An Open Source Platform for Collaborative Remote Usability Studies

Røder, Daniel L.¹ and Frokjær, Erik.²

¹²Department of Computer Science (DIKU), University of Copenhagen, Denmark.

Abstract—In the field of usability testing, remote evaluation methods have been suggested as a way to combat the high costs incurred by traditional laboratory testing. The Distributed Usability Evaluation (DUE) framework is an asynchronous remote testing suite which has yielded good results in an industrial case study of production software. The framework allows users to send video/audio reports to a centralized server on which evaluators and developers can collaborate to find their optimal solutions. In this paper we present the latest development iteration of the framework, which brings about significant improvements to allow for studies with an extensive amount (>50) of participants. The updated framework is submitted to an expert evaluation by acknowledged usability experts and scientists which concludes that the framework holds a potential for doing extensive usability studies with minimal effort. Based on the promising results, the framework is now released as an open source project in an effort to assist others in conducting long-term studies involving many users.

Index Terms—Usability evaluation; remote usability testing; instrumentation; video/audio reporting; open source; collaboration.

I. INTRODUCTION

The software industry is becoming increasingly aware of the benefits of doing usability studies both in terms of how people perceive their products, and how efficient the products function. This has however been a long and slow journey as the methods most commonly used in the industry have a large demand for both time and person-hours to execute, leaving it in many cases a luxury of larger companies. As a response to these challenges, various “discount” usability methods have been introduced, spreading also to the field of remote usability testing.

In remote usability testing the test administrator and participant can be separated in space and time. According to Dumas [1] there are several advantages both in the logistics associated with conducting the test as well as making it easier to recruit users for studies. Remote methods has been explored since the mid 90’ies and different tools has since been made available to assist in the conduction of such studies [2], [3]. In some of the earlier studies, the technology was reported as a hindrance for the success of the methods, but with the rapid development in both hardware and software the boundaries of what is possible is continually moving [1].

Bruun et al. [4] recently published an article in which they do an extensive literature review in the field of remote usability testing. They recognize the distinction in remote usability studies between synchronous methods in which user and evaluator are only spatially separated, and asynchronous methods in which the separation is both spatial and temporal. While synchronous methods do deliver greater flexibility, it is just as time consuming as ordinary lab testing as the evaluator still needs to be present throughout the evaluation. The asynchronous methods are further subdivided into different categories depending on the technique used for gathering data from the users. From the overview provided by Bruun et al. it is apparent that textual data forms are the predominant technique employed by asynchronous studies, as they span methods such as automatic log data retrieval, online questionnaires and various collaboration tools.

We present the Distributed Usability Evaluation (DUE) framework for conducting asynchronous remote usability studies, which builds on the ideas behind the user-reported critical incident technique (UCI) [5], but expands this technique. The framework consists of a process to be followed during the evaluation as well as an open source toolset based on the process, which encompasses the full cycle of a usability study. The framework separates itself from commonly used remote evaluation techniques by using audio and video recorded from the users own workstation as basis for the evaluation. These reports are gathered on a centralized platform which provides functionality to strengthen communication and collaboration among different stakeholders. Following the UCI idea, the initial evaluation is done by the users themselves, thereby limiting the work load for the evaluator.

II. PROCESS

The DUE framework is designed with the primary intention of enabling usability studies to be conducted without the extensive physical requirements incurred by traditional laboratory tests, while also minimizing the needed person-hours.

The process suggested by the framework calls for 4 different roles to participate in the evaluation:

User: The user is the source of data. They report usability issues on their own workstations and send them to the evaluation server.

Evaluator: Evaluates the data received by the users, categorizes them and assigns a rating based on severity.

Development Manager: Prioritizes the issues produced by the evaluator.

Developer: Changes the evaluated program in accordance with the issues, and closes the issue.
This division helps to clearly define responsibilities within the process, and also follows the practical approach adopted by many small to mid size companies. There is no implied limit or requirement as to how many people are participating in each role, but the workload will closely follow the number engaged in the user role. The associated toolset provides a separate interface for each of the four different roles, to further support this work process.

As stated in the definition of the roles, the users are alone responsible for reporting the usability issues. These reports are done via video and audio recordings from the users own computer. The user also provides a severity rating before the collected report is uploaded to an evaluation server. The user is thus working in his natural environment while the evaluation is running, which gives some unique advantages: Tests can run extended periods of time with minimal cost. Prolonged tests can help mitigate the learning curve effect of users working with new software. The user is in a stress free environment and will experience no pressure to perform at a certain level, as is the case with situated tests. The framework is also flexible enough to allow for the other roles to report issues, if they gain valuable insights through the process that the user themselves does not recognize.

III. TOOLSET

The process is supported by an open source toolset designed to assist both usability experts and novices in carrying out successful evaluations. The toolset consists of a client program to run on the user’s workstation as well as a server to collect the reports generated, and support the evaluation process.

The clients primary function is to provide the video/audio reports generated by the user. This is accomplished by continuously recording all activity on the user’s workstation as a screencast stored on the local machine. The client itself has only a minimal UI placed along with the windows of the application under evaluation. It stays on top of other windows, but reverts to a semi-transparent view in normal work situations, to avoid disturbing the user during normal work functions. The simple UI also serves to keep user training at a minimum, by providing a simple and intuitive interface that is easily understood and adopted by the users.

When the user indicates that an issue has been encountered, he will be prompted to describe the issue to the microphone. Video material from the last 30 seconds prior to the user marking an issue will be prepended to the explanation, thus not requiring the user to recreate the circumstances which brought forth the issue. These time limits are based on the results obtained from a case study, which revealed these parts of the video to be enough for the evaluators to correctly classify the issues [6]. The video is then uploaded to the server for evaluation.

User reports uploaded to the server will be automatically imported to the evaluation system, and await further classification by the Evaluator role. Each role has its own interface on the server which primarily means that the server automatically finds and promotes the issues ready for that particular role, but every role has access to the full database. In studies with more than one evaluator, this helps to mitigate the evaluator effect [7] as all data and all decisions made are kept on the server and accessible for everyone. To further support this notion of collaboration it is possible for every role to mark an issue as needing review by another role while adding a comment as well as sending an email with a direct link to an issue to other stakeholders. Practically this means that if e.g. a development manager does not agree with the classification assigned by an evaluator, she can voice her concerns and send the issue back for reevaluation to allow the evaluator to further explain his reasoning.

Currently the client program requires a Microsoft Windows environment to function also relying on Windows Media Encoder for the encoding of the screencast. The server runs on a normal AMP stack (Apache, MySQL, PHP) but requires Microsoft Silverlight for video playback in visiting browser.

IV. RESULTS

The framework has been successfully tested in a case study with production software under active development, in which 16 people were assigned to the user role. The study showed that the framework did produce usable issue reports while keeping the requirements significantly lower than traditional methods [6]. Since this study, the framework has undergone an extensive redesign phase, building further on the aspects of collaboration, rationale capture and a general streamlining of the toolset to make it even easier to deploy and customize. These efforts have most notably resulted in a more structured view of the user reports, a tagging system and generally better access to issue details.

The updated framework was submitted to an expert evaluation among some of the worlds leading usability experts. All invited experts are established names within the scientific community within the area of HCI / Usability and most with ties to the industry as well. 31 invitations was sent, 15 responded positively and 7 evaluations were received. This low number is largely contributed to time limitations as evaluators were only given a week to complete the evaluation. The evaluation was designed to make the experts evaluate the ideas and concepts of the process rather than the actual implementation of the toolset. To accomplish this, they had to base their reviews on a prerecorded video presentation of the framework1 rather than working directly with the toolset.

Respondents were very optimistic about the framework’s ease of use as well as its ability to be a persistent repository for all of the aggregated data through a study. A respondent writes:

[R6] “The strength is that it seems to be easy for the user to explain a problem”

while another:

[R4] “Traceability of the data captured is the obvious strength of this framework”

1Available at: http://youtu.be/Cb7zWNrm-xM
Figure 1. The four different roles of the process shown with their respective responsibilities, as well as the optional data elements available to each.

and along the same lines:

[R2] "The system also nicely stores the evaluation history which may be of remarkable value"

The biggest concern being that the initial evaluation and classification was left to the end users, rather than an experienced evaluator. One respondent notes:

[R2] "The greatest weakness, in my mind, is that the evaluation is based on analysis made by users: users decide what usability problems / strengths are"

This is an entirely valid concern, but is also a generic argument for all studies in which users are the subjects in the evaluation. Even for usability studies done in a laboratory environment, the results will be influenced by how well the user responds to the method employed. The DUE framework on the other hand has the advantage of being able to sample large groups of users, thereby increasing the chance that their combined reporting efforts are covering their concerns.

Another respondent suggests that the toolset should try to further capitalize on the users being involved in the process:

[R7] "...possibly it would be good to allow the user also to send off some comment together with the recording"

which would further allow the users to act as responsible stakeholders. The toolset does indeed support letting the user send a textual comment along with the video, but it is not the default option. A feature of letting the users annotate the uploaded video as well as provide more detailed descriptions are planned for future development.

As a last question the respondents were asked if they themselves would be interested in using the framework in their professional work as scientists or business consultants, to which 5 out of 7 replied positively. Two of the respondents note that:

[R1] "Yes, I do lots of business usability consulting, I would be interested in getting access to DUE, and be happy to send you back the feedback from our work"
Figure 2. The issue details view of the server. At the top are all relevant data entered by any of the roles presented. In the middle the video uploaded by the user is directly available, and at the bottom all uploaded reports associated with this issue is listed.

**Issue details:**

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issue ID</td>
<td>6 (send link to issue in e-mail)</td>
</tr>
<tr>
<td>Status</td>
<td>prioritized, waiting for response</td>
</tr>
<tr>
<td>Report time</td>
<td>14:09:30</td>
</tr>
<tr>
<td>User rating</td>
<td>red</td>
</tr>
<tr>
<td>Occurrences</td>
<td>3</td>
</tr>
<tr>
<td>User description</td>
<td>new Usability Video interface</td>
</tr>
<tr>
<td>Tags</td>
<td></td>
</tr>
</tbody>
</table>

**Developer:** role3  
Priority: 1, MUST fix this.  
Comment: Go with the evaluator solution.

**User video:**

![User video thumbnail]

**Supporting reports:**

<table>
<thead>
<tr>
<th>User</th>
<th>Date</th>
<th>Time</th>
<th>User rating</th>
<th>Report ID</th>
<th>User description</th>
<th>Evaluator description</th>
</tr>
</thead>
<tbody>
<tr>
<td>role1</td>
<td>18 Aug 2012</td>
<td>00:25</td>
<td>red</td>
<td>10</td>
<td>view</td>
<td>yes</td>
</tr>
<tr>
<td>role1</td>
<td>18 Aug 2012</td>
<td>00:33</td>
<td>red</td>
<td>11</td>
<td>view</td>
<td>no</td>
</tr>
<tr>
<td>role1</td>
<td>18 Aug 2012</td>
<td>00:38</td>
<td>yellow</td>
<td>12</td>
<td>view</td>
<td>no</td>
</tr>
</tbody>
</table>

**[R5]** “Yes, I think it might be interesting to experiment with, especially in relation to our work on user-reported usability problems”

While the two voicing negative opinions:

**[R2]** “I find as too big a risk to trust that users can do valid evaluations (find and report valid usability problems & strengths) by themselves. I would do the analysis myself or let some other usability professional to do it”

**[R3]** “I am rarely involved in developing the type of applications for which this was designed - I mostly work with ‘tangible’ products/prototypes ... also, I find it very important to observe users during their interaction with a system”

This shows that the majority of the responding usability experts consider the toolset to be a potentially valuable supplement to their method palette when doing usability studies. As a consequence of this, a decision has been made to release the framework as an open source project to further develop the...
method and to gather more results from its practical use. The expectation is that the DUE framework can find its place as a interesting supplement to methods already in use, by providing evaluations of a full software suite, with a large group of users. For instance a DUE evaluation can provide an overview of usability issues in a software suite, identifying areas suited for e.g. think aloud testing.

V. CONCLUSIONS

We have presented the DUE framework as a method for conducting asynchronous remote usability studies with automatic audio/video capture of usability issues aggregated from a large number of users working in their natural environment. These data are collected on a common repository with facilities encouraging collaboration among stakeholders. The framework has previously shown great potential in a case study of industrial software. With inspiration from this study the framework’s associated toolset has been further streamlined and expanded. The revised framework has been submitted for an expert review among seven usability experts. Their response further indicates that DUE can be a useful addition to the usability professionals’ toolbox. In an effort to further enable large scale usability studies of software systems, the framework is now being released as an open source project.

ACKNOWLEDGMENT

Lars Christensen for formulating the theoretic background and developing the first version of the framework[6].

REFERENCES