Subjective Usability Evaluation Criteria of Augmented Reality Applications

Valéria Farinazzo Martins[1(✉)], Tereza Gonçalves Kirner[2], and Claudio Kirner[2]

[1] Faculdade de Computação e Informática, Universidade Presbiteriana Mackenzie, São Paulo, Brazil
valeria.farinazzo@mackenzie.br
[2] UNIFEI – Universidade Federal de Itajubá, Itajubá, Brazil
{tgkirner,ckirner}@gmail.com

Abstract. This paper presents an extensive list of attributes to measure the usability of Augmented Reality applications. These attributes were collected from a systematic review of papers in journals and conferences at the global scope, arising from the productions of the last five years, such as main areas, more used attributes, AR environments, and number of papers by year. We used the most relevant studies in the literature to compose the organization and categorization of the main usability attributes discussed and used in Augmented Reality. Finally, we propose a set of questions on these Augmented Reality usability attributes, based on established questionnaires and also experience in the evaluation of the authors.

Keywords: Usability evaluation · Augmented reality · Usability testing

1 Introduction

Augmented Reality (AR) can be described as a view of the real and physical world which incorporates additional information to augment this view [1], i.e., it is a system that supplements the real world with virtual objects synthesized by computer, making these two worlds coexist in the same space, respecting the following properties: a) it combines real and virtual objects in a real environment; b) it is executed interactively in real time; c) it overlaps real and virtual objects with each other; d) it can be applied to all human senses, including hearing, touch, smell and strength [2]. Thus, this technology has the advantage of allowing the use of tangible and multimodal actions that facilitate interaction and motivate users [3, 4].

On the other hand, usability is a system quality requirement that contains aspects related to the efficiency when using the system, ease of learning, subjective satisfaction from the user and adequacy to specific patterns; it is the process of assuring interface usability and guarantee that the user's demands be meet [5, 6]. Although the aspects for usability mentioned above are conceptually clear, it is difficult to use these definitions in practice. When the evaluation is made through empirical studies, the researchers need to decide about metrics for each factor [7]. Usability metrics are usually divided into objective and subjective. The first is related to the effectiveness and efficiency of

R. Shumaker and S. Lackey (Eds.): VAMR 2015, LNCS 9179, pp. 39–48, 2015.
DOI: 10.1007/978-3-319-21067-4_5

the user with the system, while the subjective measures collect the user opinions about the system usually through questionnaires or interviews. The objective criteria can be further divided into quantitative and qualitative [6].

Although AR presents the same core usability challenges as traditional interfaces – for instance, the potential for overloading users with too much information and making it difficult determine a relevant action -, AR aggravates some of these problems because multiple types of augmentation are possible concomitantly, and proactive applications run the risk of overwhelming users [1]. There are certain peculiarities inherent in AR applications that should be evaluated in a more specific context, such as the use of markers and multimodal interaction in 3D space.

Although AR is being studied for over 40 years, only a few years researchers began to look after the formal evaluation of these systems [8]. A question that may arise is how the developers of AR systems have been following minimum criteria that can guarantee its quality in terms of usability. In many cases, AR applications have been developed without following a methodology or using a traditional software development methodology, which does not consider the peculiarities of this kind of applications. Furthermore, RA uses a natural interface that allows a non-conventional interaction, essentially making the analysis of their quality, especially related to the usability of their applications. It is worth noting that although this quality may be responsible for the success of applications, knowledge about the opinion of the users, their satisfaction and frustrations in the use of these applications is still rather limited [9, 10].

The lack of formal assessments in the area has already been pointed out by [8]. This, on the world scenario, revealed a low number of papers related with some evaluation technique of AR applications. According to this study, less than 8 % of AR-related papers, from 1993 to 2007, were evaluated according to the following parameters: perception, user performance, collaboration and usability. Moreover, they showed that only seven papers out of 169 evaluation techniques include usability.

In our work, 992 papers containing the keywords “Augmented Reality” and “Usability” in their abstract were found. However, only 58 papers contained in fact some kind of usability evaluation, especially subjective aspects of the evaluation. It is also clear that many of these studies do not address a usability study correctly.

Therefore, the focus of this paper is to present the main attributes that have been used to evaluate the usability of AR applications on the world scenario since 2008 until 2013. The papers considered most relevant, which bring more specific attributes for evaluation of AR are discussed in this paper. In addition, we propose a set of questions on these AR usability attributes, based on established questionnaires and also experience in the evaluation of the authors.

This paper is organized as follows. Section 2 discusses the methodology of study development. Section 3 presents the results and discussions of the research. Finally, Sect. 4 presents the conclusions on the subject.

2 Materials and Methods

In order to reach the main usability attributes used on AR searches, were considered the following steps: conducting a systematic review of papers in Portuguese (Brazilian), from 1998 to 2013; categorization of the attributes [11]; conducting a systematic review of

papers in English, from 1998 to 2013 (the research conducted by [8] brings the studies on the world scenario until 2007); completion of the categorization of the attributes, using papers in English; and preparation of an assessment instrument (a set of questions according to the attributes). On the whole 992 papers (227 in Portuguese and 765 in English) were collected from "Portal de Periódicos CAPES" (this Brazilian site contains a database of major scientific journals and can be configured to the area of Computing), and IEEE Xplore and ACM Digital Library to found conference papers. From these papers, 16 in Portuguese and 42 in English papers could be used for this research. The research protocol developed for this study is adapted from the models proposed by [12] and [13].

From this research, 51 attributes were found. They were divided into nine categories (System Interaction, Application Interface, Representation, Sensory and Behavioral Aspects, Motivation and Effort, Spatial Association, Internal Aspects and Configuration, General Functionality and Others Attributes), according to [11] and the second systematic review. Some of these attributes are used to evaluate the usability of any interactive computer system and any others are specific for AR applications. Subsequently, a research was conducted in studying the main usability satisfaction questionnaires: QUIS [14], PUEU [15], NAU [16], NHE [17], CSUQ [17], ASQ [18], PHUE [19], PUTQ [20], and USE [21].

In the next stage, these 51 attributes were mapped to these existing questionnaires, as indicated in Table 1. Some attributes had more than a question mentioned by the same questionnaire and sometimes the same attribute was in different questionnaires. Others attributes could not be mapped to any existing questionnaire. It was necessary to standardize questions for the new questionnaire. Questions not mapped had to be created.

In addition to finding the main attributes for assessment of RA applications, it was desirable to know:

- What usability attributes are being more widely used, general and specific?
- Which areas are being more developed on usability analysis issues?
- Which kind of AR systems has been used more frequently: markers or markerless?
- Which kind of AR environment has been used more frequently: desktop or mobile?
- How many users, on average, has been used to perform a usability test for RA applications?

3 Results and Discussion

In the process of evaluating usability of AR applications, papers were selected by its importance and it is possible verify, from 2008 to 2013, that their distribution, through the years, took place, as shown in Fig. 1. One might understand a gradual increase in the number of papers that focus some kind of usability evaluation in the last year. The areas covered by these papers can be seen in Fig. 2.

It was found that the greatest amount of usability evaluation work occurred in the "Education" area; Secondly, we have "Base –AR" that means tests performed with new algorithms, techniques and user's tools utilized for evaluation; thirdly, we have "Arts".

Figure 3 presents the types of AR environments. Figure 4 presents the information about the use or not of markers.

Table 1. Attributes and questions of augmented reality

Attributes	*Questions*	*References*
	SYSTEM INTERACTION	
A1. Ease of handling marker/object	1. It is easy for handling marker/object	Authors
A2. Use of mouse and others unconventional devices	2. The control of cursor is compatible with movement	PUTQ
A3. Adequacy of audible and visual prompts	3. The audio and text information (such as on-line help, on-screen messages and explains) provided with this system was clear.	NHE, QUIS, CSUQ
A4. Latency/Response Time	4. The response time is appropriated, for example, the images appear briefly on the screen.	Authors
A5. Performance of the interface	5. It is possible to complete the tasks and scenarios quickly using this system.	PUEU, CSUQ, QUIS
A6. Depth Perception	6. It is possible to perceive that objects are three-dimensional	Authors
	APPLICATION INTERFACE	
B1. Feedbacks sent to the user	7. The application provides timely feedback about all processes, system status.	NHE, QUIS, PHUE
B2. Ease of learning the application	8. Learning to operate the application would be easy for me.	PUEU, NAU, USE, QUIS, CSUQ, PUTQ
B3. Ease of use	9. Performing tasks is straightforward.	PUEU, USE, CSUQ, QUIS
B4. Visibility of system status	10. The application informs about its progress.	QUIS
B5. Control over the activity	11. I can feel in control of the activities and tasks.	USE
B6. Flexibility and efficiency of use	12. I would find the application to be flexible to interact with.	PUEU, USE, QUIS, CSUQ
B7. Consistency and Standards	13. The colors, labels, objects, audios, feedbacks of the application are consistent. 14. It is difficult to make mistakes or errors in the application.	NHE, PHUE, PUTQ
B8. Error Prevention	15. It is difficult to make mistakes or errors in the application.	NHE, PHUE
B9. Minimal Action	16. The application requires the fewest steps possible to accomplish what I want to do with it.	USE, PUTQ
	REPRESENTATION	
C1. Presentation of Information	17. The information is presented in a very organized and easy way to view.	QUIS
C2. Appearance and arrangement of the elements of the screens including text,icons, graphics and colors	18. The organization of objects (text, icons, 3D object) on the application screens was clear.	CSUQ
C3. Artistic aspects able to express beauty, style and elegance	19. The interface of the application is pleasant.	CSUQ, USE
C4. Visual Realism	20. The application presents visual realism.	Authors
C5. Minimalist design and information overload	21. The scene is clean, only necessary objects are represented.	Authors

(*Continued*)

Table 1. (*Continued*)

C6. Quality and relevance of 3D objects	22. The 3D objects are well-designed and relevant.	Authors
C7. Faithful representation of 3D objects	23. The 3D objects faithfully represent reality.	Authors
C8. Quality and relevance of the animations	24. The animations are well-designed and relevant.	Authors
C9. Accuracy	25. The application displays a few errors, allowing a good interaction.	NAU, Authors
C10. Audio Aspects	26. Audio aspects of the application are good quality.	Authors
C11. Media Integration	27. The application integrates multiple media: audio, text, video	Authors
SENSORY AND BEHAVIORAL ASPECTS		
D1. Capacity of immersion and user participation	28. The application provides capacity of immersion and user participation.	Authors
D2. Clarity and intuitiveness of the application behavior	29. I can easily navigate through the application, because it is clear and intuitive.	Authors
D3. Effectiveness	30. It is possible to complete the tasks and scenarios effectively using the application.	PEUE, CSUQ
D4. Efficiency	31. The application helps me being more effective.	PUEU, NAU, USE, CSUQ
MOTIVATION AND EFFORT		
E1. Cognitive Load	32. The application minimizes user memory load.	NHE, PHUE, PUTQ
E2. Perception of the importance of the topic	33. It is important using this application.	Authors
E3. Level of user satisfaction	34. I am satisfied with it.	USE, CSUQ, NAU
E4. Ability to promote fun	35. It is fun to use.	USE
E5. Fatigue and eye strain	36. If I use the application for a long period of time, I feel tired, and my eyes feel tired.	Authors
E6. Comfortable	37. I felt comfortable using this application.	CSUQ
E7. Confidence in use	38. I felt confident of using the application.	Authors
E8. Attractiveness/ Motivating	39. Use of the application is pleasant	PUEU, CSUQ
E9. Recommendable	40. I would recommend it to a friend.	USE
SPATIAL ASSOCIATION		
F1. Navigation	41. Browse application is easy, as it is easy to find and interact with information and desired objects.	Authors
F2. Spatial compatibility of knowledge	42. The virtual objects perfectly integrate with the real environment.	Authors
F3. Mapping	43. Give the user a way to preview where to go, what will happen. Give the user a way to review / return to previous contexts.	PHUE
INTERNAL ASPECTS AND CONFIGURATION		
G1. Application Access	44. It is easy to access the application.	Authors
G2. Quick access to information	45. The information I need are easily found and accessed.	Authors
G3. Need for calibration	46. Before I start using the application, the system starts the calibration process	Authors
GENERAL FUNCTIONALITY		
H1. Software utility and meeting intended goals	47. The application accomplishes the goals it sets itself.	Authors
H2. Well integrated Features	48. I can understand what features of the application that the markers represent, as well as realize the integration of different media (video, text, sound, three-dimensional objects)	Authors

(*Continued*)

Table 1. (*Continued*)

	that make up the application	
OTHERS ATTRIBUTES		
I1. Ease of collaboration	49. The application allows me to collaborate with others easily.	Authors
I2. Security	50. The application does not allow the user to make mistakes that he cannot undo, or enter areas that do not. have a clear exit.	Authors
I3. Creativity support	51. The application supports creative activities of the user, allowing, for example, that the user can interact in different ways and to modify the environment.	Authors
I4. Portable	52. I can use the application on various platforms such as smart phones, desktops, laptops and tablets.	Authors

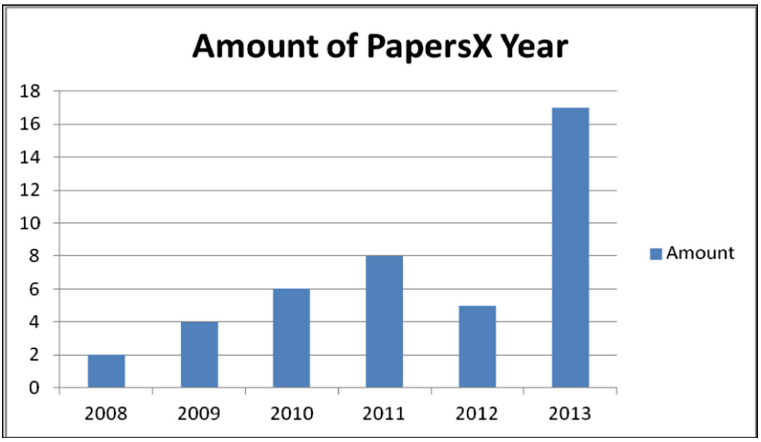

Fig. 1. Amount of papers distributed over the years 2008 to 2013

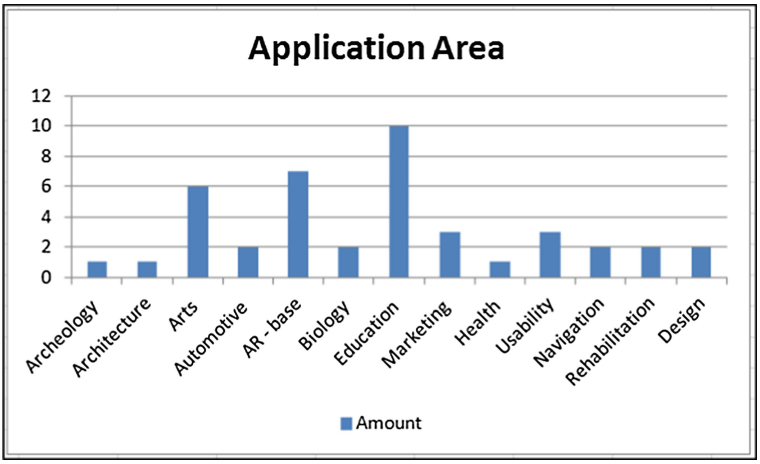

Fig. 2. Amount of papers distributed in the 13 areas

Through Fig. 3, it is possible to see that even the development of AR focuses on applications for desktops.

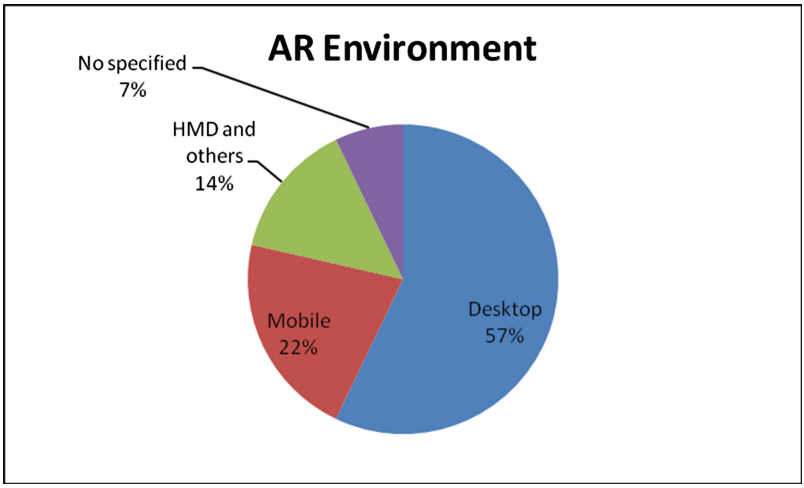

Fig. 3. Types of AR environment

On the other hand, it is possible to see that markerless use surpassed the use of AR markers (Fig. 4).

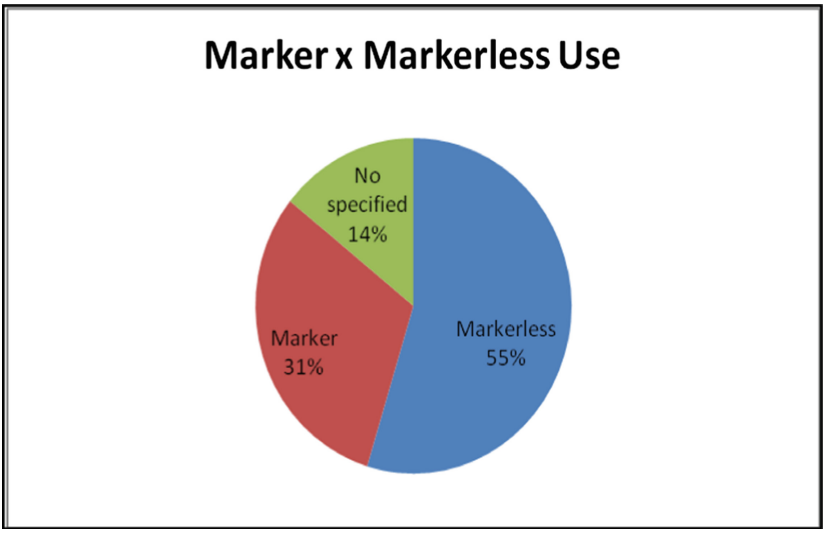

Fig. 4. Use of marker or markeless

We could see, through the systematic reviews, that the most commonly used attributes (which appeared at least ten times) for these papers are: attractiveness, ease of learning the application, ease of use and level of user satisfaction.

Other information obtained by analyses of these papers makes reference on the users' quantity during usability tests: in average 30,167 people are used. This information was obtained from the information read the papers with the number of users for testing (71 % of papers containing such information).

3.1 Integration of Criteria and Peculiarities of AR

Our initial studies were performed from a systematic review of evaluation of usability of AR systems in the Brazilian scenario, indicated by [11]. This study ranked the usability attributes for AR applications according to eight dimensions. Throughout this study it was possible to establish:

- Interaction with the system: refers to the mechanisms that allow the user to interact with the system (markers, audio, mouse, unconventional devices).
- Application Interface: is related to issues, such as interface /system. It is presented to the user in terms of ease of use and learning, flexibility of use, among others.
- Representation: relates to aspects perceived by the user, such as the appearance that the interface presents to the user.
- Sensory and Behavioral Aspects: are related to how the interface can be intuitive, promote user adaptation, in addition to immersion.
- Motivation and Effort: relates to how the system can hold the user's attention, motivating him to use and reuse the application.
- Spatial Association: refers to the distribution in space of the virtual and real environment, overlapping them. Therefore, the virtual objects inserted onto the scene should have proportional sizes to the real environment.
- Internal and Configuration aspects: refers to aspects that allow the application to be ready for use.
- General Functionality: is related to the criteria of software utility when it meets the objectives and requirements that are proposed.

Including the second systematic review, now with papers on the world stage (written in English), were found 51 attributes. A category the most, called "Others attributes" was inserted to the categorization, incorporating attributes that had no relation to the other original eight categories [11]. These attributes mapped in questions are presented in Table 1.

4 Conclusions

This paper presented a study on the topic of usability evaluation of Augmented Reality applications in the world scenario, since 2008 until 2013.

The central goals of this paper were to investigate the state of the art of AR evaluation and also extract the most relevant attributes that have been used in the world scenario to evaluate AR applications.

Evidence shows that 94 % of researched papers have used the word "Usability" and "Augmented Reality", but not really deal with this subject or use erroneous and /or irrelevant criteria. Only 58 out of 992 papers researched in journals and conferences contained some relevance and could be used in order to contribute to the assessment area.

Many of the general attributes are presented to evaluate interactive applications, such as: ease of learning the application, ease of use and level of user satisfaction.

On our first systematic review, we find 42 usability attributes for Augmented Reality, in the Brazilian scenario. On our second systematic review on the world stage, nine attributes were found. Moreover, 85.71 % of the attributes found in the first systematic review were present in the second systematic review. These attributes were divided into nine categories.

A second goal of this work was to propose a questionnaire for the attributes found by systematic review. This questionnaire is grounded in questionnaires established in the literature and the authors' experience in evaluation of application of Augmented Reality.

It is worth mentioning the prevalence of researches on usability evaluation of Augmented Reality with markerless (Fig. 4), especially in papers related to conferences. This stands out the need to check if there are specific attributes for this class of applications that should be addressed.

As future work, we intend to study and propose attributes to evaluate the usability of Mobile Augmented Reality.

References

1. Singh, M., Singh, M.P.: Augmented reality interfaces. IEEE Internet Comput. **17**(6), 66–70 (2013)
2. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Comput. Graphics Appl. **21**(6), 34–47 (2001)
3. Kirner, C., Kirner, T.G.: Virtual reality and augmented reality applied to simulation visualization. Simul. Model. Curr. Technol. Appl. **1**, 391–419 (2007)
4. Azuma, R.T.: A survey of augmented reality. Presence **6**(4), 355–385 (1997)
5. Nielsen, J.: Usability Engineering. Elsevier, The Netherlands (1994)
6. Sommerville, I.: Software Engineering, 9th edn. Addison Wesley, Boston (2011)
7. Skov, M.B., Stage, J.: supporting problem identification in usability evaluations. In: Proceedings of the 17th Australia conference on Computer-Human Interaction Citizens Online Considerations for Today and the Future Computer-Human Interaction Special Interest Group (CHISIG) of Australia, pp. 1–9, November 2005
8. Dünser, A., Grasset, R., Billinghurst, M.: A survey of evaluation techniques used in augmented reality studies. In: Human Interface Technology Laboratory, pp. 5–1, New Zealand (2008)
9. Martins, V.F., Soares, L.V., Mattos, V.B., Makihara, E.L., Pinto, R.N., de Guimarães, M.P.: Usability metrics for augmented reality applications. In: Computing Conference (CLEI), XXXVIII Latin American, pp. 1–10. IEEE (2012)
10. Martins, V.F., de Guimarães, M.P., Correa, A.G.D.: Usability test for Augmented Reality applications. In: Computing Conference (CLEI), XXXIX Latin American, pp. 1–10. IEEE October 2013

11. Martins, V.F., Kirner, T.G., Kirner, C.: Estado da Arte de Avaliação de Usabilidade de Aplicações de Realidade Aumentada no Brasil. In: Anais do WRVA 2013, Jataí, Brasil, (2013)
12. Biolchini, J., Mian, P.G., Natali, A.C.C., Travassos, G.H.: Systematic review in software engineering. system engineering and computer science Department COPPE/UFRJ, Technical report ES, 679(05), 45 (2005)
13. Kitchenham, B.: Procedures for performing systematic reviews. Keele, UK, Keele University, 33, (2004)
14. Chin, J.P., Diehl, V.A., Norman, K.L.: Development of an instrument measuring user satisfaction of the human-computer interface. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems pp. 213–218. ACM, May 1988
15. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989)
16. Nielsen, J.: Heuristic evaluation. Usability Inspection Methods **17**(1), 25–62 (1994)
17. Lewis, J.R.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int. J. Hum. Comput. Interact. **7**(1), 57–78 (1995)
18. Lewis, J.R.: Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. ACM SIGCHI Bull. **23**(1), 78–81 (1991)
19. Perlman, G.: Practical heuristics for usability evaluation. (last retrieved on 1st) (1997)
20. Lin, H.X., Choong, Y.Y., Salvendy, G.: A proposed index of usability: a method for comparing the relative usability of different software systems. Behav. Inform. Technol. **16**(4–5), 267–277 (1997)
21. Lund, A.M.: Measuring usability with the USE questionnaire. Usability Interface **8**(2), 3–6 (2001)