' data entry '. The best performing RTA method; produced the highest number of words and comments.

When testing dynamic content, e.g. a flash animation or software that allows a lot of user manipulation, a static cue is not possible and a screen video with or without eye movements should be used instead. Moreover, if the purpose of the testing is to evaluate the static, visual aspects or terminology of a site,

a static cue would be the preferred choice as this allows the user to spend more time focusing on each individual page rather than their own interaction with the pages.

4.2.1 Using a Gaze Plot as a Cue in an RTA Interview

A gaze plot is a static image showing the pattern of how a participant viewed a page (see image below). If certain settings are selected in the software producing the gaze plot, it can also show in which order the participant fixated on different areas of the page. The level of transparency of the fixation dots can often also be regulated by the software. By choosing opacity of 50% the participants will be able to see what is behind the dots which then allow them to identify what it was they were looking at.

If you choose to conduct a gaze plot cued RTA, you will be showing the participants images of all the eye movements they made on each webpage. Showing a static image allows participants to talk about any problems encountered on each webpage, and the interviewer to ask questions. Once the interviewer feels that the participant has exhausted all their comments on a page, the interview can proceed to the next page.

Using a static gaze plot is suitable if you want to evaluate details like site terminology as it allows participants to read and reflect on what they see during the RTA in their own pace. In addition, the image being static also allows participants to point at and discuss certain areas of the page without it being replaced by another page during the discussion as it might if video was used.



A gaze plot is not ideal for evaluating a dynamic website or software where user manipulation is an important part of the interface. This is because interactions cannot be successfully represented on a static image. In addition, the gaze plot might not be the best option for evaluating a process as showing the images one after another does not represent the flow of the actual process and important aspects such as transitions between pages or stages might not be explored or commented upon.

4.2.2 Eye Gaze Video as a Cue

An eye gaze video shows a video recording of the participant's interaction with the page on which the eye movements and fixations are also shown (see image on the right). As with the static gaze plot, the opacity of the fixation dots can be altered even when an eye gaze video is used.



If you choose to replay a video of the participant's eye movements during the RTA interview it is important to consider the video settings. In Tobii Studio 2 the user can control some aspects of the video, such as play/pause, rewind/fast forward and playback speed. These features allow participants to pause if they need more time to say what they want to say or to fast forward if they feel they

have said everything they need to during a specific section.

Using an eye gaze video replay as a cue has proven useful for identifying a broad spectrum of usability issues. As it shows an entire user journey in sequence, it is particularly suitable for evaluating dynamic content or processes. Additionally, as it shows eye movements on the screen as they happened, the participant is presented with context around an interaction with the interface and might therefore recall the reasoning behind an action that they might not have thought of if they only saw the static gaze plot. Because the continued process is shown, comments collected during RTA using gaze plot video often mention different aspects of the data entry and layout.

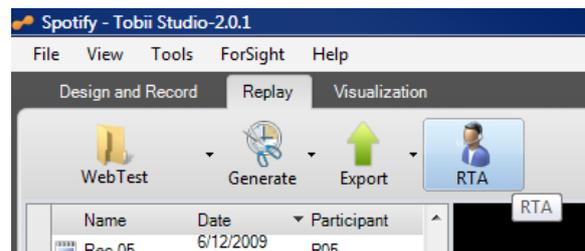
Although participants have the option to pause, rewind/fast forward or change the speed at which the video is replayed, some feel that they have to keep the commentary of their actions and thought process in sync with the video replay. In addition, a replay of a video takes roughly as long as it took to record. Consequently, if you need more flexibility in controlling how long a participant spends on each task in the study, it might be more suitable to use a static gaze plot as the interviewer can urge the participant on to the next page or choose to spend more time on a specific page.

4.2.3 Other Cues

There are of course other cues not discussed in this document. Two of these are not to use any cue at all or using a video cue without eye tracking data, but both of these have shown to provide fewer comments and less information about usability problems than when using a gaze plot or an eye gaze video. Therefore, to gain the most information from your participants, it is recommended that a gaze plot or eye gaze video is selected as a cue when using RTA in combination with eye tracking.

4.3 How Can a RTA Session be Recorded?

If you are using Tobii Studio 2 Enterprise version, there is an easy-to-use RTA function included that allows you to quickly set up the RTA recording of your choice (gaze plot cued, eye movement cued, video cued etc.). By clicking the RTA button you can name the RTA recording and then initiate a screen recording where you can present the participant with whatever cue you want to use from Tobii Studio or on the computer. The participant's reactions and comments about the task they just completed are then video and audio recorded in Tobii Studio. Participants can now both discuss the task they completed and answer any pre-set questions you have created, while it is all recorded by Tobii Studio.



5 What Should I Do Once the Testing Is Completed?

5.1 How Can the Data Collected be Used?

One advantage of using an eye tracker is that you can extract objective data from the recordings by using metrics such as time to first fixation, observation count, etc. This enables you to combine objective data with the subjective data gathered from the RTA interview and other participant comments. The eye tracker allows you to generate heat maps that can be used to demonstrate problem areas, specific areas of interest or gaze plots that illustrate typical behaviors displayed by the participants. By analyzing eye tracking data in the context of participants' comments about the user interface and interaction with the site or software evaluated, valuable insight into usability problems and other issues can be derived.



5.2 How Can the Findings be Presented?

You might want to present the data to someone who might be making changes to the website or software, or who is responsible for decisions about the website's or software's appearance. How you choose to present the data depends on your audience and the kind of information that they want or need. Using a combination of heat maps, gaze plots, videos from the recording or the RTA interview, user videos and comments, and statistics is a good way to provide an overview of your usability testing findings. Make sure to include a clear explanation of what heat maps and gaze plots are as well as what information can be extracted from them as a part of your presentation.

6 Recommended reading

- Bojko, A. (2006). Using Eye Tracking to Compare Web Page Designs: A Case Study. *Journal of Usability Studies*, 1(3), 112-120.
- Cowen, L., Ball, L. J. & Delin, J. (2002). An Eye Movement Analysis of Webpage Usability. In: *People and Computers XVI - Memorable yet Invisible: Proceedings of the HCI 2002*. Springer-Verlag Ltd., London.
- Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*, 35(3), 379-383.
- Hyrskykari, A., Ovaska, S., Majaranta, P., Rähkä, K-J. & Lehtinen, M. (2008). Gaze path stimulation in retrospective think aloud. *Journal of Eye Movement Research*, 2(4), 1-18.
- Jacob, R. J. K. & Karn, K. S. (2003). Commentary on section 4: Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyona, R. Radach, H. Deubel (Eds.) *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movement*, pp. 457-470. Elsevier Science: Oxford.
- Maughan, L., Dodd, J. & Walters, R. (2003). Video replay of eye tracking as a cue in retrospective protocol...don't make me think aloud! *Scandinavian Workshop of Applied Eye-Tracking*, 2003.
- Peyrichoux, I. & Robillard-Bastien, A. (2006). Maximize Usability Testing Benefits with Eye Tracking. SIGCHI Conference Paper.
- Tullis, T. & Albert, B. (2008). *Measuring the User Experience*. Morgan Kaufmann Publishers: Amsterdam.