

# **Affective computing vs. usability ?**

## **Insights of using traditional usability evaluation methods**

Charlotte Wiberg  
Department of informatics, Umeå University  
S-901 07, SWEDEN  
Charlotte.wiberg@informatik.umu.se

### **ABSTRACT**

Evaluating affective interfaces in order to provide input for designers is a challenge for the CHI community. One question is to what extent traditional evaluation methods, used for evaluating traditional usability, are applicable at all, and if they need revision. The purpose of this project was to gain an understanding in how applicable traditional usability evaluation methods are in understanding people's experiences of affective systems, in this case Entertainment web sites. Empirical techniques as well as inspection methods were used on a number of web sites. The results show that the methods are applicable but need revision. When it comes to the development of inspection methods, the challenges include finding proper heuristics to support the experts in using Heuristic Evaluation, providing conditions for experts which bridge the gap between evaluation and authentic use, developing complementary methods for use in combination with existing methods etc. In empirical evaluation of entertainment in the context of web usability, the most crucial aspect might be to consider how to arrange a setting that is as natural and authentic as possible when evaluating fun, as this seems to be important for the results. Overall, the results of the study clearly show that important aspects of affective interfaces can be revealed by using traditional usability evaluation methods – aspects which should be considered early in the design phase.

### **Background**

Research about affective computing and interfaces rarely discuss methodological issues regarding evaluation. More, the focus is on the character of the system, interface or interaction *per se* (Wiberg, 2003). There are however a few studies with a more specific focus on evaluation of affective interfaces, for instance, on how users react to entertainment technology (e.g., Pagulayan et al, 2003, Karat & Karat, 2003, Desmet, 2003). Some researchers argue that completely new methods are needed to deal with fun and pleasure in the area of HCI (c.f. Thomas and Macredie 2002). This might well be true, but as we have so little knowledge about how traditional usability evaluation works in the context of fun and entertainment work, it is difficult to argue for new approaches. Further studies are much needed (c.f. Carroll & Thomas, 1988; Thomas & Macredie, 2002; Pagulayan et al, 2003, Karat & Karat, 2003, Desmet, 2003; Nielsen, 2003, Monk et al, 2002). Arguably, usability evaluation methods can have a substantial impact on designing pleasurable and enjoyable systems and web sites (c.f. Pagulayan et al, 2003, Nielsen, 2003). Therefore, even though extending traditional usability to include evaluation of fun an entertainment appears to be a sensible research objective in the context of existing HCI/Human Factors research, this research provides little guidance on how this objective can be accomplished. This conclusion constituted a point of departure for the work presented in the thesis and the purpose was to gain an understanding in how applicable traditional usability evaluation methods are in understanding people's experiences of affective systems, in this case Entertainment web sites (EWSs).

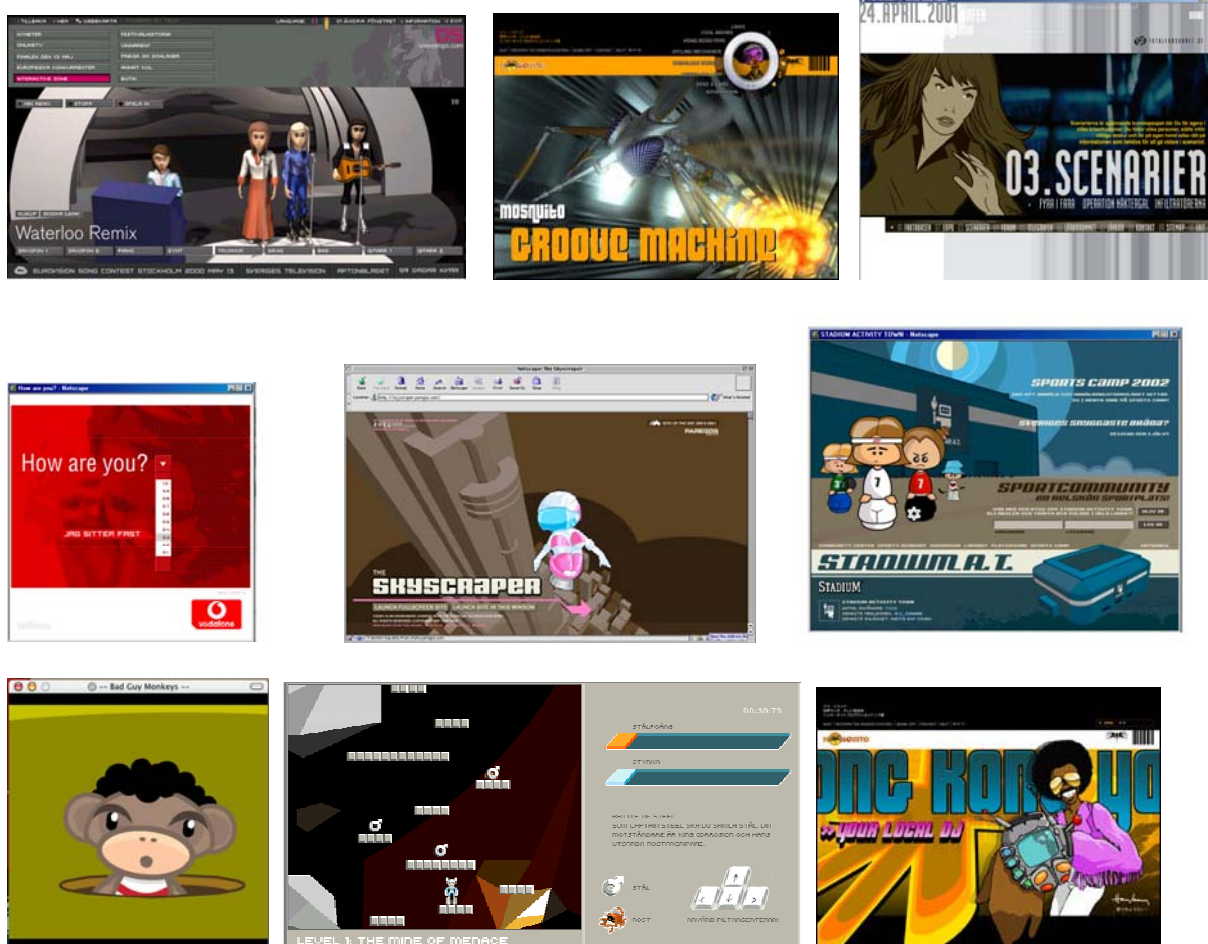
### **The study**

An entertainment web site has many faces – it is almost every web site which include any type of feature or content which is entertainment related. However, in this study we defined EWSs by describing commonly included features. Features commonly found in entertainment web sites are presented below:

1. *Entertainment information* – information about the theme of the web site, jokes etc.
2. *Downloadable items* – screensavers, pictures etc.
3. *Small 'stand-alone' games* – 'Memory' or such.
4. *Other features dependent on plug-in technology* – Re-mixing of music etc.

5. *High quality graphic design*
6. *Edutainment content*
7. *Communication with others* – chats, virtual meeting rooms etc.

The EWSs in the study included one or more of the above mentioned features and the were: (1) Eurovision Song Contest, (2) Mosquito, (3) Totalförsvaret (Total Defence), (4) Skyscraper, (5) 'How are you?' – Vodafone, (6) Activity Town – Stadium, and (7) Jernkontoret – Captain Steel.



The overall strategy in the study could be described as follows: Common usability methods were used in the study for evaluation of entertainment web sites to assess their suitability for elucidating relevant information about EWSs. The aim of applying traditional usability methods was to establish whether they needed further revision and re-design. The findings of the study indicated that the methods needed to be further revised and re-designed to become more applicable. Therefore, the methods were revised and re-designed accordingly. The re-designed methods were subsequently applied in evaluation of the same, or additional, entertainment web sites, to establish whether the re-design resulted in any differences in applicability. In other words, the aim of the new application was to find out if the changes in the methods resulted in changes in the outcome of the evaluation. The methods were judged on the basis of the applicability, i.e. to what extent the methods could inform design of EWSs. This was compared to earlier steps in the study. Finally, on the basis of the results from the study, an improved methodology for evaluating entertainment web sites was presented. The methods evaluated in this study was *Think aloud protocol*, *Interviews*, *Questionnaires*, *Heuristic Evaluation* and *Design Walkthrough*.

### **Empirical usability evaluation of entertainment web sites**

The empirical evaluations conducted in this study produced a number of findings about the aspects of evaluation procedures that should be taken into account when evaluating fun and entertainment in case of entertainment web sites. Below, the findings are summarized, first with regard to the specific conditions examined in the study followed by a general methodological discussion.

Conducting an empirical usability evaluation involves making numerous decisions about the concrete design of the evaluation procedure, which would be optimal for the purposes and general context of the evaluation. A number of aspects, or “dimensions” of the design of evaluation procedures were identified in usability evaluation research as important and potentially problematic. Our study was designed to provide empirical evidence about the importance and relative advantages of these dimensions by comparing various controlled conditions employed in, the study. This evidence can be summarized as follows:

#### *Pairs vs. individuals*

Testing entertainment web sites in pair settings works well, particularly when children and teenagers are tested. In some cases a pair session design must be regarded as an unauthentic situation, that is where the web sites are mainly intended for individual users.. However, results from the study show that authentic use of web sites designed for single use often occurs in pairs. This was clearly shown in the part of the study that involved teenagers, who use EWSs in collaboration with others, for instance, at school. In these cases evaluating in pairs is more authentic than single user evaluation. When testing pairs of subjects, such things as domination, ‘showing-off’ and competition within the pairs must be taken into account. Furthermore, when testing pairs, it is important to always be aware of what is being evaluated, i.e. the interaction between the subjects or the interaction between the subjects and the web site.

#### *Structured vs. unstructured activities*

Traditionally, the use of structured tasks is a common approach in the context of usability evaluation. In this study, the subjects were asked to make evaluations that included both structured and unstructured tasks. As many EWSs are exploratory in nature, providing subjects with unstructured tasks appears to be a reasonable approach. However, depending on the type of entertainment web site evaluated, a structured approach with specified assignments for subjects to complete is not such a bad approach. The main reason for this, according to the results from the study, is that some subjects are frustrated when the assignment is too unstructured or free. Breakdowns occurred in some sessions for this very reason. However, in highly exploratory web environments or where only one task is concerned, for instance in web sites which are comprised only of a game of some kind, the use of unstructured tasks is a more applicable approach.

#### *Testing children vs. adults*

Children as subjects are more spontaneous and more willing to explore. In successful evaluations, where no breakdowns related to the evaluation per se occur, it is possible to obtain data of high quality from them. If children are the target group of the evaluated EWS, some aspects might be impossible to test on any other group. However, it might still be worthwhile to also include adult users in these evaluations, since adults are often better at thinking in abstract terms and verbalize more easily.

#### *Written vs. oral answers to questions regarding entertainment*

Oral answers are to be preferred when asking questions regarding entertainment, because of the subjectivity of the answers and the possibility of asking follow-up questions.

Finally, our tests showed the importance of being situated and intuitive as an experimenter, if useful results are to be obtained when testing entertainment. As a subject, to laugh in a silent crowd is difficult.

### **Evaluation of entertainment web sites using inspection methods**

The main findings from the last part of the study where inspection methods were used, refined and revised, can be summarized as:

*Providing experts with general information about web sites* - It is important that the information about the intentions and aims of the EWSs evaluated is as extensive as possible for valid judgments about the web site to be evaluated. All included experts agreed on the necessity of extending the information about the

web sites' intended target group as well as on the originators' goals for the web site, as interpreted by the designers.

*Changes in the heuristics – language and functional related heuristics* - There seems to be a relation between functional aspects and fun and entertainment aspects in EWSs. The number of heuristics changed from eight to ten in the last evaluation using inspection methods. The additional heuristics were function related. Overall, the experts were positive to this change, and the results from the evaluations of the web sites also show that the heuristics were widely used, which may indicate the need for this type of heuristics – even when entertainment web sites are evaluated.

*The 'free-surf' approach is retained in the methodology* - When evaluating the usability of any system, it is always important to set up a use situation, which is as authentic as possible. This is also true for evaluations of EWSs. However, as some of the experts commented, it may be more difficult in these cases. The 'free-surf' approach was highly valued by the experts in evaluating entertainment web sites. The reason for this was that the evaluation session, as designed in the first place, turned out to be far from an authentic use session of entertainment web sites. The experts could not escape from the fact that they were evaluating the web site and not entertaining themselves. This differs from evaluating pure function, where the difference between evaluation and use is less. For this reason, the 'free-surf' approach remained in the overall methodology for evaluating entertainment.

*Ranking of suitability of heuristics in meta-evaluation* - The meta-evaluation was introduced into the study mainly to serve as a tool for the study of methods as 'objects of study' and not to supply any information to the process of evaluating entertainment web sites as 'objects of study'. However, the experts implied that this was a valuable tool, even in the latter case. The reason lay in the nature of the entity 'entertainment web site', which must be considered highly individual. In some entertainment web sites, playability is very important and in others playability is not applicable at all. The meta-evaluation was seen as a tool to mediate this applicability in each case.

*A possibility to give an overall judgment or review* - Fun and entertainment are difficult to judge just by investigating their parts. It needs a more holistic approach, where the greater whole is bigger than the sum of its parts. This part of the overall methodology came from an idea developed by some of the experts early in the study. The differences between the concepts of 'evaluation' and 'reviewing' were highlighted, where evaluation often is seen as 'revealing problems' but reviewing is more about 'giving an overall judgment'. In the context of entertainment this seemed relevant. This approach was tested in the last inspection method evaluation, and two types of results indicated its importance in the methodology. (1) The overall judgment, given in the reviews, did not always correlate with the balance between positive and negative comments given in the Heuristic Evaluation, i.e. the rate of negative comments could be high, but the overall review might still be positive. (2) The second type of result indicating the importance of the overall review was the number of positive responses from the experts. In general, they were very positive about the presence of this approach in the overall methodology.

## **Future work**

Overall, findings from this study point out that valuable findings for designers regarding aspects of fun and entertainment in entertainment web sites can be obtained if evaluations are conducted using applicable usability evaluation methods. Because of this, is extremely important to continue the effort to develop methods and techniques for usability evaluation – both for inspection methods and empirical usability evaluation methods. When it comes to the development of inspection methods, the challenges include finding proper heuristics to support the experts in using Heuristic Evaluation, providing conditions for experts which bridge the gap between evaluation and authentic use, developing complementary methods for use in combination with existing methods etc. In empirical evaluation of entertainment in the context of web usability, the most crucial aspect might be to consider how to arrange a setting that is as natural and authentic as possible when evaluating fun, as this seems to be important for the results. In addition, it is crucial to consider carefully the level of intervention. The setting up of such conditions as testing in pairs and providing unstructured tasks are just one step on the road to success in evaluating fun and

entertainment in the context of web usability. Further steps have to be taken and other conditions have to be explored.

The strongest impression after all the use sessions conducted in this study as well as in other related projects, is how extremely important the findings are, when external sources are used as subjects in an empirical usability evaluation. No matter how many design awards or prizes the designs or designers have won, and no matter how experienced the expert conducting expert evaluations is – it will always be impossible to predict everything that happen when authentic users of a system are investigated. Even if we feel dissatisfied with our methodologies, and even if we have to struggle to meet challenges in designing and conducting these evaluations, the effort is always worthwhile, considering the interesting and important results this type of evaluations produce.

## References

- Carroll, J.W. & Thomas, J.C. (1988). Fun. SIGCHI Bulletin. January 1988, Vol.19 No.3.
- Desmet, P. (2003). Measuring emotion: Development and application of an instrument to measure emotional responses to products. In *Funology: From usability to Enjoyment*. (eds.) Blythe, M., Overbeeke, K., Monk, A.F., Wright, P. Human-Computer interaction series Vol.3. Kluwer Academic Publishers. (111-123)
- Karat, J. & Karat, C-M. (2003). That's entertainment! In *Funology: From usability to Enjoyment*. (eds.) Blythe, M., Overbeeke, K., Monk, A.F., Wright, P. Human-Computer interaction series Vol.3. Kluwer Academic Publishers. (125-136)
- Monk, A. (2002). Fun, communication and dependability: extending the concept of usability. Closing plenary at HCI2002, (for Human factor Advanced Module)
- Nielsen, J. (2003). User empowerment and the fun factor. In *Funology: From usability to Enjoyment*. (eds.) Blythe, M., Overbeeke, K., Monk, A.F., Wright, P. Human-Computer interaction series Vol.3. Kluwer Academic Publishers. (103-105)
- Pagulayan, R.J., Steury, K.R., Fulton, B., Romero, R.L. (2003). Designing for fun: User-testing case studies. In *Funology: From usability to Enjoyment*. (eds.) Blythe, M., Overbeeke, K., Monk, A.F., Wright, P. Human-Computer interaction series Vol.3. Kluwer Academic Publishers.
- Thomas, P. and R. D. Macredie (2002). "Introduction to The New Usability." *ACM Transactions on Computer-Human Interaction*, Vol. 9(No. 2): 69-73.
- Wiberg, C. (2003). A Measure of Fun. Extending the scope of web usability. Ph. D. thesis. Department of Informatics. Umeå University.