UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA
C O M P U T A C Ã O

UFMG
UNIVERSIDADE FEDERAL DE MINAS GERAIS

The Value of Informal Evaluation in Collaborative Systems – A Case Study

RT.DCC.009/2006

RAQUEL OLIVEIRA PRATES
ALBERTO BARBOSA RAPOSO

SETEMBRO
2006

Abstract. This paper describes a case study in which informal evaluations have rendered valuable information about a videoconferencing tool, leading to a review of the interface, as well as changes to the evaluation plan. The study points to how informal evaluation can be cost effective and provide designers with useful information about their design, as well as the evaluation process.

1 Introduction

In 1988, Grudin pointed out the importance of evaluating CSCW systems, and the challenges involved in performing the task [4]. Almost two decades after the issue being raised, it is still an open issue to the CSCW community. Many studies have invested in understanding how evaluation has been done [8], what are the challenges involved [6] and proposing new frameworks [1, 5] or methods [2, 9]. In spite of all the effort, there is still no agreement when it comes to deciding what to evaluate, which method to use, or when. There is still the need to continue building up the knowledge about evaluation of CSCW systems, in order to provide designers of such systems with useful information that can support them in making decisions regarding evaluation.

One of the few aspects there is agreement regarding evaluation of CSCW systems is that it is difficult and expensive [1, 4, 5]. This has motivated a number of studies on how to evaluate CSCW systems at a lower cost [2, 9, 13]. Most of these works have proposed low budget methods that could be applied to evaluating CSCW systems. In this paper we present a study that illustrates the value informal evaluations can have in the design process of a CSCW system. By doing so we are contributing to building up knowledge about evaluation of CSCW systems, and specifically about the benefits and limits of low cost informal evaluations.

Our study planned at evaluating a videoconference system called CSVTool [10] developed to be used as a standalone tool or integrated to other CSCW systems and support direct communication among group members. The designers of CSVTool wanted to evaluate the quality of the interface being offered to users. Initially, the evaluation consisted of 2 different steps: (1) an informal inspection evaluation made by HCI specialists; (2) a formal controlled user test. Step (1) was concluded as planned and was used to define the focus of the user test and plan the user test. The test was planned for groups of 3 people and then 2 pilot tests were done to fine tune the material for the test. The problems identified in these two informal evaluations led to the decision to deal with the problems found, and delay the formal test with users. The issues that led to this decision were related mainly to the identification of scale-up problems – problems that prevent some aspect of the system from being evaluated [13] – and the high cost of the controlled user test.

The results of our study point at how informal evaluations can be cost effective and provide designers with useful information, especially if used early in design process as a formative evaluation. It mainly adds value when it identifies severe prob-

lems that could generate major problems in use, or scale-up problems that could cause a formal evaluation to fail in assessing intended relevant aspects of the system. An informal evaluation should not be the only one applied to the design of a system, but our study shows how it can allow for more evaluation steps to be taken during the design process, and thus, improving the quality of the system before a formal evaluation. It can also support the design team in deciding what, how or when to evaluate the system. Although our study illustrates the value the informal evaluation had in CSVTool project, the results are still preliminary and part of an ongoing research that aim at better understanding how to make better use of informal evaluations during the design of a CSCW system.

2 CSVTool

CSVTool (Collaboration Supported by Video Tool) is a collaboration tool developed specifically to be used in a large oil & gas company [10]. Since the company is present in many regions, including administrative offices and production fields, strategies for communication among these places become necessary. CSVTool has been conceived in cooperation to the company research center to comply with their necessities and technological requirements, such as platform-independence. Although the tool has additional collaboration resources, such as desktop content transmission, its main focus is communication, providing audio, video and textual channels to group members (Figure 1).

Among the characteristics of CSVTool, it has been designed to be used both as a standalone, generic videoconference tool, as well as integrated to other collaborative tools, already used by the professionals at the company. Most of these professionals used single-user systems as part of their daily work, but were not users of collaborative systems. Usually they resort to face to face, or telephones for communication, or at the most to simple communication systems, such as chats. The goal is to enable easy utilization, initialization and high adaptability and coupling to applications with collaboration resources in a distributed environment. To use the CSVTool, the first step is starting a collaborative session. Once the session is in place, users must connect to the other users they intend to talk to. Once the connection is established, the videostreams exchange among the participants is automatically started by CSVTool.

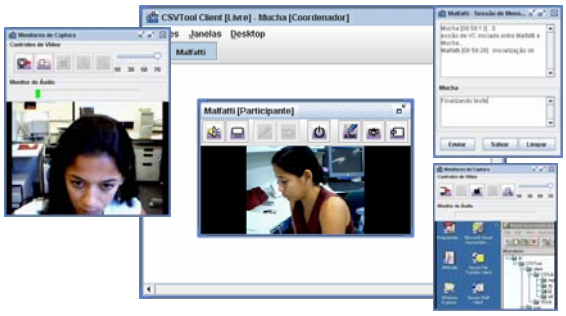

Figure 1 - CSVTool user interface

Currently, a pilot version of CSVTool is being used by technicians in the specific area of geosciences, mainly in its standalone version, but also integrated to a distributed host application aimed at the collaborative construction of earth-models for application in geosciences [12]. In this paper, the evaluation has focused on the quality of the interface for final users of the standalone version.

3 Performed Evaluation

The goal of the CSVTool design team was to evaluate the quality of the interface and whether users could easily interact with it. HCI specialists were contacted to conduct the evaluation. The HCI specialists decided that the best way to proceed would be to do an informal inspection of the tool, and based on the issues identified decide what to focus the evaluation on and which method to use. The decision was to focus on members' awareness and social issues, and to do so the method chosen was a user test with groups of 3 members. We next describe each evaluation performed and why it is considered informal, as well as their main findings.

3.1 Specialists Informal Inspection

The inspection was carried out by two HCI specialists, by going through CSVTool available functions and identifying potential problems the users could experience. In fact this is the basic procedure for any inspection evaluation [11], however, what defines this one as informal is the fact that specialists did not adopt any specific method for performing it. The inspection was solely based on their expertise. As CSVTool is a synchronous system, the evaluators had to work together in a session in order to inspect it.

Findings. The main issues raised by the inspectors are presented in this section.

Invitation. In order for participants to connect to each other, one of them must invite the other for a conference (metaphor: phone call). In case the other refused the connection at that time, the inviting participant did not have any feedback about it. Also, the invitation did not expire, nor could it be withdrawn. Thus, once invited the participant could contact the other one at any moment after that (as long as CSVTool session had not been terminated).

Communication Channels. Participant had fine control over video and audio communication, making it possible to interrupt the sending or receiving of a channel at any given time. However, they had no such control over the instant message channel, so it could be intrusive or interrupt a participant, without the participant sending the message being aware or having the intention of doing so.

Communication Channel Feedback. The user has control over the communication with each participant in the conference. What that means is that for each one she has

control over receiving and sending video and audio channel. Participant had to look at 4 different buttons, and which state they were in (4 different states were possible). There is no overall view that could help them quickly see their audio and video connections to everyone else, and no information at all regarding connection between other participants in the conference.

Social Issues. Evaluators identified that these problems could lead to issues of a social nature, by forcing breakdowns in social and cultural norms, such as politeness and privacy. For instance, when a participant could not immediately accept the invitation and denied it. If later on he became available, he could not inform the other ("I am now available") or then invite her. If he tried to invite her, he would be automatically connected to her. This could be intrusive and unexpected and could be upsetting. Also, in face to face situations usually people do not leave others talking to themselves, even if they are not interested in what they are saying, due to social norms. In CSVTool, if a participant were speaking to many others, each one could interrupt the receiving audio (stop listening) without knowing the others had done the same, and just leave the speaker talking to himself. Even just turning off the receiving of someone's audio could be perceived by the other as "I am not interested in what you have to say".

Conclusions. After discussing with the design team coordinator the findings of the inspection, he explained that the decision to give users such fine control over the communication channel, was because they believed users could have bandwidth problems, and thus, would at some moments have to favor one channel over the other. The control gave users the opportunity to decide in each situation, which channel would be the most relevant to keep. The evaluators had not understood through the interface the intent to offer the fine control as a way to manage net resources, but rather to manage communication goals. Thus, one aspect relevant to evaluate was how (most) intended users would interpret the fine control of the channels. Note that understanding the channels' control as means to manage physical resources could minimize some of the potentially negative social effects.

Based on the inspection findings and design intent, the decision was to perform user tests in groups of 3 members with the goal of evaluating: (1) Could users maintain a fluent conversation using CSVTool? (Without user-system or user-system-user communicative breakdowns?); (2) Would users understand the design intent? (Fine communication control meant to manage bandwidth needs?); (3) Would users be aware of their own communication connections with the other participants in a small conference? (4) How would users react to potentially negative social situations?

3.2 Controlled User Test

The user test was planned for small groups of 3 in which one of them would be interviewed on a recent social issue being discussed in the country at the time. A script was prepared, containing the role of each member, and some questions for the interview. The participant being interviewed would give his own opinions on the topic. The script involved inviting other participants to the conversation, the questions to be

asked, and some changes on the status of the communication channels. Some of the situations to be observed were: one participant that had been invited, came in late into the conversation (denied the invitation and later on came in and started participating); and at one point the participant being interviewed was left speaking to himself (the others disabled the receiving of his audio).

The test execution involved 3 users, 3 evaluators and one member from the CSVTool design team to simulate the "late" participant. Each evaluator would observe each one of the participants. Also, if necessary, they could help the participant follow the script (e.g. reminding him it was his turn). Users would be placed in separate rooms, and could only communicate through CSVTool. Before starting the videoconference, users were given an overview explanation about CSVTool, which included the presentation of the control of the communication channels and how its status was represented in the interface. Also after the test, a group interview was conducted with all the test participants and evaluators.

Test participants would be volunteers with the same profile of intended users, that is, expert users of technology, but not developers. A pilot test was done with members of TeCGraf who were software developers, but who were not involved with CSVTool project and had never seen the system.

During the first pilot test, a misunderstanding evaluators had had about CSVTool's invitation function became clear. They had thought a user was invited to participate in an ongoing conversation, and thus would be connected to all other participants of that conversation. However, the invitation is only for a conversation with the inviting participant. During the pilot test the problem was overcome, but the script had to be changed and a second pilot test was executed.

Findings. The pilot tests, besides allowing for the fine tuning of the test plan and material, pointed out some interesting issues regarding the test goals.

Communication Awareness. None of the interviewed participants noticed that the others had interrupted his audio, and that for a while he had been talking to himself[1]. The other participants also did not realize that the interviewed participant had been left speaking to himself[2]. Although, CSVTool's interface showed the status of the channels, participants had low awareness of their communication status with others.

Social Issues. Since users did not notice they were talking to themselves, the impact of potentially negative social issues could not be observed during the videoconference, but users were asked about it at the group interview. The interviewed participant of the first pilot test was visibly embarrassed to learn that he had been talking to himself. All the users, except one, said they would not have been pleased to find out they were talking to themselves. Only one participant said he would imagine the others were having bandwidth problems. The delay in coming into the conversation of one of the participants, was noticed, but did not bother them.

---

[1] One participant interrupted the audio receiving in the middle of the interview, and the other right after the last question. So the problem could not be attributed to the amount of time he was not being listened to.

[2] CSVTool did not depict this information, but evaluators wondered if participants could have perceived the situation through indirect clues obtained through audio or video of the others.

New Issues. The realization that users only knew who they were talking to, but were not aware who the others were talking to, was a new potential problem to be considered. For instance, in pilot test 1, there was a breakdown when the delayed participant (P4) connected to one of the interviewers (P1), and started asking a question. P1 believed P4 was connected to all the others and did not realize P4 was talking to him. That led P4 to repeat a few times her question, while P1 just waited for someone else to answer.

Evaluation Cost. Although evaluators were aware of the high cost of the evaluation, it proved itself even more expensive. During the first pilot test a technological issue came up – CSVTool and the screen capture system interfered with each other – and took some time to be solved for the second pilot test. The cost for adjusting the use of both systems added to preparation cost, but would not add to the time for each actual test. The cost for each test involved finding the volunteers with the desired profile, coordinating an available time for all 3 of them, the time for 3 evaluators (having one for each user was proved necessary during the pilot test) per test, the time for 1 member of CSVTool design team to play the role of the late participant. Each test took about one hour, including the use of CSVTool and group interview. Thus, the plan for a qualitative study including 5 groups would be highly expensive.

Conclusions. The pilot test findings led us to reconsider the initial evaluation plans. The comments during the group interview indicated that users may not necessarily interpret the disabling by other member of a channel as a bandwidth problem, and that negative social impacts could arise and were worth evaluating. However, due to the low communication channel awareness, there was a chance that during the test evaluators would not be able to observe these impacts and would only be able to collect information regarding the problem discussing it hypothetically during group interview, as had happened in the pilot tests.

Due to the high cost of the test with volunteers and the scale-up problems identified, the evaluators thought the best course of action for the moment would be to postpone the tests. The recommendation to the design team was to work on providing users with a better way to quickly perceive the status of their communication channel with everyone else they were talking to, and also to decide whether they wished to communicate more explicitly to the user their intent to express the channel interruption as a bandwidth problem. Without this, these scale-up problems could prevent the evaluation of the social issues as planned.

Pilot tests are used to fine tune planned test, and their results are not supposed to be considered to any other effect. However, instead of discarding the collected data, evaluators used it to identify scale-up problems, and change the evaluation plans. Thus, although the pilot tests were conducted rigorously as a part of a formal method, the formal methodology was not completed, and the pilot tests by themselves should be considered informal evaluations.

4 Discussion and Next Steps

In this paper we described the study in which informal evaluations were conducted in CSVTool and how they helped evaluators identify scale-up problems, and replan the evaluation process. Thus, their results supported evaluators in dealing with some of the challenges of evaluating CSCW systems: identifying problems that could have caused the evaluation to be ineffective [13], and mapping factors to be considered to the choice of an appropriate methodology [5]. Although the CSVTool study illustrates one success story, it can be perceived as a proof of concept of the value of informal evaluations, and it contributes to building up knowledge about CSCW evaluation, as well as pointing to some research directions.

Informal evaluations can be done at a lower cost than even low-budget methods. In CSVTool the informal inspection done by two HCI experts identified some potential problems, but was not sound enough to clear the misunderstanding evaluators had regarding the invitation function. Thus, for such a tool a formal inspection would have needed at least 3 people to examine it. The exact number of evaluators would depend on the method adopted. For instance Heuristic Evaluation recommends 3 to 5 evaluators to generate independent reports that later are consolidated[3] [7]. This would mean 9 evaluators, and a lower cost version could consider 1 HCI expert and 2 non-experts in each group (or experts in other areas such as CSCW, or the system's domain). Even considering methods that required just one evaluation report [3], it would mean at least 3 people, one of them being an expert, and would certainly take up more time than what was spent in the informal inspection. The cost of having executed the formal evaluation with users would have been much more expensive, as discussed in section 3.2.

As appealing as the potential low cost for informal evaluations can be, the fact that it is achieved at the expense of the application of a formal methodology has its toll. Potential problems may go unnoticed, as happened in the informal inspection. The issues associated to the one-to-one invitation feature, were only identified during the pilot test. Also, conclusions based on information gathered during an informal evaluation can not be rendered as representative. For instance, in the pilot test only 1 out of 6 participants interpreted the channel disabling as a bandwidth problem (as had been intended by designers). Based on that we may state, that some users may not interpret the channel controls as intended by designers, but not that most users will favor a different interpretation. This may not be enough to support the design team in making their cost-benefit decisions as to what their next actions should be.

It is clear that informal evaluations should not be used as substitutes for formal evaluations, but rather they should add to the evaluation process, as part of the formative evaluation, or supporting decisions on which formal methods to use and when. The next steps in this research aim at better understanding the usefulness and limits of informal evaluations. This effort will involve applying informal evaluations to other CSCW systems, and comparing their results to those of this study. As experiences

[3] This recommendation is aimed at single-user systems, and the study proposing its extension does not make any recommendations in that direction [2]. However, maintaining the same number of independent reports would probably be a valid assumption.

with informal evaluations are collected, an investigation on the possibilities of identifying situations in which they could be useful could be performed. To better understand its limits it will be necessary to conduct both informal and formal evaluations of CSCW systems at the same points in the design process and compare their results.

References

1. Araujo, R. M., Santoro, F. M., Borges, R. S. (2004) `A conceptual framework for designing and conducting groupware evaluations', Int. J. of Computer Applications in Technology, Vol. 19, Nos. 3/4, pp. 139-150.
2. Baker, K., Greenberg, S. and Gutwin, S. (2001) `Heuristic evaluation of groupware based on the mechanics of collaboration', Proceedings of EHCI'01. Lecture Notes in Computer Science Vol 2254, p123-139, Springer-Verlag.
3. da Silva, E.J.; de Souza, C.S.; Prates, R.O.; Nicolaci-da-Costa, A.M. What they want and what they get: a study of light-weight technologies for online communities. Proceedings of CLIHC 2003. Rio de Janeiro: Editora PUC-Rio, 2003. p.135-146.
4. Grudin, J. (1988). Why groupware applications fail: Problems in design and evaluation. Office: Technology and People, 4(3), 245-264.
5. Neale, D. C., Carroll, J. M; Rosson, M. B. (2004) Evaluating computer-supported cooperative work: models and frameworks. . Proceedings of CSCW 2004, ACM, 112 - 121
6. Newman, J. and Huang, J. (2003) WETICE 2003 - Evaluating collaborative enterprises workshop report', Proceedings of 12th IEEE WETICE'03, USA.
7. Nielsen, J. (1994) "Heuristic Evaluation", in Mack, R. & Nielsen, J. (eds.) Usability Inspection Methods. New York, NY: John Wiley & Sons, 1994, 25-62.
8. Pinelle, D. and Gutwin, C. (2000) `A review of groupware evaluations', Proceedings of 9th IEEE WETICE 2000 , USA.
9. Pinelle, D. and Gutwin, C. (2002) `Groupware walkthrough: adding context to groupware usability evaluation', Proceedings of CHI 2002, Minneapolis, USA.
10. Pozzer, C. T., Lima, L. S., Raposo, A. B., Vieira, C. J. G., (2005) A Multi-user Videoconference-based Collaboration Tool: Design and Implementation Issues. The 9th International Conference on CSCW in Design – CSCWD 2005, p. 547-552, U.K.
11. Preece, J.; Rogers, Y.; Sharp, E. (2002) Interaction Design: Beyond Human-computer Interaction. New York, NY: John Wiley & Sons. 2002.
12. Reis, L. P., Raposo, A. B., Paul, J. C., Bosquet, F., An architecture for collaborative geomodeling. CRIWG 2005**,** Brasil. Lecture Notes in Computer Science, Vol. 3706 – Groupware Design, Implementation, and Use, p.121-136. Springer-Verlag, 2005.
13. Twidale, M., Randall, D., Bentley, R. (1994)Situated Evaluation for Cooperative Systems. Proceedings of CSCW '94, 441-452.