Evaluating Entertainment: Implications for usability tests conducted on entertainment web sites.

Kalle Jegers & Charlotte Wiberg

Department of Informatics, Umeå University Kalle.Jegers/Charlotte.Wiberg@informatik.umu.se

Abstract: Evaluating usability is sometimes troublesome. Evaluation of systems containing entertainment makes it even more complex. The question one might ask is if the traditional methods is useful at all when dealing with entertainment. We report from the first iteration of three, in an empirical usability evaluation series on entertainment web sites. The sites evaluated are Mosquito and Eurovision Song Contest. The approach we used was to use traditional usability evaluation methods and elaborate with specified conditions. This in order to be able to give guidance in how to use such methods for informing in design projects. The conditions elaborated were Intervention of testers, number of evaluated users at the same time, usage of more strictly approach with given tasks vs. a more explorative scheme and written vs. oral answers on questions concerning user's experiences. The paper conclude by giving implications for evaluation of entertainment web sites; (1) Elaboration with methods, questions and level of intervention must be refined in the start up as this type of sites are unpredictable in it's nature. (2) Verbalizations of experiences are very hard to get from the users. Overall, the general finding was that traditional usability is important in this context as well. Even if the web sites are entertainment based - do not forget about traditional usability. If a general level of understanding of basics as navigation and structure of information, the user experience when it comes to the entertainment might as well be lost.

1 Introduction

In this paper we offer implications for usability evaluation of web sites in general, but more specifically focused upon entertainment web sites. With entertainment, we here refer to sites designed in purpose to give the visitor an experience of amusement, rather than just displaying various kinds of information, as on many traditional web sites.

This relatively new kind of functionality these entertainment focused sites strives to

offer, seems to be in direct opposition to the functionality offered by web sites and systems focused on information retrieval. Since the established techniques and concepts used in usability engineering and usability evaluation in most cases are concerned with problems faced by information retrieval designs, they may not be sufficient for measuring the aspects of this new kind of functionality. A traditional usability test often consists time measurement as a measure of success. On the web this might not only be a proper measure, as surfing is different from searching (c.f Olsson, 2000). On entertainment web sites it would be directly misleading to use a clock to gain knowledge of whether the site is entertaining or not. The measure might be just the opposite, the longer stay the more entertaining environment. Or not! Sometimes the time do not correspond with anything at all concerning user satisfaction. One might even argue that traditional usability aspects are not relevant at all when evaluating and designing entertainment.

Since entertainment web sites are relatively new kind of phenomena, few attempts to apply existing methods for usability evaluation involving users have been made. Therefore, some questions considering methodological aspects arise: What will the consequences be, when existing evaluation methods are applied to entertainment web sites? Is it possible to measure and evaluate this new phenomenon with existing techniques? How will the basic properties of the available techniques affect a situation where entertainment is tested?

We wanted to focus on the most basic properties of evaluation with user tests. In order to do so, we narrowed our approach to deal with the basic properties of the Thinking-Aloud method, since it is a well-known and established method for retrieving qualitative, subjective data from the user.

Thinking-Aloud is often combined with different tasks for the user to solve. Besides the technical consequences when Thinking-Aloud is applied to entertainment web sites, we wanted to investigate the effects of the more formal task solving approach as well. In summary, the conditions we chose to focus upon in this first iteration are:

Level of intervention from the test crew/person. (Shneiderman, 1998)

Number of tested subjects at the same time and if co-operation could give a more natural situation for the 'think-aloud' method in this type of testing.(Shneiderman, 1998)

Written or oral answers on more discursive questions.

Free surf or task analysis; which is more suitable when user testing entertainment?(Vora, 1998)

2 Related work

Below, related work made of usability in general of interest, of web usability in general and finally of entertainment is shown.

The traditional definitions of usability (Nielsen, 1993; Shneiderman, 1998) tend to focus on factors considering user productivity and performance, in order to ensure timeand cost effectiveness in the situations where the system is applied. The majority of the most central measurable human aspects of usability evaluation are defined in terms of time for carrying out specific tasks, speed of performance and rate of errors made by the user (Shneiderman 1998).

Related work around web usability can be divided into two groups, usability tests of web sites and methodological papers. Below examples of both kinds are shown.

Schneiderman (1997) discuss usability aspects related to the web and says that, as in any media, criteria for quality vary with the genre and author's goals. His idea is that there are some web-related criteria that may be seen as more general, like visual appeal, comprehensibility, utility, efficacy and navigability. However, he continues to warn about these high-level goals and point out that a categorization of the web is needed to find more fulfilling criteria to test. Schneiderman gives an example on how to chategorize the web: (1) By originator's identity. (2) By the number of web pages in the site. (3) By goals of the originators, as interpreted by the designers (4) By measure of success. Schneiderman's work is genuine and valuable. However, for us in our evaluation of entertainment sites, not very applicable. Another work done in this research direction is a discussion around how web sites have other kind of characteristics than traditional interfaces. (Laskowski & Downey, 1997). Here, similar aspects like Schneiderman discuss are elaborated. Both of these papers are on a general level.

Gaines et al. (1996) discusses dimensions of problems on the web and tries to categorize sites from the concepts of utility and usability. They come up with a layered framework. Ratner (1998) tries to come up with some conclusions around novice and expert users in learning environments using Netscape. She stresses that even if the goal of the educators, have a specific goal and that the students seem to be a homogenous group, they are not. This must be taken into account in design of such web based learning environments. These problems are very critical and must be taken into consideration. We do not investigate them further here.

So far most usability studies of web sites have focused mainly on efficiency aspects (e.g. the time it takes for a user to find a piece of information in a relatively large site).

Information retrieval is the far most common target for usability testing at web sites. This is because this activity often is seen as central at the web in general (Spool et al, 1999). Usability tests may be conducted in numerous ways, including all from one single technique to a whole repertoire of approaches. It is important to be aware of what to measure. Two common approaches to measuring usability are the following (Redmond-Pyle & Moore, 1995): (1)*Performance tests*, where users use the system to perform a task, and their effectiveness are measured. Common measures are speed, accuracy and/or

errors.Examples are Borges et al (1996) and Borges et al (1998), where they first conduct heuristic evaluation on a number of university sites. Re-design of some of them are then conducted and finally task analysis is done where users are measured when doing tasks. The usability team then ended up with a list of guidelines as a result of their test. (2) *Attitude surveys*, where user satisfaction and user perception of the software is captured. Common ways of capturing data are questionnaires or interviews. A typical example of this is Spool et al (1999) and their huge usability test of big corporate sites with main focus on e-commerce. This report covers the study of nine sites, and here the tests are much wider. Instead of using the clock in measuring, the test team uses interview forms before and after combined with observations. The users got tasks, but interest was more shown on ways of finding information, instead of how quickly the information was retrieved. Grose et al. (1998) shows with a two folded study that web style guides differ from traditional style guides and stress the fact that this must be investigated further.

These examples show how usability engineers handle the web. However, the web is, as Schneiderman stated, very general. Therefore, we have to take further notice of such related to our scope, entertainment. Some work have been done by Claudio Pinhanez and his team and they discuss issues of entertainment web sites and provide a list of references. They look into literature regarding play, pleasure, engagement, and flow theory. (Pinhanez et al 2000; Karat, C-M. Et al. 2001). Pinhanez and his team have done a good work. Our work could be seen as another view on the same problem. We use others sites, they built their own. We try to find different types of entertainment sites. They use a more narrow scope of sites. We use other types of methods. Overall, our research could be seen as complementary rather than competing. We also agree with Pinhanez et al when they say that no relevant work on the subject have been done before.

3 Entertainment – what do we mean?

Before moving on, we would like to say something about entertainment in itself. "Entertainment' is not easy to give a single definition of. What is entertaining is different in different times, cultures, media and more. (Dyer, 1992). However, one way of, at least, separating it from other types of experiences is given by Pine & Gilmore (1999) where they place *entertainment, esthetics, education* and *escapist* as four realms of experience. They separate them as below:

Entertainment; This type of experience appears when people passively absorb the experience through their senses, i.e when listening to music.

Education; This is also a type of experience, where the user absorbs the experience. However, education is more of a participation type of activity.

Escape; This realm involve much greater immersion than entertainment or education experience, and they are often popular opposites of the above pair of experiences mentioned. Examples of escape type of environments are games, theme parks, chat rooms, a physical room of paintball.

Estheticism; Here, the users immerse themselves into the experience. However, they have little or no effect on the experience itself. Art galleries are good examples of such experiences.

Another way of

4 Test Design

Inquiries of participation were sent out by e-mail to approximately 80 persons of different gender, age and backgrounds. Of the 30 persons who responded, 13 were students at different departments of Umeå University and the rest were employees at the department of Informatics, Umeå University. As a reward for participating, the subjects received a ticket to the cinema after the tests were conducted.

The tests were performed at three different locations: in a small conference facility, in a computer lab and in a small office, all three located at the department of Informatics, Umeå University. To present the web sites, two laptop computers were used: one with external mouse, one with external mouse, keyboard and monitor. The tests were recorded on digital video camera, and MiniDisc recorder.

The number of tested users at the same time: how does the number of users in the test situation affect the outcome of the Think-Aloud method?

- Free surf with supportive tasks vs. strict task solving with predefined tasks: in what way are the users affected by the different approaches? How do the approaches affect the Think-Aloud method?
- Level of intervention from the test crew: how much help does a user need and in what situations?

The subjects were assigned on two different test groups, with two different sub groups and two conditions. In all: 10 users to the Mosquito test group and 20 users to the Eurovision Song Contest test group (See figure x below).

	Free surf	Task solving
Single user	3	3
Pair of users	-	4

	Free surf	Task solving
Single user	8	6
Pair of users	2	4

The Mosquito test group: Distribution of test subjects over categories.

Figure 1: Distribution of subjects over test groups.

The Eurovision Song Contest test group: Distribution of test subjects over categories.

Before each test, the subjects went through a short practice session in order to introduce the Think-aloud method. The Method was then used throughout the whole sessions, for all subjects, in all test groups.

In both test groups, the subjects answered written pre- and post test questionnaires, considering background factors as computer skill, age and gender, as well as experiences from the test situation. In addition to the written post-test questionnaires, the Eurovision Song Contest test group had a short verbal interview, considering experiences from the test. The Mosquito test group answered the same questions, but in a written form.

The test crew estimated the level of intervention on a 1-5 point Visual Analogue Scale, directly after the tests were completed. On the scale, 1 indicated no intervention at all, and 5 indicated a level of intervention where the user needed some kind of assistance in 90% of the time or more.

5 The two sites evaluated

This article offers the first report from an on-going research program, *Joyride*, which has the main goal to provide guidelines for evaluating entertainment web sites. In the project, a combination of participant observation at the design company, interviews with all types of actors involved in the design as project leaders, art directors, animators, illustrators, programmers and so on. Further, from basic usability engineering (c.f. Mayhew, 1999) and related work in usability testing of entertainment - and entertainment environments, we shaped different ideas of what aspects to test further.

The project contains of three iterations of user tests, in total for the whole project, more than 150 user tests. One purpose of the iterations is to refine the next iteration coming, and so far we have conducted the first iteration, on two web sites. The paper reports from totally 40 hours of evaluation.

The two sites were chosen as being two traditional entertainment web sites. One of them, Eurovision Song Contest (ESC), was an event site, which acted as support for the event with the same name in the year 2000. The second site, Mosquito was a support site for a TV-show on the Swedish National Television. Both of the sites were designed by Paregos AB. Below the sites are described.

5.2 The Mosquito web site

This site has received numerous prizes both nationally and internationally. This made it an interesting site to test. Quoting the description of the web site, from the corporate site of Paregos, the purpose of the production was as follows:

"Paregos accepted Swedish Television's challenge to create the web version of Mosquito as an extension of the TV-show and as a meeting place for those who like the program. The result is a flash site that has been awarded with several prizes in the media business, chosen the site of the summer by the magazine Resumé and won the Prix Italia prize for "the best innovative"

solution". But mostly, it has been a high-octane, crazy, wonderful meeting place for all the "Mosquitoes".

(http://www.paregos.com)

The web site is in active use, and can be reached from the official web site of SVT. (http://www.svt.se) Required plug-ins for the browser are: Macromedia Flash®, QuickTime® and Beatnick®.



Figure 2: The Mosquito home page.

Located in the upper right corner, is the equivalence to a main menu, here called "The Remote control". The remote control follows the user through all the sections of the site, and is the main tool for navigating. The remote control can be adjusted in size and moved around by the user. Next to the remote control is a small, animated mosquito, that links to a short QuickTime® movie. The bar code right to the animated mosquito is a link to Paregos AB corporate web site. In the upper left corner is a menu containing functions to change general settings for the site, a help function that display an additional layer of short information about the different objects on the page, and an exit button, since the site open up in "full screen" format when loaded.



Figure 3: The links page.

The sections in the left side of the tunnel links to different sites outside the Mosquito web site. Each section is a link to a different web site. When the user moves the mouse cursor over a section, a text with the address that the current section links to is shown.



Figure 4: The cool movies page.

At the bottom of the page is five small boxes, that links to small QuickTime® movies. When a box is clicked, the movie chosen is shown in a small, separate window.



Figure 5: The Hong Kong yoyo page.

When this section opens, a device is loaded and appears in the hand of the animated male person, as shown in the figure. The device produces different sounds, when the user moves the mouse cursor over different buttons and levers.

In addition to the pages described above, the site contains pages with possibilities to create and send your own Flash[®] animated postcard. More, a design gallery is found, which is a *Beatnik*[®] device for creating larger music loops with alternative mixes, a download section with screensavers and start up screens, and a page with information about the TV-show and the people behind it.

5.3 Eurovision Song Contest

Quoting the description of the web site from the corporate site of Paregos¹, the purpose of this site was:

"Swedish Television and Aftonbladet wanted a web site for the Eurovision Song Contest that was not just a pale copy of the television show and they wanted it to present the sponsors in a sensible way. The site were steadily the most visited for the weeks before and after the competition. The visitor can compete in a Song Quiz (with other visitors) and be his/her own DJ by mixing his/her own version of ABBA's Waterloo, and so on."

(http://www.paregos.com)



Figure 6: The ESC Home Page

The web site uses some plug-ins for the browser as *Macromedia Flash*®, *Beatnick*® *and Shockwave*®. The main menu is found in the upper left. In the very upper corner, a *back* and a *home* button is found. This as the browser goes into 'full screen mode' when the site starts. At the upper right, changing of language (Swedish or English) could be done as well as information about the design team behind the web site could be found. An *Exit* button is put in the very upper right corner. Further below, in the pink area, there are some high lighted entries. They end up at the same pages as in the menu - but they are more specific in their descriptions. In this area there are also some clickable illustrations, which also are entries to underlying pages. In the very below there are links to external sources. The menu have entries like 'Welcome', 'On-line TV', 'Participate', 'Interactive zone', 'Past years', 'Other fun' 'Site map' 'Ask about ESC' 'News' and more. Under all these entries, lower level entries are found as well.



Figure 7: The' past years' page

The entry 'past years' contains historical facts about the contest; old winners, flops, different kinds of charts as well as more general trends from different decades.



Figure 8: The interactive zone

For instance under 'Interactive zone' you find 'Screen saver', 'mix your own song' and 'pop-quiz' could be found.



Figure 9: The 'Waterloo remix' page

The 'waterloo remix page' is a page where the visitors can remix and record their own version of the song Waterloo by Abba. The dolls on the picture dance, sing and play their instruments. The buttons in the bottom are the remix buttons. They turn red when they are active.

As the page is no longer in use, as already said, the page were used in it's present state. By that, we mean the waterloo remix, information retreival on the 'past years' and interactive postcard. All the film clips and other more Stockholm Event specific material had been taken away.

6 Empirical findings

In the case with Mosquito, the free-surf approach was fairly successful. In most of the cases, the single users never asked for the tasks offered as support. When tasks was

requested, the subjects only needed one or two to get going. There was only one case where the subject requested all of the pre-defined tasks.

The think-aloud method presented some difficulties. Some subjects showed signs of stress, uncertainty and had difficulties verbalizing already in the practice session. These subjects also continued to have problems verbalizing their thoughts throughout the rest of the test session. This affected the level of intervention highly, as the test crew had to encourage the subjects at several times with short intervals, to start verbalizing. In general, the subjects had difficulties verbalizing continuously throughout the test. None of the single users managed to keep a continuous verbalization throughout the whole session, without encouragement from the test crew.

The subjects who worked in pairs solving the pre-defined tasks, had few difficulties verbalizing, compared to those who worked alone free-surfing. These subjects also required less intervention, since the subjects discussed their thoughts with each other in a way that eliminated the need for verbalization encouragement from the crew. However, the outcome of the verbalizations differed between the single-user group and the group working in pairs. The single-users more often expressed subjective, emotional thoughts and gave more information on what impressions the different aspects of the site created. The outcome of the task solving pair' verbalizations, tended to express various strategies and problem solving approaches, rather than subjective thoughts and impressions of the site. This may just illustrate the different instructions the two groups were given, or perhaps that a goal orientated approach of thought is easier to verbalize than a more intuitive process of thinking, based on subjective impressions.

7 Conclusions

Below, some of our findings are summarized. To be noted is that there are both findings related to the fact that there is entertainment web sites we test as well as findings of more general kind regarding usability testing. It might be argued that conclusions should only consider entertainment testing. However, it is our strong opinion that they could not, and not even should not, be separated. This because of the fact that traditional usability testing is the basis, even if we as here test entertainment. The findings of more general kind must not be forgotten and are therefore included in our list of findings below. We

A carefully designed task structure may provide important information, even when the web site is entertainment focused. However, the test leader must be observant to the fact that the subject might want to elaborate more than in traditional tests using tasks.

- Test subjects have difficulties verbalizing their thoughts when interacting with this kind of web sites. From this follow a relatively high level of intervention from the test crew, that need to be considered in the design of the evaluation.
- The behavior of the test persons in contact with this entertainment focused functionality are often very unpredictable and presents a number of factors that are difficult to control during the test. Therefore, the evaluator must be very

careful and flexible in the test design, and be prepared to change and refine methods during the evaluation in order to compensate for unforeseen behavior and effects.

- The free-surf approach is not to be trusted alone, but is effective when complemented by additional tasks offered as support.
- Subjects working together reduce the level of intervention, but differs from subjects working alone in terms of content in their verbal reports.

Even if web sites are entertainment focused, traditional usability methods may provide useful tools for practical evaluation, if chosen with respect to the functionality of the site to be evaluated.

References

- Bevan, N. (1998). Usability Issues in Web Site Design. In Proceedings of UPA'98. Washington DC.
- Borges, J. A., Morales, I., Rodrigues, N. J. (1996) Guidelines for Designing Usable World Wide Web Pages. In Proceedings of ACM Conference on Human Factors in Computing Systems, CHI'96.
- Borges, J. A., Morales, I., Rodrigues, N. J. (1998). Page Design Guidelines Developed Through Usability Testing. In Forsythe, C., Grose, E., Ratner, J., Human Factors and Web Development. Lawrence Erlbaum Associates, Publishers. Mahwah, NJ, USA.
- Dix, A., Finlay, J., Abowd, G., Beale, R. (1998). Human-Computer Interaction. Prentice Hall, Europe.
- Dyer, R. (1992). Only Entertainment. Routledge. London, UK
- Gaines, B. R., Shaw, M. L. G., Chen, L. L-Y. (1996). Utility, Usability and Likeability: Dimensions of the Net and Web. In Proceedings of WebNet96. San Francisco, CA, AACE (Association for the Advancement of Computing in Education.
- Grose, E, Forsythe, C., Ratner, J. (1998). Using Web and Traditional Style Guides to Design Web Interfaces. In In Forsythe, C., Grose, E., Ratner, J., Human Factors and Web Development. Lawrence Erlbaum Associates, Publishers. Mahwah, NJ, USA.
- Karat C-M., Pinhanez C., Karat J., Arora R., Vergo J. (2001). Less Clicking, More Watching: Results of the Iterative Design and Evaluation of Entertaining Web Experiences In Proceedings of Interact'2001, Tokyo, July 2001.
- Mayhew, D., J. (1999). The Usability Engineering Lifecycle a practitioner's handbook for user interface design. San Francisco, CA, Morgan Kaufmann Publishers, Inc.
- Mayhew, D.J. (1998). Introduction in In Forsythe, C., Grose, E., Ratner, J., Human Factors and Web Development. Lawrence Erlbaum Associates, Publishers. Mahwah, NJ, USA.
- Nielsen, J. (1993). Usability Engineering. Academic Press.

- Nielsen, J. (1999). User Interface Directions for the web. In Communications of the ACM. Vol.42, No.1.
- Olsson, C. (2000). To measure or not to measure: Why web usability is different from traditional usability. In Proceedings of WebNet2000, San Antonio, TX
- Pine II, B. J. & Gilmore, J.H. (1999) The Experience Economy. Work Is Theatre & Every Business a Stage. Harvard Business School Press, Boston.
- Pinhanez, C. Karat, C-M., Vergo J., Karat., Arora R., Riecken D., Cofino T. (2001) Can Web Entertainment Be Passive? To appear in Proc. of International World Wide Web 2001 (IWWW'01), Web and Society track, Hong Kong, May 2001
- Preece, J. (1994). Human-Computer Interaction. Addison-Wesley .
- Ratner, J. (1998). Easing the Learning Curve for Novice Web Users. In Forsythe, C., Grose, E., Ratner, J., Human Factors and Web Development. Lawrence Erlbaum Associates, Publishers. Mahwah, NJ, USA.
- Schneiderman, B. (1997). Designing information-abundant web sites: issues and recommendations. International Journal of Human-Computer Studies. No. 47, (pp.5-29)
- Shneiderman, B., (1998) Designing the User Interface: Strategies for effective Human-Computer Interaction, Addison-Wesley
- Spool, J., Scanlon, T., Schroeder, W., Snyder, C., DeAngelo, T. (1999) Web Site Usability: A Designer's Guide. Morgan Kauffman Publishers Inc.
- Vora, P. (1998). Human Factors Methodology for Designing Web Sites. In Forsythe, C., Grose, E., Ratner, J., Human Factors and Web Development. Lawrence Erlbaum Associates, Publishers. Mahwah, NJ, USA.