

Fun in the Home: Guidelines for Evaluating Interactive Entertainment on the Web

Charlotte Wiberg, PhD

Department of Informatics
Umeå University,
901 87 Umeå, SWEDEN
charlotte.wiberg@informatik.umu.se

Abstract

In recent years, we have witnessed a dramatic change in how ICT (Information- and Communication Technologies) is used, and examples of usage of ICT are allocation of leisure time, learning, man-to-man communication, e-shopping and much more (Bradley, 2001). Today, the computer has moved to the kitchen, the living room, and even the garage, to support the activities going on there – food recipes are browsed from the kitchen and in the living room a TV-show is displaying supporting web sites for chatting, presentation of further information, voting, etc. Children have PC's, on which they communicate with others, play games, search the web and so on. Because of this change in usage of ICT, the focus of design concerns has expanded from predominantly functional aspects of ICT systems to overall user experience. The trend towards experience has direct implications for usability evaluation. Since experience is considered an important aspect of the quality of various products, it should be evaluated. One possible, if not universally accepted, approach is to consider evaluation of experience as a case of usability evaluation. However, the existing methods cannot be employed. When the focus is on experiences rather than on more functional aspects of systems, a revision of usability methods is required. This paper presents a novel approach in how to conduct Heuristic Evaluations (Nielsen, 1993) on ICT mediating interactive entertainment. New heuristics, so called *funology heuristics*, were developed and empirically tested on, what is called, entertainment web sites. These funology heuristics were also combined with novel methodological approaches in order to fulfill the requirements when evaluating fun on the web.

1 From efficiency in the office to fun in the home

Traditionally, in the 70's and 80's, computers, then known as EDP (electronic data processing) were used mainly in work life. This technology was used to streamline administrative work processes and more and more people were involved using mainframe computers and personal computers (PC's) (Bradley, 1986; Bradley, 2001). The quality of applications was mainly measured in efficiency for the organization – the more efficient the system was for the organization the better. Some researchers pinpointed the interaction between the user and technology as significantly important in relation to quality of the systems. The reason for this was mainly that negative effects started to show up, like fatigue and physical and psychological stress. Researchers started to focus on these problems and disciplines like Psychology, Psycho sociology, Sociology, Ergonomics and Computer Sciences made research efforts, focusing on the problems. The sub-disciplinary research field of HCI (Human-Computer Interaction) was born (c.f. Preece, 1994; Monk and Gilbert, 1995).

A key concept in HCI is usability. It concerns, for instance, making systems safe, easy to learn and easy to use (Preece, 1994). It originates from 'software psychology' in the 1970s, which was a related discipline to experimental psychology (Shneiderman, 1980 in Ehn & Löfgren, 1997). The first usability study reported in HCI research, was presented by Roberts and Moran (1982), and was an evaluation of text editors. Here, the first attempt to divide the concept of usability into various dimensions was made. Thus, usability was divided into:

- *Time* to perform edit tasks by experts
- *Errors* made by experts
- *Learning* of basic edit tasks by novices
- *Functionality* over all possible edit tasks

Throughout the years, the concept of usability has been described and conceptualised in a similar way as above. Only slight changes have been done. One of the most commonly used description of usability nowadays originates from Jakob Nielsen (1993). Here, usability aspects is specified as (1) Learnability, (2) Memorability, (3) Error rate, (4) Efficiency and finally (5) User satisfaction. Overall, the concept of usability has focused mainly on to what extent a user understands an ICT system or not.

As mentioned above, HCI is a sub-disciplinary research field and researchers in Psychology, and more specifically Experimental Psychology, can draw on a long line of various kinds of experiments in human behavior. These experiments have also had a major impact on the evaluation of usability in the HCI community in that many of the methods used in HCI originate from Experimental Psychology. Then methods in the latter field have focused mainly on first generating a hypothesis and then gathering quantitative data in order to test the hypothesis. The focus on data is vital, arising as it did from a collective reaction to the theories of “armchair” psychologists at the turn of the century (Monk and Gilbert, 1995). In general, usability evaluation methods are divided into two types of methods, i.e. empirical evaluation methods, or user tests, and inspection methods, or expert evaluation methods (c.f. Karat, 1997). These both types of methods vary in pros and cons. The empirical methods includes real users which is the ideal situation in finding true usability problems, of course. However, in some cases it is not suitable to evaluate with end users, for instance in early stages in the design cycle, as users cannot understand how a final version of a system might work based only upon immature prototypes. The inspection methods, on the other hand, have their main disadvantage in that they simulate real use, since experts, most often HCI-experts, conduct them. These cannot completely foresee every real world problem. The strength of these methods, however, is that they can be used in early stages in the design process, as experts most often has no problem in understanding how the final versions of a system will work based on a prototype. Another advantage is that they are less context dependent since they do not really include a user situation – they are a simulation *per se*.

In recent years, we have witnessed a dramatic change in how ICT (Information- and Communication Technologies) is used, and examples of usage of ICT are allocation of leisure time, learning, man-to-man communication, e-shopping and much more (Bradley, 2001). Technological development has rapidly given us the opportunity to use technology everywhere, since it has become mobile, embedded, ubiquitous, etc. (c.f. Wiberg, 2001). ICT has also increasingly become adopted in our homes – overall, since the 70's the usage has moved out of the office into our leisure time wherever we are. In the 90's, the home usage of ICT included mainly administrative activities, conducted on a PC. Today, the computer has moved to the kitchen, the living room, and even the garage, to support the activities going on there – food recipes are browsed from the kitchen and in the living room a TV-show is displaying supporting web sites for chatting, presentation of further information, voting, etc. Children have PC's, on which they communicate with others, play games, and search the web and so on. A backbone of ICT usage in the home is the web (or to be more exact the world wide web) (c.f. Petrov, 1997).

Because of this change in usage of ICT, the focus of design concerns has expanded from predominantly functional aspects of ICT systems to overall user experience. This is partly due to a wider scope of current applications of IT, which now include not only systems with function as their most critical aspect, but also systems designed to target the overall user experience. The shift in design focus also partly originates from more demanding users, who are becoming increasingly mature in their use of IT systems, and therefore also expect the systems to provide them with experiences and not just functionality.

The trend towards experience has direct implications for usability evaluation. Since experience is considered an important aspect of the quality of various products, it should be evaluated. One possible, if not universally accepted, approach is to consider evaluation of experience as a case of usability evaluation. However, the existing methods cannot be employed. When the focus is on experiences rather than on more functional aspects of systems, a revision of usability methods is required. The traditional usability evaluation methods used in research and in large corporations, deal adequately with functional aspects, but to date are not particularly suitable for dealing with the experiential aspects of systems. Evaluating users' experiences and not just users' understanding of the system requires new procedures, indicators and analysis. In other words, the basic concepts and methods of usability need to be further developed before they can be applied to this new field.

Limitations and possibilities can be identified for usability evaluation of new technological artefacts, systems, applications and services used and explored at home. The problem is twofolded: (1) Traditionally, usability have

been evaluated in standard usability labs with PCs installed to be used in more or less designed and controlled experiments. This environment is very much different from today's usage of ICT in the home when it comes to environmental aspects – we use ICT in combination with other media as TV, newspapers and telephones. An ethnographical approach could be considered since this technique is regarded as a useful approach in evaluation of situations where contextual aspects are critical. Conducting ethnographic studies in organizational settings could be difficult but could be done fairly easy. However, to conduct the same type of evaluations in home settings is far more complicated. Because of this, the method might disqualify as suitable in the case of evaluating ICT for home usage. (2) Furthermore, the usability of a system was earlier mainly highlighting aspects such as efficiency, learnability, errors and other aspects related to understanding. Today, when experiences is in focus, these aspects are not enough because of two reasons; The quality measurements above need to be complemented with aspects such as 'funability', aesthetics and so on. Furthermore, experiences are seen to be of a holistic kind of phenomena, i.e. even if correct aspects are found and measured in such an environment, the complete experience is bigger than the sum of the parts. Something happens in the combination of the parts in an experience. This is problematic in that it might be difficult to evaluate since it is to some extent undiscovered as an entity or criterion in the measurement or evaluation of the usability.

Since much of the ICT use in the home somehow relate to the web, it seems reasonable to choose the object of study on the web in order to give some kind of guidelines in how to evaluate ICT in the home. For this, a specific type web sites is chosen, i.e. entertainment web sites (EWSs). These are web sites with the purpose of mediating interactive entertainment in different shapes. Furthermore, since inspection methods have the advantages of being simulations of real use it is reasonable to believe that these have great potential in being applicable in this case. The reasons for this is that as they simulate real use of a system, this helps us in relation to the contextual problem, as described above. This suits our purposes in striving to find suitable evaluation approaches for home use of ICT, in this case EWSs. However, the inspection methods still include quality measurements connected to the understanding- and efficiency paradigm rather than covering experience- and entertainment dimensions. This issue has to be taken care of.

This paper presents a novel approach in how to conduct Heuristic Evaluation (Nielsen, 1993), which is perhaps the most widely known inspection method. This method was invented in early 90's to be used as a method for systems where aspects such as learnability, efficiency, error rate were in focus and where it, for some reason, was not suitable to include end users. Another highlighted advantage of this method was that it was less time-consuming than empirical evaluation, and therefore less expensive. As the use of ICT have shifted to include more than only efficiency aspects, the traditional heuristics in Heuristic Evaluation are to some extent out of date. New heuristics is needed. Because of this, an extensive study was conducted and a new list of heuristics was developed where entertainment is in focus. These heuristics cover aspects important to evaluate in ICT mediating interactive entertainment. The list of heuristics is also combined with methodological considerations covering, for instance, rating of importance of each heuristic in every single case as well as some other aspects, important to incorporate when evaluating interactive entertainment.

2 Entertainment web sites

Web sites come in different shapes and have different purposes. Briefly, they can be divided into different types, such as information retrieval web sites, where the main purpose is to expose information to the visitor. Other types of purposes can be education, e-shopping, etc. A broad term for one type of web sites where the purpose is interactive entertainment is entertainment web sites (EWSs). Features commonly found in entertainment web sites are presented below:

- *Entertainment information* – information about the theme of the web site, jokes etc.
- *Downloadable items* – screensavers, pictures etc.
- *Small 'stand-alone' games* – 'Memory' or such.
- *Other features dependent on plug-in technology* – Re-mixing of music etc.
- *High quality graphic design*
- *Edutainment content*
- *Communication with others* – chats, virtual meeting rooms etc.

Below, some examples of entertainment web sites is presented:

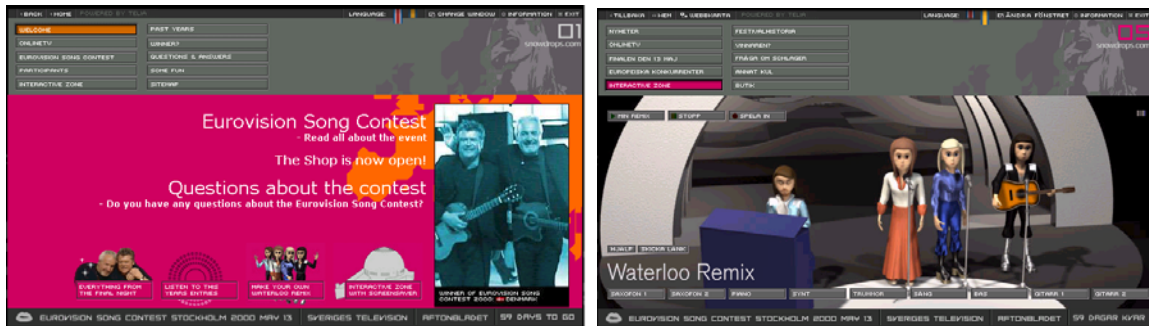


Figure 1 & 2: Eurovision Song Contest web site year 2000

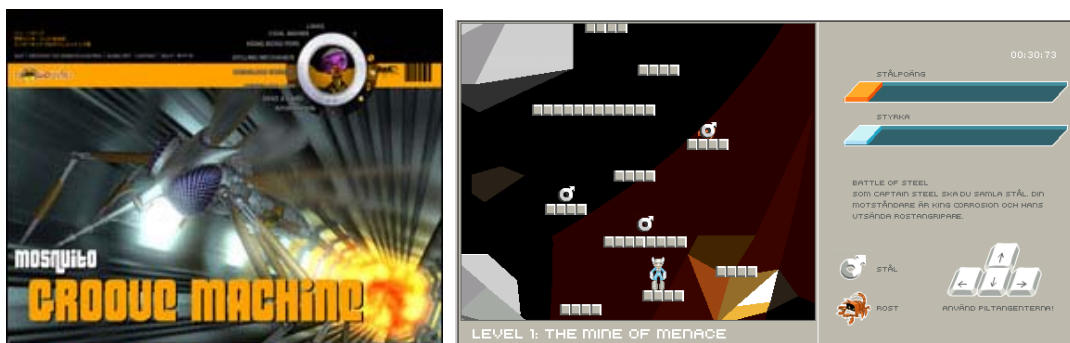


Figure 3 & 4: TV-show Mosquito web page and a game 'Captain Steel'

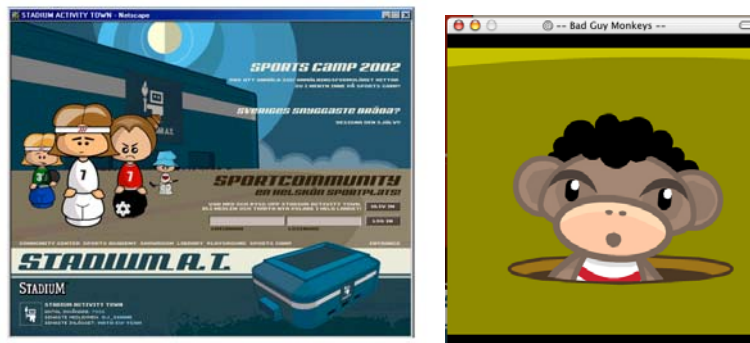


Figure 5 & 6: The Stadium Community 'SportsCommunity' including the game 'Bad Guy Monkeys'

3 Heuristic evaluations of entertainment web sites

In the study, a large number of entertainment web sites were evaluated applying Heuristic Evaluation. The study design could be understood as being iterative, i.e. evaluations were conducted with the traditional approach as designed by the originators in early 90's. After this iteration, methodological revisions were done and new evaluations of other entertainment web sites were conducted. New modifications of the method were done and final evaluations were conducted, which gave input to the last methodological changes. These stages are further described later in this section.

3.1 The heuristic evaluation method

The inspection method called Heuristic Evaluation were originally designed by Nielsen and Molich (1990) and the process of Heuristic Evaluation can be described as follows: A specific list of design guidelines, frequently called

'heuristics' is used as a basis for evaluation of a system or application. The evaluator, often a UI expert or designer, reviews the system and comments on usability problems in relation to each heuristic. The most commonly used list of heuristics in Heuristic Evaluation nowadays is as follows. (Nielsen, 1993);

- (1) Visibility of system status,
- (2) Match between system and real world,
- (3) User control and freedom,
- (4) Consistency and standards,
- (5) Error prevention,
- (6) Recognition rather than recall,
- (7) Flexibility and efficiency of use,
- (8) Aesthetic and minimalist design, and
- (9) Help users recognize, diagnose and recover from errors
- (10) Help and documentation

3.2 The structure of the study

The study consisted of three main phases – (1) use of traditional usability evaluation methods, (2) refinement and re-design of these methods, and (3) use of refined and re-designed usability evaluation methods. All three phases will be further developed and described in detail below. An overview of the complete study is shown in Figure 7.

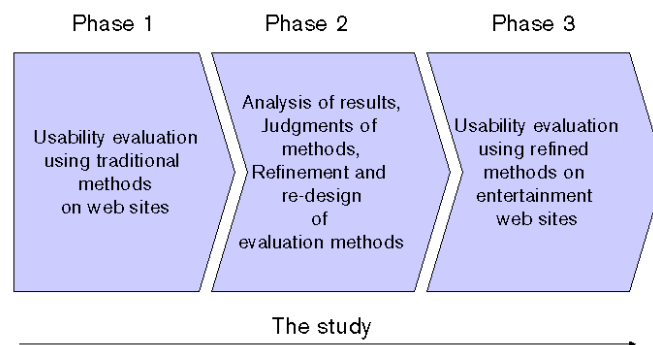


Figure 7. An overview of the study.

3.2.1 Phase 1 – using traditional method

In this phase, two parallel expert groups were employed to evaluate two web sites, one entertainment web site and one so-called information retrieval web site (IRWS). The second web site was used as a control site. The two groups differed in that the first group had ten (10) so-called experienced experts, i.e. HCI researchers and lecturers and the second group included twenty (20) novice experts, i.e. higher-level undergraduate students of informatics, with a specific focus on HCI. The students had a theoretical knowledge about usability evaluation but seldom or never used inspection methods before. To compensate to some extent for the novice nature of the experts in this group, the students were allowed to work in pairs. A total of ten (10) student groups were given the task of conducting a Design Walkthrough and a Heuristic Evaluation of the two web sites. As a help, they received handouts containing full descriptions of what they were to do together with forms on which to report problems and make other comments. The experts were subsequently asked to propose new heuristics, suitable for evaluating fun, which were also included in the documentation. Finally, while the experienced experts were interviewed about the sites and the evaluations, the student experts were given the task of writing a report of their findings and suggestions and completing the handouts.

3.2.2 Phase 2 – Refinement and re-design of evaluation methods

The results from the first phase of the overall study were used as input for the second phase – the refinement and re-design phase. This phase of the study is presented below:

In this phase the data consisted of completed handouts, with reported problems and other comments about the web sites together with experts’ suggestions for new heuristics. The written reports from the student experts and the answers from interviews with experienced experts provided further input about more general methodological aspects. All this data material constitutes input for the interpretation and analysis in this second part of the study, conducted by evaluators in the research project. The analysis and interpretation of the data generated a new output in the form of a new set of heuristics for Heuristic Evaluation for further exploration in the last phase of the study, as well as more general suggestions concerning how to evaluate fun using Inspection Methods.

3.2.3 Phase 3 – Usability evaluation of EWSs using new and refined evaluation methods

In this phase, two entertainment web sites were evaluated by the same experienced experts as in the first phase of the study. They used an approach called ‘free surf’ –which could be compared with Design Walkthrough (c.f. Karat, 1997). The experts then conducted Heuristic Evaluation with a new set of eight (8) heuristics – only fun related. Further, experts conducted a ‘meta-evaluation’ of the suitability of each heuristic for the evaluated web site. Experts were given a set of handouts similar to those in the first phase, which they used to report findings and suggestions for further revision of the methodology. Finally, the experts were interviewed about both their results from the evaluations of the web sites and their proposals for further developments of the evaluation methodology. This data material, i.e. the handouts and the answers from the interviews, were then subjected to further interpretation and analysis, and evaluators made new changes in the evaluation methodology. A further round of inspection method evaluations of EWSs was conducted by experts. Once again, two entertainment web sites were evaluated using the ‘free surf’ methodology approach and Heuristic Evaluation, this time with a revised set of heuristics. The number of heuristics was now ten (10) the earlier tested eight heuristics with additional two heuristics covering understanding and more traditional heuristics. In addition, a meta-evaluation of the suitability of the heuristics for each of the web sites was also used in this final round. An overall review of the complete web site was also added to the methodology to comply with the demand made by experts in the earlier phases. Finally, when the evaluations were complete, the experts answered a questionnaire in free text. The data from all the methods included in the revised methodology were then analyzed and conclusions were drawn from this third phase of the study.

4 Funology heuristics

As a result of the last phase of the study, a list of, so called, *funology heuristics*, suitable for evaluating entertainment web sites, is presented. It is worth mentioning that these are tested being applicable, in the sense that they verbalize the aspects of web sites of which expert evaluators only have a ‘guts feeling’, only on entertainment web sites. However, it is reasonable to believe that these heuristics also is applicable to other interactive ICT with a focus on entertainment. Below, the list of *funology heuristics* is displayed:

Heuristic number	Name of heuristic Description
1	Visual impression vs. expectations On entertainment web sites it is important to create an expectation, for instance by providing a ‘fancy’ and attractive design and suitable music. However, it is also important that the expectations created are fulfilled. If they are not, the whole impression of the web site is spoiled, and the user might feel disappointed or cheated. Do not let the visual impression create expectations the interaction cannot meet.
2	Exploratory design The design should entice users to explore the web site. The users should be motivated to stay on the web site and explore the possibilities it has to offer.
3	Playability – gameplay It is important to clearly visualize the gameplay, otherwise there is a risk that the user will feel cheated in that he/she had expected to be able to do other things than are actually possible.
4	Durability and lifetime – amount of content Is there enough content for a longer ‘visit’ or ‘stay’? This is important as this type of web site is often expected to support or entertain a user, frequently for a longer period. It should not feel as if the web site has been ‘emptied’ after just a short period of use.

5	<p>Coherence between chosen design and desired mediated feeling</p> <p>Does the choice of colors reflect the mood mediated, as intended by the designer? Is the design appealing? Are the colors attractive? How should the menus be designed and where should they be placed? Does the web site catch the user's attention and curiosity and make a generally strong impression. (For intentions about the mediated mood and purpose of the web site – as stated by the designers – please check the description of the web site in the documentation)</p>
6	<p>Clarity of genre – design for the right target group</p> <p>The web page must be designed, i.e. be given form and filled with content, relevant for the intended target group. This also means that it is important to discover what target group really is the receiver. (For information about the intended target group of the web site please see the description of the web site in the documentation)</p>
7	<p>Balance between information and entertainment</p> <p>If entertainment is one goal to be accomplished on a web site, there should be an attempt to achieve a balance between entertainment and information. Entertainment web sites are often so-called information containers for banners and entertainment material, for instance multimedia content etc. However, the total amount of content on a web site is limited, and there must be a balance in order to achieve a good result. The content must be relevant to the purpose of the web site. (For information about the purpose of the web site please see the description of the web site in the documentation)</p>
8	<p>Originality and freshness (uniqueness)</p> <p>It is important that the information is relevant and 'fresh'. Otherwise the feature has no reason to be on the web site.</p>
9	<p>Consistent navigation</p> <p>Does the navigation work in a consistent way, externally and internally? Externally, based on general standards and guidelines for (this type of) web site, and internally, i.e. is the navigation structure of the web site consistent throughout the whole page structure?</p>
10	<p>General function-related aspects</p> <p>Does the web site have an overall functionality which is comprehensible? Does the system give feedback about what is going on? Can a user find out by him/herself if mistakes have been made and remedy them? Are there any so-called emergency exits for users if they have made mistakes? What about help and documentation?</p>

It is also worth mentioning that this list is developed to be used with another methodological idea and that is a meta-evaluation form, in which the evaluator judges the importance or applicability of each heuristic for the specific object of study. The results of the heuristic evaluation of an EWS is then to be understood as a combination between the reported results based on the funology heuristics and the rating of the importance of each heuristic in the single case.

5 Conclusions

The findings of this study indicate that traditional usability evaluation methods can be applied to evaluation of entertainment and fun in the context of web usability. More specifically, the heuristic evaluation method were found to produce relevant and potentially useful evidence when applied to evaluation of entertainment web sites. However, it was also found that existing methods have serious limitations and methods needed to be further developed and revised in order to become more applicable. It was demonstrated that the revision of the methods resulted in an increased quality of the methods as analytical tools for usability evaluation. Therefore, the main conclusion of the study can be formulated as follows; while the underlying concepts and principles of web usability can be employed in evaluation of entertainment and fun, specific evaluation methods need to be revised.

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

5.1 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.

5.2 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.

5.3 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.

5.4 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.

5.5 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.

6 Discussion

Overall, findings from the study in this thesis 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 evaluation methods. Because of this, it 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.

7 References

- Bradley, G (1986). Psykosocial arbetsmiljö och datorer. Stockholm: Akademiliteratur
- Bradley, G.(Ed.) (2001). Humans on the Net. Information and Communication Technology (ICT) Work Organization and Human Beings. Stockholm: Prevent.
- Ehn, P. & Löfgren, J. (1997) Design of Quality-in-use:Human-Computer Interaction Meets Information Systems Development. In Helander, M., Landauer, T.K, Prabhu, P.(eds.) Handbook of Human Computer Interaction Second, completely revised edition. Elsevier Science B.V.
- Karat, J. (1997). User-Centered Software Evaluation Methodologies. Handbook of Human-Computer Interaction. M. Helander, T. K. Landauer and P. Prabhu, Elsevier Science B.V.: 689-704.
- Monk, A. And Gilbert, N. (1995). Inter/disciplinary research. In Monk, A. and Gilbert, N. (eds.) Perspectives on HCI. Diverse Approaches. Academic Press. London
- Nielsen, J. (1993). Usability Engineering. Academic Press.
- Nielsen, J. & Molich, R. (1990). Heuristic Evaluation of User Interfaces. Proceedings of CHI'90: 249-256.
- Petrov, P. (1997). En kritisk granskning av undersökningar om Internetanvändare. TRITA-NA-D9713 • CID-24, KTH, Stockholm, Sweden 1997
- Preece, J. (1994). Human Computer Interaction, Addison-Wesley.
- Roberts, T.L. & Moran T.P.(1982). Evaluation of Text Editors. In Moran, T.P.(ed.)Eight Short Papers in User Psychology. Xerox Palo Alto Research Centers, Palo Alto, CA
- Wiberg, C. (2003). A Measure of Fun. Extending the scope of web usability. Ph. D. thesis. Department of Informatics. Umeå University.
- Wiberg, M. (2001). In between Mobile Meetings: Exploring seamless ongoing interaction support for mobile CSCW, PhD-thesis, Department of Informatics, Umeå University, Umeå, Sweden.