Usability Evaluation and Study of a Video-Conferencing Service Provided via the Virtual Conference Centre

Borka Jerman Blažič
Jožef Stefan Institute, Faculty of Economics, Ljubljana, Slovenia
jerman-blazic@ijs.si

Tanja Arh
Jožef Stefan Institute, Ljubljana, Slovenia
tanja@e5.ijs.si

Abstract—**Usability evaluation is a core component of user-centred design (UCD) approach and aims primarily to evaluate effectiveness, efficiency and satisfaction when users interacting with a product/service to achieve their goals and needs, influencing their decision to its future adoption (i.e., user acceptance). The focus of this research work was to assess the usability of videoconferencing system with traditional usability method (usability testing) in order to get good assessment of its learnability and applicability. Based on the five tasks performed by 29 participants from five countries, the overall evaluation of the Virtual Conference Centre (VCC) was satisfactory and implicit a high level of reliability.**

Keywords - usability evaluation; task scenarios; videoconferencing system; collaborative environment.

I.   INTRODUCTION

During the last years, sites like Google, Flickr, Youtube, LinkedIn, Facebook, Myspace, Skype and many more have developed extremely successful mass services which have led to a service paradigm known as Web 2.0. Web 2.0 is more an evolution in service design than a revolution. It proposes the use of a set of basic principles such as: "Web as platform", putting the user in the centre (rather than the technology); it uses simple user interfaces, aggregate knowledge and wisdom of crowds by using folksonomies, uses social software, makes proper and extensive use of URIs (Uniform Resource Identifiers) and HTTP using REST (Representational State Transfer) and Web applications, uses peer-to-peer and Grid networks, etc.

On the other hand, the use of videoconferencing and collaboration tools in the Internet is slowly taking up due to the availability of more bandwidth and to the development of better integrated and more usable tools. Those tools do not inter-work in most cases and are used mainly today for realising relatively simple tasks like working meetings, substituting the popular conference calls, to connect remote speakers on co-located conferences or meetings, etc. The integration of videoconferencing and collaborative tools within a Web 2.0 service is leading to gather use of those tools.

The growth in videoconferencing and collaboration tools and systems is heralding also a shift in the nature of Human Computer Interaction [1]. Use of new media technologies such as videoconferencing systems is becoming a part of everyday communication and entertainment amongst individuals. Indeed, user satisfaction with technologies related to distance and collaborative applications is an integral part of usability [6], which is defined as the extent to which people can use the product *quickly* and *easily* to accomplish their tasks [2]. Usability testing is to make sure that users can find and work with the functions of the product to meet their needs. One of the most important outcomes of usability test is a list of problems, which entail changes and thus improvement of the product. Usability evaluation of videoconferencing and collaboration tools is traditionally conducted by means of task performance measures and subjective measures such as questionnaires, interviews, etc.

The paper is organized as follows: the background and production description section introduces the application studied; the evaluation methodology and participants section describes the approach used. The next section of the paper deals with the task scenarios and the last section discuss the results. The paper ends with short conclusion.

II.   II. BACKGROUND AND PRODUCTION DESCRIPTION

The Virtual Conference Centre (VCC) was designed following the Model-View-Controller Model together with agile software design methods such that the logic and the models of the various functionalities can be developed independently of the views. The software design frameworks *Ruby on Rails and RESTful* were chosen, as they are the most efficient rapid prototyping tools for website design.
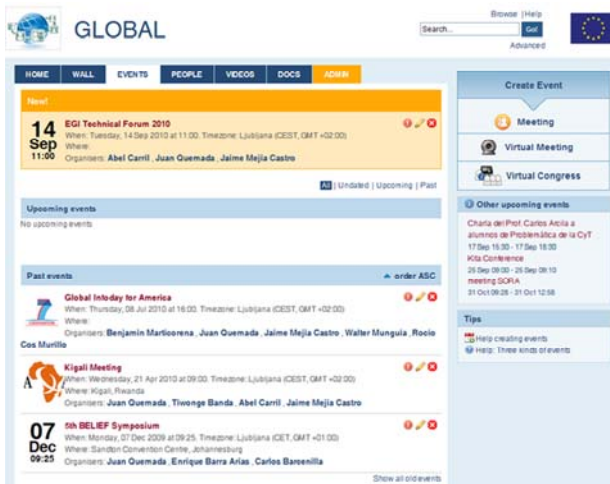
Figure 1: Event Space in the VCC

The VCC includes the following main features:

- A single unified point of access to the virtual auditorium features enabling access to the different functionalities through the available tabs: *Home, Events, Posts, People,* and *Spaces.* It also contains a specific button for direct access to the GLOBAL project, another for description of the project partners, a *Login* and a *Registration* button.

- *Spaces* are the means to organise different projects and different topics. Each space has a repository with public and private documentation and events (this functionality was not included in the usability testing). The space's public area can be customised to be the public face of the project in the VCC.

- In the spirit of openness, any user may register to the VCC. The registration procedure includes spam protection. The VCC registration allows the user to access publicly available spaces, documents, event announcements, etc. More importantly, it allows granting access to areas of the VCC that are marked *private* by a particular group of people.

- A profile display for every registered user such that he/she can update the information stored, such as password, address, email contact, project interests, etc. Profiles will facilitate user networking and partnership building.

- The *People* tab displays the registered users for quickly finding specific contacts and can only be seen by registered users.

- The *Spaces* tab shows the available VCC spaces. Each *Space* has a space administrator. The space administrator can add users to or delete them from a specific space, allow an "open to all"

registration policies, creates user groups under that space, and even delete the complete space.

- The *Events* tab is the event scheduler and the main part of the VCC. It allows the registered user to create events and organise event documentation in a clear accessible manner assigning access privileges to each item. Any VCC visitor, even unregistered users, may see the event calendar with all the public events, incentivising new registrations. Only registered users can access private events of the spaces he is a member of.

- The *Posts* tab gives access to the repository of shared documents and a blog-like or forum-like upload system based on posts to fill it. These posts may have multiple attachments. In future versions of the VCC other upload mechanisms will interface the user with the repository.

Virtual Conference Centre has been developed by the GLOBAL project from EU FP7 program.

III. III. EVALUATION METHODOLOGY

Standard user test procedures were adopted [5]. Jožef Stefan Institute (JSI) playing the role of *General UT Coordinator* was responsible for the overall coordination of UT, the compilation and analysis of the data, and the production of a usability report to be sent to the development team (UPM). The key role of *Local UT Coordinator* was to coordinate the conduction of the usability tests in his or her site and to ensure the required data would be collected and sent to General UT Coordinator. To ensure the uniformity of the testing procedures and thus the validity and reliability of the testing results, a document entitled "*GLOBAL Usability Test – Questionnaires & Task Scenarios*" describing in detail the goals, instruments, participant requirements, procedures, data handling, etc. Local UT Coordinators were supposed to follow the Guidelines with minimum deviations. The language of communication between General UT Coordinator, Local UT Coordinators and Local Testers is written English. Hence, it is expected that all of them possess certain level of English proficiency.

IV. PARTICIPANTS

The profile of the participants was characterized by several elements:

A. *General characterization of real user population*

Companies, researching and administrative staff of research institutes and higher education institutes, including director, professor, lecturer, tutor, and researcher, who have certain level of experiences and knowledge with regard to ICT can be users of the VCC. Characteristics which were common to all test participants were:

- Possessing experiences of using software applications;
- Possessing some basic knowledge of ICT and about videoconferencing systems, etc.;
- Possessing good English proficiency, at least a high level of reading comprehension.

B. Characteristics of the testing sites

English language version of the VCC has been tested in UbuntuNet in Malawi, Zentrum für Soziale Innovation in Austria, Jožef Stefan Institute in Slovenia, U. Politécnica de Madrid (UPM) in Spain, University College London in United Kingdom and CLARA in Peru.

C. Characteristics of the testing participants

The number of participants involved in the usability tests was 29 (8 female and 21 male). All of them had the educational level of at least the first university degree. Their participations were voluntary. Prior to working out the task scenarios with the VCC, the participants were required to complete a Pre-test Questionnaire on demographic data (gender, age, job title). This questionnaire also reflects the average level of competence in ICT (M = 3.80) and the average level of competences in videoconference systems (M=3.09). We convert the nominal to interval scale, with left anchor "1" indicating the lowest level and right anchor "5" the highest level of the attribute in question. None of the participants had interacted with the VCC before they took part in the usability tests. These demographic data are relevant for interpreting the results of usability tests.

D. Profile of the local tester

Ideally, Local Testers should be usability specialists or highly experienced in conducting usability tests. However, it may not be easy to get them, especially when the resource is tight. Alternatively, those who meet the following criteria can assume the role of Local Tester:

- Experienced in conducting experiments with human participants,
- Has some knowledge in Human Computer Interaction (HCI),
- Fluent in the native language and English,
- Motivated to do research,
- Good at observation,
- Able to manage several tasks at one time.

It was recommended to employ the same Local Tester to conduct all the testing sessions to ensure the consistency of data recording and interpretation.

V. TASKS SCEANRIOS AND PROCEDURE

A set of five tasks covering the core functionalities of the VCC was employed. The tasks were presented in English. All test participants possess a reasonable level of English reading ability. Below is the list of the tasks:

- *(T1) Obtaining a user account for a Virtual Conference Centre (VCC)*
- *(T2) Creation of new space in VCC and joining an existing one*
- *(T3) Creation of new event*
- *(T4) Modifying the event in event manager*
- *(T5) Sending private message*

Each of the above five tasks was translated into task scenarios, which render the test more realistic and problem oriented (e.g., *You are organising a workshop or other distributed event like a meeting. Therefore, you need to create an event in the space you created in task 2. Log yourself into the VCC and create a new event. The event must be marked as Isabel event)*. In addition, for each of the task scenarios, quantitative usability goals in terms of task completion time and number of errors were set, which were benchmarked by an experienced user of the VCC. They can serve as references or baselines for data analysis. System Usability Scale (SUS) (Brook, 1996) and Feedback Questionnaire (FQ) (After-Scenario Questionnaire) [4] were employed. Test participants were welcomed and briefed about the goal and procedure of the usability tests, which was followed by an explanation of the equipment to be used. Participants were asked to perform a set of selected task scenarios that cover most frequent as well as critical functionalities of the VCC.

After each task, participants were asked to complete the After-task questionnaire, consisting of four questions (Q1-Q4), which were derived from the literature on usability research [4]. A 7-point Likert scale was employed with left anchor indicating lowest level of satisfaction and right anchor the highest. Q.3 and Q.4 evaluated the same two variables, which are nonetheless phrased. After completing all the five tasks, participants were asked to complete Post-test questionnaire entitled "System Usability Scale (SUS)" which consists of 10 questions and has psychometric properties.

VI. ANALYSIS AND RESULTS

We have categorized collected data along two dimensions: (i) qualitative vs. quantitative and (ii) objective vs. subjective (Table 1). Some analysis results of these data types are presented in subsequent sections.

TABLE 1: TWO DIMENSIONS OF DATA TYPES

	Qualitative	Quantitative
Objective	A list of usability problems (UPs) derived from the participants' notes	Time-on-Task Effectiveness & Efficiency
Subjective	Participants' comments in SUS	Responses to the questionnaires: • Pre-test Q. • Feedback Q. • Follow-up Q. (SUS)

Descriptive statistics – mean and range – comprehends six quantitative objective and subjective usability measures: duration (min), perceived ease of use,

perceived efficiency, perceived difficulty, perceived time-consumingness and task completion rate (%).

A. Quantitative data

Different quantitative measures were taken, including time-on-task, proposed time, mean time and standard deviation. The two usability metrics – **effectiveness** and **efficiency** – were derived effectiveness and efficiency.

Furthermore, effectiveness and efficiencies per task were computed. Effectiveness denotes the rate that a task is completed successfully *without* assist from any help desk – unassisted completion rate. Efficiency is calculated through dividing an unassisted completion rate by its corresponding unassisted mean time-on-task. In the Table 2, for the sake of comparison, both unassisted and assisted completion rates together with their corresponding mean time-on-task are displayed. The average effectiveness over five tasks was 88.57 %. All participants could complete all the given tasks, with or without assist from the Local Tester. The average efficiency over five tasks is 21.52 %/minute, ranging from 7.14 %/min (Task 4) to 33.33 %/min (Task 1). In fact, Task 4 (Modifying the event in event manager) was proved to be problematic.

B. Time on tasks

Each participant was required to perform five tasks. Based on the data of the 29 participants, the value (average time) of this variable is 30.65 minutes, with the range from 15.00 (JSI-P5) to 81.00.

TABLE 2: EFFECTIVENESS & EFFICIENCIES PER TASK

Task	Total Completion (with or without assist)		Effectiveness (%) (tasks without assist)		Efficiency (%/min)	Total no. assists
	Rate (%)	Mean Time	Rate (%)	Mean Time		
1	100 %	4,72	100.00%	3,00	33.33	0
2	100 %	4,79	100.00%	4,00	25.00	0
3	100 %	6,20	85.71 %	5,00	17.14	1
4	100 %	9,48	57.14 %	8,00	7.14	3
5	100 %	5,44	100.00%	4,00	25.00	0

Altogether, 145 tasks were performed and 145 (100 %) were successfully completed. As shown in Table 2, among the 5 tasks, Task 4 (Modifying the event in event manager) was found to be most problematic. Indeed, the average time-on-task for Task 4 is 9.48 minutes, exceeding the benchmarked upper bound (i.e., 6.00 minutes) by 63.2 %. The range of time-on-task for Task 4 is large, spanning from 4.00 min (UPM-P2) to 20.00 min (UBN-P1 and UBN-P6). As a matter of fact, Task 3 (Creation of new event) is quite similar to Task 4 (Modifying the event in event manager) except the additional scheduling. As evidenced by the data, the average time-on-task for Task 3 is less than that for Task 4. All participants performed Task 3 much faster. However, in some other cases, the reverse could be observed (e.g., UCL-P2). This may be attributed to the fact that this user found the task very confusing and without logic.

One of the most important results from the Usability Tests is the list of Usability Problems (UP) identified when the participants interacted with the system to achieve the given tasks (Table 3). The six testing sites have collected different sets of UPs, with a number of them being overlapped and the others being unique. We compiled and integrated the six lists of the usability problems into a complete list (see Table 3). The implications of individual columns are:

- *Usability problem (UP)*: It is the identifier of individual UP.
- *Task ID*: It denotes in which task the UP was identified. For instance, Task ID 2(5) means that this UP was identified in Task 2 (Creation of new space in VCC and joining an existing one) and number 5 in brackets indicating how many of the test participants found the problem.
- *Descriptions of Usability Problem*: Detailed explanations what the UP was and how the UP was identified.
- *Severity:* There are three levels:
  - *Moderate usability problems* are those that significantly hinder task completion but for which the user can find a work-around.
  - *Severe usability problems* are those that prevent the user from completing a task or result in catastrophic loss of data or time.
  - *Minor usability problems* are those that are irritating to the user but do not significantly hinder task completion.

As shown in Table 3, there are altogether 21 the most important usability problems (UP). Some of the UPs have frequency only once and the highest frequency is 12 for UP13 (Task 3: Create a new event and invite people to this event) - users had difficulty in inviting people to the event, as well as the option to invite people during the event creation. In a scenario-based usability study, participants use a computer application to perform a series of realistic tasks. The FQ is a 3-item questionnaire to assess participant satisfaction after the completion of each scenario [3]. The items address three important aspects of user satisfaction with system usability: *ease of task completion, time to complete a task,* and *adequacy of support information (online help, messages, and documentation)*. Each item is rated with a 7-point Likert scale, with 1 being "Strongly disagree" and 7 "Strongly agree". The items are phrased in a positive manner. Hence, the *higher* the score, the more the user is satisfied with the system. The questionnaire takes very little time for participants to complete. Table 4 shows the results of FQ of the five tasks. Q1.1 addresses the ease of task completion for Task 1 as perceived by a user; Q1.2 addresses the degree to which the user is satisfied with the time to complete Task 1.

TABLE 3 IMPORTANT VCC USABILITY PROBLEMS

UP	Task ID	Descriptions of Usability Problem	Severity
UP1	1(9)	The participant hardly found the »Login« button on the start page.	Moderate
UP2	1(5)	The Login/Register button at the top of the page is quite inconsistent. Clicking the button you get the login page.	Moderate
UP3	1(3)	No alternative way for registering or troubleshooting while registering: contact info who to turn to if something goes wrong (e.g. not receiving confirmation mail).	Minor
UP4	1(2)	The facility for password reminder transmission via email did not work. The request was submitted successfully, a confirmation was displayed, but no email with the password was received.	Severe
UP5	2(2)	Topic titles/names are presented very user unfriendly (e.g. spaces/spacesname/post/200).	Minor
UP6	2(1)	User spaces should be allowed to be created only by registrated users (those who has confirmed registration).	Minor
UP7	2(9)	The purpose of groups within spaces should be more clearly explained.	Severe
UP8	2(2)	No connection to other users of VCC from your space (invitations).	Minor
UP9	4(7)	Confusion between "Comment" and "Post".	Severe
UP10	4(4)	Adding attachments to posts/comments is a little confusing – only one item can be added – no confirmation is given.	Minor
UP11	4(7)	When displaying "The post can't be empty" for e.g (or other system info) there's no exact explanation on the specific error made in the post.	Moderate
UP12	General (2)	Sometimes operations are performed without giving feedback/notice to user.	Moderate
UP13	3(12)	The "invite people" option is not intuitive to find, as well as the option to invite people during the event creation.	Moderate
UP14	3(9)	Private Message to complex to sent.	Moderate
UP15	General (6)	Help is not available.	Moderate.
UP16	General (3)	No information about what VCC is what it offers, what are its features.	Minor
UP17	General (3)	The participant hardly found the link to the support & help.	Minor
UP18	General (5)	Look of the VCC; everything happens in the most right part of the screen.	Minor
UP19	Invitation	The system crashed without any warning or error messages	Severe
UP20	General (3)	The Admin tab should be differentiated from the other tabs (e.g. Home, Posts, Events, etc).	Minor
UP21	General (5)	Clicked the back button of the browser and lost all the data, very frustrating.	Severe

TABLE 4. RESULST OF THE FQ TASKS

Qn	Q1.1	Q1.2	Q1.3	Q2.1	Q2.2	Q2.3	Q3.1	Q3.2	Q3.3	Q4.1	Q4.2	Q4.3	Q5.1	Q5.2	Q5.3
Mean	5,41	5,66	5,07	5,90	5,90	5,59	5,21	5,38	4,72	4,38	4,48	3,76	4,48	4,79	4,07
St. Dev	1,72	1,72	1,44	1,08	1,35	1,38	1,29	1,40	1,36	1,68	1,60	1,62	2,03	2,38	1,96
Min	2,00	2,00	2,00	3,00	2,00	1,00	2,00	2,00	2,00	1,00	2,00	1,00	1,00	1,00	1,00
Max	7,00	7,00	7,00	7,00	7,00	7,00	7,00	7,00	7,00	7,00	7,00	6,00	7,00	7,00	7,00

Q1.3 addresses the adequacy of support information for Task 1 as perceived by a user. The same sequence is for Task 2 to Task 5. For Task 2 (Creation of a new space in the VCC), the ease of completion was rated as 5.90, the degree of satisfaction with the completion time was 5.90 and the adequacy of support information was 5.59. These ratings imply that the users generally were quite satisfied with this particular Task. Task 1 (Obtaining a user account in the VCC) has similar ratings, but of lesser degree. Task 4 (Modify and event in event manager) and task 5 (Send private message) imply that the users generally were not so satisfied with this particular tasks.

Based on the five tasks performed by 29 participants with different cultural and academic backgrounds as well as various levels of experiences and knowledge in information technologies and video-conferencing systems, the overall evaluation of the design of the VCC (beta version) was satisfactory. The English language version of the VCC has been tested independently in six different sites. The average effectiveness and average efficiency of the five tasks over the 29 participants are 88.57 %.

VII. CONCLUSION

The results of the study show that the users' performance is highly acceptable to be improved to render it suitable for a wider scope of users, especially those who have limited experience and competence in ICT and in the domain of video-conferencing systems.

ACKNOWLEDGMENT

REFRENCES

[1]. T. Brinc, D. Gergle and S. D. Wood, Usability for the Web: Designing Web Sites that Work. San Francisco, USA: Morgan Kaufmann, 2002.

[2]. J. Brooke, SUS: A 'quick and dirty' usability scale. In W. Jordan, B. Thomas, B.A. Weerdmeester, I. L. McCleland (Ed.), Usability evaluation in industry (21, pp. 189-192). London, UK: Taylor & Francis, 1996.

[3]. J. R. Lewis, Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. SIGCHI Bulletin, 23(1), 1991, pp.78-81.

[4]. J. R Lewis, IBM Computer Usability Satisfaction Questionnaires:Psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction, 7(1), 1995, pp. 57-78.

[5]. J. Nielsen, International usability testing. In E. del Caldo & J. Nielsen (Eds.), International user interface. New York: John Wiley & Sons, 1996.

[6]. A. S Patrick, The Human Factors of Mbone Videoconferences: Recommendations for Improving Sessions and Software. Ottawa. CRC Technical report, 2006.