Virtual Usability Testing

Pascal Haazebroek
Cognitive Psychology
Leiden University
Wassenaarseweg 52
2333 AK, Leiden
+3171 527 3654
pascal@haazebroek.nl

ABSTRACT
In this short paper we present a novel remote, asynchronous method of usability testing which combines the strengths of usability testing with the benefits of modern user behavior logging techniques.

Keywords
Usability evaluation, Remote usability testing, Tools, Asynchronous, Web usage logging

INTRODUCTION
In order to assess website usability, evaluators employ both analytical (e.g., based on usability heuristics) and empirical methods (e.g., usability tests with actual test users). The latter type of methods typically requires elaborate effort to invite, instruct, guide and monitor test users. In order to overcome some of these burdens, usability evaluators can resort to remote usability testing [1]. This method involves calling test users by phone, instruct and guide them remotely and logging their screen using video capture software. The obvious advantages include the flexibility of allowing test users to participate that are not able to come to the usability lab, not having to have a laboratory and even the comfortable and realistic setting of the personal computer environment at home or at work. However, these advantages are quite limited as remote usability testing still requires the evaluator to sit in real time, one on one (albeit remotely) with a single test user and monitor the test. In addition, the need to install and/or run screen capture software does not always appeal to test users.

Now, recent advances in web technology (e.g., [2]) allow the logging and analysis of user behavior in asynchronous fashion. That is, during usage of a website, user behavior is measured and stored in a log file. After collecting the data, the log file is analyzed. A well-known example is Google Analytics. A small adjustment to the website code allows its owner to analyze a variety of usage metrics, such as which page is viewed most often, what sites visitors are coming from and which pages stop the average visitor from actually ordering a product. These measures are primarily focused on between-page behavior and typically do not allow the analysis of user behavior from a usability point of view. The few web usage logging tools (e.g., [3]) that do focus on within-page behavior and usability are mostly faced with the problem of not knowing the context of use and/or the users' goals at the time of the logged behavior. This renders the logged data often of limited usefulness.

In this short paper we present a novel remote, asynchronous method of usability testing which combines the strengths of usability testing with the benefits of modern user behavior logging techniques.

VIRTUAL USABILITY TESTING
The proposed method consists of four components: (1) logging web events, (2) defining tasks, (3) conducting the test, (4) analyzing the data. Each of these components is discussed below.

Logging web events
Modern web browsers are equipped with JavaScript engines that allow the running of scripts on the background. We employ this ability – much like the technique presented in [2] – by adding additional event handlers to the Document Object Model that allow to send simple messages to the log server while the user is interacting with website elements, such as hyperlinks, form fields and buttons. By attaching these handlers to browser events of mouse movement, mouse clicks, key presses et cetera, it is possible to send and receive detailed information of user behavior *within* the webpage, rather than between webpages. This technique only requires the addition of a single line of JavaScript code to each webpage that is to be tracked. Test users do not need to install any type of additional software in order to participate in the 'Virtual Usability Lab'.

Defining tasks

Before user behavior is monitored, first those tasks the evaluator is interested in, need to be identified, much like how this is done in composing a test script before conducting a classical usability test. Such tasks may involve the most important flows of interaction on a website (e.g., selection in product lists, product details, product ordering, payment), but may also include occasional tasks that are still quite important (e.g., customer service). In addition to these tasks, a test also includes questionnaire items in order to acquire the test users' opinion on certain aspects (e.g., opinions on the site esthetics or the perceived ease of use).

Conducting the test

When the test is defined, the study is made active and test users can sign up for participating in the usability test in our virtual lab. One could restrict the testing to only those users that actually use the website normally (e.g., by putting a 'help us improve our website' ad on the website), or match the website with a selection from a test user database (e.g., testing university websites with students or aspired students). In principle, the virtual usability test follows the course of action of a classical usability test: after choosing to participate in the test, a test user receives an introduction, some basic questions and a series of tasks to accomplish on the website. During the completion of these tasks, the detailed information of the test user's behavior is logged rather than observed in person. The test ends with a questionnaire. Depending on the website's traffic one could gather enough data within a day or within a week time. In a sense, the evaluator sets up the test as a behavioral scientist sets up his/her experiment: test participants receive an instruction and give responses to presented stimuli. During the test, the participant's behavior is measured precisely and analyzed afterwards.

Data analysis

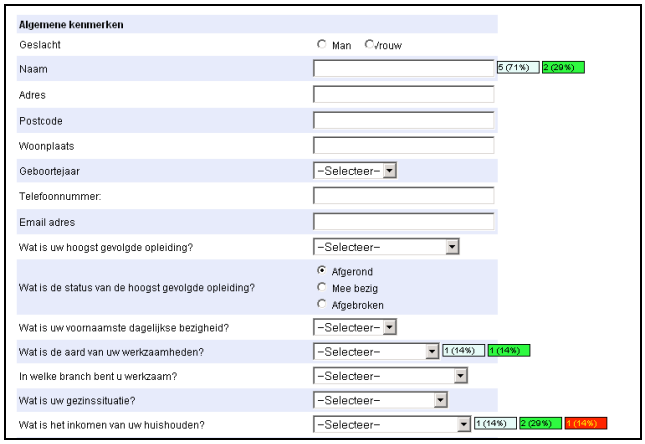
When enough data is acquired, the test is made inactive and data analysis can begin. As mouse behavior (i.e., both movement and clicks) is logged, the evaluator can determine for each task how much time was required to complete the task. As the goal of a task (e.g., ordering a specific product) is known, the evaluator can also determine task success rates. Finally, because log data is available, the evaluator can examine what happened in trials that have high reaction times, allowing the identification of possible conflict conditions or competing interaction paths (e.g., alternative hyperlinks, buttons or menu items that were chosen over the 'correct' option, or even lengthy mouse trajectories that could point to moments of 'doubt' in the interaction).

PRELIMINARY RESULTS

We have conducted some preliminary tests with a small number of websites and a handful of test users to validate the technical aspects of our approach. Figure 1 shows the click surface of a webpage, indicating which clicks were successful and which were incorrect. Figure 2 shows a webpage with a form. Next to the form fields percentage labels are added, indicating how often a field was the first one to be edited, how often it was skipped and how often it was the last one to be edited. This gives a rough idea of how users in general fill out the web form.

Figure 1. Correct and incorrect clicks within the webpage.

Figure 2. Form fields with their 'usage percentages'.



Importantly, the implementation of the Virtual Usability Lab is work in progress and needs to be extended with facilities that allow the evaluator to specify tasks so that this illustrated within-page behavior can be interpreted in a meaningful way.

DISCUSSION

Where traditional usability testing focuses on qualitatively assessing the behavior of a small group of test users and where traditional website usage logging focuses on simple statistics of mostly out-of-context bulks of data, our Virtual Usability Testing method combines the strengths of these two worlds: we define a task context for observed behavior

and are able to gather moderately large amounts of data without the need for real time behavior monitoring by the evaluator. Of course, usability testing has its own merits (e.g., being able to observe the test user, ask additional questions on the fly et cetera) as is the case for web usage logging (i.e., statistics on the general behavior of all users of the website instead of test users). Indeed, Virtual Usability Testing is not meant to replace these methods. However, we do believe we introduce new possibilities to the toolkit of a usability evaluator: by defining measures on within-page and between-page behavior, the evaluator can test the design against expectations or even usability benchmarks. In addition, new usability engineering processes are made possible. For instance, one could envision an ongoing improvement of a website, where after each small change to the design the website's usability is re-tested in the virtual lab. Changes in design should reflect changes in the measured user behavior, giving designers quantitative feedback on the design improvement in terms of actual user behavior measurements within meaningful task contexts.

REFERENCES

1. Andreasen, M.S., Nielsen, H.V. Schrøder, S.O. and Stage, J. What happened to remote usability testing? An empirical study of three methods, in *Proceedings of CHI '07* (San Jose CA, May 2007), ACM Press, 1405-1414.
2. Atterer, R., and Schmidt, A. Tracking the Interaction Users with AJAX Applications for Usability Testing, in *Proceedings of CHI '07* (San Jose CA, May 2007), ACM Press, 1347-1350.
3. Vargas, A., Weffers, H. and Vieira da Rocha, H. A method for remote and semi-automatic usability evaluation of web-based applications through users behavior analysis, in *Proceedings of Measuring Behavior 2010* (Eindhoven, The Netherlands, August 2010), ACM Press, 253- 256.