ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/286562759

A Checklist to Evaluate Augmented Reality Applications

Conference Paper · May 2014

DOI: 10.1109/SVR.2014.17

CITATIONS
20

READS
4,171

2 authors:

Marcelo De Paiva Guimaraes
Universidade Federal de São Paulo
121 PUBLICATIONS **194** CITATIONS
SEE PROFILE

Valéria Farinazzo Martins
Universidade Presbiteriana Mackenzie
71 PUBLICATIONS **163** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Uso da Realidade Virtual na reabilitação de pacientes com AVC e avaliação utilizando conectividade cerebral View project

Objetos de aprendizagem interativos: conceito, ontologia e uso View project

All content following this page was uploaded by Marcelo De Paiva Guimaraes on 03 February 2016.

The user has requested enhancement of the downloaded file.

A checklist to evaluate Augmented Reality Applications

Marcelo de Paiva Guimarães
Universidade Aberta do Brasil-UNIFESP/Programa de Pós-graduação da Faculdade Campo Limpo Paulista
São Paulo, Brasil
marcelodepaiva@gmail.com

Valéria Farinazzo Martins
Faculdade de Computação e Informática
Universidade Presbiteriana Mackenzie
São Paulo, Brasil
valeria.farinazzo@mackenzie.br

***Abstract*— Augmented reality applications merge virtual content with the real world, with real-time interaction and they have inherent characteristic, such as the lighting conditions, use of sensors, and the user position. These applications are very different from conventional applications that use mouse and keyboard, so they require a usability evaluation to verify whether they achieve their goals and satisfy users. We developed a checklist to measure the usability of augmented reality applications in a practical manner. This checklist was developed adapting the ISO 9241-11 and Nielsen Heuristics for an augmented reality context, and criteria created by us. Five experts evaluated two applications to test the checklist. In both cases, the evaluation clearly identified key problems in the application design. They also evaluated the checklist. Analyzing the results, we could determine that the checklist appears to be a possible solution to evaluate usability of augmented reality applications.**

Keywords*— *Augmented Reality, Heuristics, ISO 9241-11, Usability, Usability Inspection

I. INTRODUCTION

Augmented Reality (AR) is a technology that overlays virtual content (i.e., 2D graphics, 3D graphics, sound, and video) onto the real world, with real-time interaction [1]. This technology enhances the capacities of the human senses via a device, such as a mobile phone, a tablet, or a desktop. AR has been employed in many domains such as entertainment, industrial, military, commercial, health, and marketing applications [2-7].

Like other products, the success of AR depends on the final acceptance by users. Thus, the success of an AR application is achieved if the quality is known, which requires the development of a usability measure solution. Usability measurement is implemented using formal and practical methods, which involves the assignment of values to subjective aspects. These methods must include procedures that compare the information obtained with the application goal, by measuring variables such as effectiveness, efficiency, and satisfaction, where representative test users perform representative tasks in the user context. These measures are not conclusive per se, but they can be used to directly infer the degree of success achieved by the application.

Although traditional usability evaluation methods may be able to discover some problems with AR applications, none of the current methods really fit the specific needs of such systems. Dunser and Billinghurst [8] showed an overview of methods that have been used to evaluate AR applications. Usually, the evaluation is for a specific AR application. Kostaras et al. [9] and Zainuddin et al. [10] presented guidelines for the evaluation of AR applications based on their strengths and weaknesses, using interviews, questionnaires, surveys or usability testing. However, they did not describe the heuristics or metrics used for such evaluations.

Dunser, Grasset, and Billinghurst [11] showed that less than 8% of the AR studies published between 2003 and 2007 included evaluations of the following parameters: perception, user performance, collaboration, and usability. Indeed, only 7/169 studies included techniques for usability evaluation. Thus, we created a checklist based on ISO-09241 (part 11-guidance on usability [12]) to measure the usability of AR applications in a practical manner. This standard defines usability and identifies the relevant information required for a usability evaluation, and explains how to measure the user performance and satisfaction of Visual Display Terminals. The definition of usability in ISO-09241-11 is:

> *"Extent to which the users of products are able to work effectively, efficiently and with satisfaction."*

This standard suggests that usability can be measured during or after use of a system. It specifies a framework based on the three components of usability and their relationships: the goals of the use of the product; the contexts of its use (a description on the application scope); and usability measures. The standard also defines general variables for measuring usability: effectiveness aims to measure how accurately and completely the goals can be achieved; efficiency is related to the resource costs; and satisfaction measures the degree to which users are free of discomfort and their attitude toward the use of the application. Usability requires a detailed understanding of the user context for a product and the purposes for which usability is being assessed. A weakness of this standard as a quality model is that it is too abstract and not explicitly concerned with time (usability changes over time). A detailed method to evaluate usability is not part of this standard so a solution is required that can be applied repeatedly to the same application as part of an iterative design process. This

978-1-4799-4261-9/14 $31.00 © 2014 IEEE
DOI 10.1109/SVR.2014.17

CPS Conference Publishing Services

standard also provides few information about how to interpret scores from specific usability metrics.

Our approach involves the use of heuristics [13] to measure usability to ISO-09241-11 in the context of AR applications. How to decide the optimal or even appropriate heuristics for usability evaluation of AR applications is a question. Several previous studies have used heuristics to evaluate specific types of applications, such as handheld projection-based application websites [14], handheld projection-based application [15], Voice User Interfaces (VUIs) [16], Game design [17] and Virtual Reality (VR) [18]. Like ISO-09241-11, these studies did not consider the specificities of AR applications. General usability heuristics consider details such as user satisfaction and learning curve (where usability changes over time), but they ignore concepts such as interactions with virtual objects in a real world. Thus, we adapted the Nielsen heuristic [19] and added new specific heuristics to evaluate AR applications, reducing the problem of detecting cosmetic flaws or false positive in the data. The heuristics defined by Nielsen [19] were created primarily for desktop applications in the 1990s. These heuristics cannot be used directly to evaluate AR applications, but are flexible enough to be adapted to AR context.

Heuristic evaluations aim to define the current state of software in order to improve it. The heuristic evaluation applied to an interative design, which progressively refines the application through the feedback result from the early evaluation, can help the application reaches an acceptable level of usability. The task of defining heuristics with suitable indicators is complex and expensive. Each heuristic must be associated with an accurate goal and context, or the results may be invalid.

Figure 1 shows an overview of all the components used to create our checklist. We identified the main components of ISO-09241-11 (application goal, context of use, and usability measures) and adapted them to an AR context. The usability measured was implemented using heuristics adapted for AR applications (Nielsen heuristics and new heuristics). We validated the checklists based on evaluations of two applications. Our initial findings indicate that the checklist can support experts in detecting usability flaws and providing severity ranking of the same.

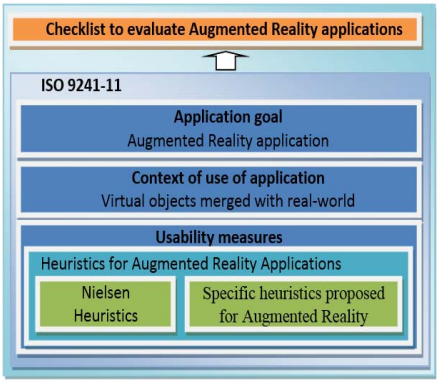

Fig. 1. Usability framework for augmented reality applications.

This study provides a practical solution for measuring the usability of AR applications. This paper is organized as follows. Section 2 describes the ISO 9241-11 components adapted for an AR context. Section 3 explains the heuristics proposed by Nielsen applied for AR context and others created by ourselves. Section 4 presents the checklist created to evaluate usability of AR applications. Section 5 describes the applications used to test the checklist and the results after applying the checklist. Section 6 presents the checklist evaluation. Finally, Section 7 presents our conclusions.

II. APPLICATION OF ISO 9241-11 TO AR

There are several definitions of usability, such as those found in [13,20,21], but we used those established by ISO 9241-11, which defines it as the ability of a product to be used by users to achieve specific goals based on the effectiveness, efficiency, and satisfaction in a specified user context [12].

According to the ISO 9241-11 standard, it is necessary to specify the application goals and the user context to measure the usability of an application. Figure 2 shows all of the sub-components required to assess the usability in AR context. The description of each component is as follows:

1. Application goals: these goals are related to the aim of the application, given the merging of virtual objects with the real world;
2. Context of use is related to:
   (a) User: a description of important user characteristics (e.g., skill, experience, education, and training). Lack of previous knowledge is an example of the characteristics of AR application users;
   (b) Task: activities undertaken to achieve the application goal (e.g., showing the marker to a webcam is an example of task required by our application test);
   (c) Equipment: describes all of the equipment needed. In general, AR applications utilize devices such as cameras or sensors, but some applications require the creation of objects in advance, such as markers or a colored ball;
   (d) Environment: description of the physical environment requirements (e.g., workplace, furniture, temperature, and humidity). If an AR application utilizes a camera, it is dependent on the lighting, position, and user location.
3. Usability measures: the next section describes our solution for verifying the variables used to measure usability (efficiency, effectiveness, and satisfaction).

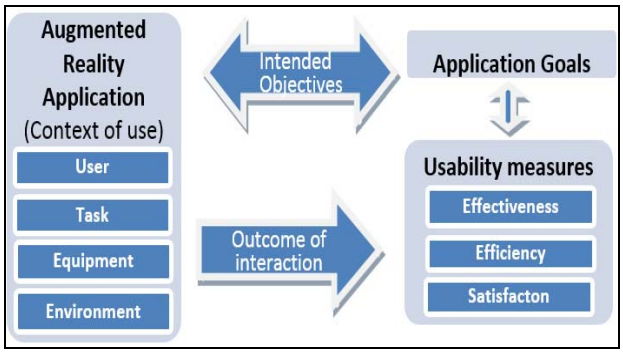

Fig. 2. Usability framework adapted to augmented reality applications.

III. HEURISTICS FOR MEASURING THE USABILITY OF AR APPLICATIONS

To create the usability measures, we reinterpreted the heuristics set proposed by Nielsen [19] and added new specific heuristics, considering also the context of use and the application goals described by the ISO 9241-11. This reinterpretation was essential for creating the checklist (next section). There are other sets of usability heuristics, which are not mutually exclusive (e.g., Norman [22] and Tognazzini [23]), but the Nielsen heuristics are the most popular because of their applicability, simplicity, and low cost. However, they require trained usability experts. Our reinterpretation of these heuristics in the context of AR is as follows.

- Visibility of the system status: evaluates how the system is seen by the user. Users must receive feedback about what is occurring in the system. AR applications utilize tracking systems to determine the virtual content position in a real scene, which must be fast and reliable otherwise users will become lost while interacting with the application;
- Match between the system and the real world: the system design should follow real-world conventions, thereby making information appear in a natural and logical order. Users must interact with the virtual content in the same way as they would in the real world. The object scale and animation must be coherent with the scene;
- User control and freedom: the application must provide freedom so the user can perform actions and undo incorrect actions. If a user presents the wrong marker to the camera, the system must support simple marker replacement, if possible alert the user about the mistake. Actions such as undo and redo must be simple;
- Consistency and standards: the application interface layout and the user interaction must be consistent. Users must interact in the same way with all virtual objects. Each marker must be associated with a specific action or virtual object to avoid mistakes;
- Error prevention: the application must be designed to avoid mistakes and to prevent undesired actions. If this occurs, it is essential that users receive readily understandable messages, such as: “the 3D loader is not working appropriately”;
- Recognition rather than recall: this establishes whether a user can run the application in an intuitive way. The marker functionalities and positioning in the scene must be easy to memorize;
- Flexibility and efficiency of use: users must be able to interact with the application in a fast and flexible manner. Novice users must be able to interact easily with the AR application and interactions with expert users must be facilitated, e.g., they are not required to watch the video instructions whenever the application is started. Furthermore, the user and the marker must be positioned easily in the environment;
- Aesthetic and minimalist design: the system should not provide irrelevant information during the dialogue with the user. Irrelevant information competes with relevant information and eventually the user’s attention is focused on unimportant aspects. The presence of many virtual objects or markers in the same application can lead to an overload of information;
- Help users recognize, diagnose, and recover from errors: the system must indicate problems precisely and make suggestions in a constructive manner. For example, if an unacceptable marker is detected, the system must provide guidance to solve the mistake;
- Help and documentation: it is better if a system can be used without documentation, but the provision of a good procedure and documentation is helpful. Information should be easy to find and should be focused on the user’s task in a concise manner. For example, the system must explain how each marker works. Explanatory videos for AR users are interesting solutions.

We extended the Nielsen heuristics set to the evaluation of AR application, including the effectiveness, efficiency, and satisfaction, according to ISO 9241-11. This allowed us to overcome a drawback of this heuristic method, which initial focus were on desktop applications. The new heuristics we created were as follows:

- Accuracy: how accurate is the system during interactions. The position of the virtual content in the image is determined by the tracking system and it must not vary;
- Environment setup: AR applications require special devices such as sensors and/or cameras. Furthermore, markers may be necessary, such as fiducial markers. The environment setup must be as simple as possible;
- Satisfaction: this measures the degree to which the AR application surpasses user expectation. Interaction is an important aspect of AR applications, and the user must have positive attitudes toward the system.

IV. CHECKLIST TO EVALUATE USABILITY OF AR APPLICATIONS

This study addresses the research question:

Is it possible to develop an usability checklist that is applicable to a broad variety of AR applications?

The question is addressed in a contextualized way. The heuristics set presented in the last section was used to build a checklist. Several tools are available to support usability measurement, such as Ergolist [24], GLIST [25], and SUMI [26]. These tools comply with ISO 9241-11, but they cannot be used directly to evaluate AR applications, so it was necessary to make adaptations.

Our checklist is an informal inspection tool, which is intended to be practical and answered by experts (in around 20 minutes). For each heuristic, we created various items in our checklist. We aimed to identify usability problems and to classify and quantify the problems encountered. We also selected and prioritize the problems that needed to be corrected. This checklist can be applied to finished products or during development.

Much of the inspection work entails classifying and counting the number of usability problems encountered via the interface (user interface problems can lead to reduced usability for the end user). Degrees of severity are given to the problems, which allow to deal with higher priority problems first. During the evaluation, each checklist item is classified according to the following severity rate, which was adapted from Nielsen [19]: 0 (zero) = not a problem or not applicable to this application; 1 (one) = fixed if

extra time is available; 2 (two) = minor problem; 3 (three) = major problem; and 4 (four) = must be fixed or improved. Items classified as high priority (4) must be fixed or another solution should be adopted. Evaluators are also asked to provide their comments about each item.

Each checklist item was classified based on its effectiveness (Table 1), satisfaction (Table 2), and efficiency (Table 3)). This checklist helps the evaluator not to forget heuristics and it enhances the evaluation objectivity, reliability, and reproducibility.

TABLE I. CHECKLIST: VERIFIABLE VARIABLE EFFECTIVENESS

Item	Heuristic
Do you know what is going on during all of the interactions?	Visibility of system status
If the camera or sensor detects more than one marker in the scene, is it possible to specify one?	User control and freedom
Is it possible to execute "redo" or "undo" easily? (i.e., return to a previous state without the virtual object)	User control and freedom
Does the application achieve the goal?	Satisfaction

TABLE II. CHECKLIST: VERIFIABLE VARIABLE SATISFACTION

Item	Heuristic
Is the number of virtual objects in the scene appropriate?	Aesthetic and minimalist design
Is the number of interaction options satisfactory? (marker, keyboard, mouse, joystick)	Aesthetic and minimalist design
Is the user guide satisfactory? (video, text, audio)	Help and documentation
Are you satisfied with the interaction solution?	Satisfaction
Are you satisfied with the freedom to move around during interactions? (e.g., you don't need look directly at the camera constantly)	Satisfaction

TABLE III. CHECKLIST: VERIFIABLE VARIABLE EFFICIENCY

Item	Heuristic
Is the loading time of virtual objects in the scene satisfactory?	Visibility of system status
Are the virtual objects merged correctly with the real world? (position, texture, scale)	Match between system and the real world
Is the virtual object animation coherent with the real world?	Match between system and the real world
Are actions/feedback standardized? (e.g., borders are added to the outside of the tracked object)	Consistency and standards
Is error prevention enabled? (i.e., if the user shows an unexpected marker, is an error message presented to the user?)	Error prevention
Is it easy to remember the application's functionalities? (i.e., is it easy to memorize the functionalities of each marker?)	Recognition rather than recall
What is the learning curve like for novice users?	Flexibility and efficiency of use
Can expert users utilize the application in an optimized manner? (e.g., can they skip introductory videos)	Flexibility and efficiency of use
Is it easy to stand the marker in an appropriate position and orientation to be detected by the camera/sensor?	Flexibility and efficiency of use
Is the user instructed about what to do during the interaction? (e.g., show the marker to the camera or is there a manual)	Help users recognize, diagnose, and recover from errors
Are there specific requirements? (camera, marker, mobile, GPS, user position, lighting, print, calibration)	Environment configuration
Is the tracker system stable?	Accuracy
If the tracker system detects more than one object in the scene, does the application continue to function correctly?	Accuracy

This checklist was created following up the heuristics presented in section 3.The initial version was proposed by the authors, then, using brainstorming techniques, it was discussed with 5 experts. Finally, 5 other specialist in usability and AR evaluated two applications and this checklist.

V. AR INSPECTION

Five usability experts with >5 years experience in evaluation and AR application development applied the checklist to two AR applications from the area of marketing. Each evaluator took around 20 minutes to test each application. The performance tasks defined were to explore all features - previously known - of each application. There is many debates

about ideal sample size (number of experts) to reach effectiveness in a usability study. Nielsen suggested five [27]; Hwang and Salvendy [28] suggested 10±2; Schmettow [29] doubt that 80% of problems can be found with only 10 users or even with 10 experts. We adopted five because both applications are simple and it is not yet known the ideal sample size for finding general AR problems. We must keep in mind that application of our checklist must be a lightweight process, cheap, fast, and easy to apply.

A. Checklist used for the automotive application

The first case study was from the automotive industry and it involved promoting a car. A fiducial marker was utilized to interact with a vehicle (Figure 3). After initialization, the application instructed the user to show the marker (car wheel image) to the camera, which then showed an interactive screen above the recognized marker for the 3D vehicle. Turning the wheel turned the car, so the users could observe the design from any angle they preferred. After clicking on the car, the user could turn on the car lights and book a test drive. The usability problems encountered are shown in tables 4, 5 and 6. The tables show only checklist items that at least one evaluator recognized a problem.

Fig. 3. Automotive application.

TABLE IV. VERIFYING THE EFFECTIVENESS VARIABLE FOR THE AUTOMOTIVE APPLICATION

Checklist	Severity degree	Expert comments
Does the application achieve the goal?	1,2,1,2,1	It achieves the goal, which is simple. It was expected to have a more ambitious goal. The user could customize the vehicle, change the car color, open the door, and drive the car.

TABLE V. VERIFYING THE SATISFACTION FOR THE AUTOMOTIVE APPLICATION

Checklist	Severity degree	Expert comments
Is the number of interaction options satisfactory? (marker, keyboard, mouse, joystick)	4,4,4,3,4	The lack of special effects leads to disinterest.
Are you satisfied with the interaction solution?	2,3,1,3,1	It is difficult to hold the marker with one hand and use the other hand to click the mouse.
Are you satisfied with your freedom to move around during interaction? (e.g., you don't need to look directly at the camera constantly)	2,1,2,1,1	It is necessary for the marker to be seen by the camera.

TABLE VI. VERIFYING THE EFFICIENCY VARIABLE FOR THE AUTOMOTIVE APPLICATION

Checklist	Severity degree	Expert comments
Is the loading time of virtual objects in the scene satisfactory?	0,0,1,0,1	About 5 seconds, which is considered high
What is the learning curve like for novice users?	1,2,1,0,1	The application is not very intuitive. It is available as a demo video and a tutorial (images).
Can expert users utilize the application in an optimized manner? (e.g., can they skip introductory videos)	3,2,3,3,2	The application is simple, so novice and expert users interact in the same way.
Is it easy to stand the marker in the position and orientation required for detection by the camera/sensor?	2,1,2,2,1	Users need to spend time placing the marker in the camera viewing field.
Is the user instructed about what to do during interactions? (e.g., showing the marker to the camera or is there a manual?)	0,1,0,1,0	A video and a manual are available to the users via the home page, which are good.
Are there specific requirements? (e.g., camera, marker, mobile, GPS, user position, lighting, print, calibration)	1,1,2,1,2	The user has to print the marker.

B. Checklist used for the fashion application

The second application promoted jewelry for the fashion industry and it was markerless. The user could virtually try on bracelets, earrings, and necklaces. The application identified the positions of the user's face, neck, or arm and overlaid 3D accessories on them. Figure 4 shows this application. The usability problems encountered are shown in tables 7, 8 and 9. The tables show only checklist items that at least one evaluator recognized a problem.

(a) Virtual earrings (b) Virtual bracelets

Fig. 4. Augmented reality application for the fashion industry.

TABLE VII. VERIFYING THE EFFECTIVENESS VARIABLE FOR THE FASHION ACCESSORY APPLICATION

Checklist	Severity degree	Expert comments
If the camera or sensor detects more than one marker in the scene, is it possible to specify one?	2,3,2,2,3	If another user appears on the scene, the system becomes lost.
Is it possible to execute "redo" or "undo" easily? (i.e., return to a previous state without the virtual object)	1,2,3,3,1	The user can change the type of attachment and accessory at any time.

TABLE VIII. VERIFYING THE SATISFACTION FOR THE FASHION ACCESSORY APPLICATION

Checklist	Severity degree	Expert comments
Are you satisfied with the freedom to move around during interactions? (e.g., you don't need to look directly at the camera constantly)	3,2,2,1,2	I would like more freedom to move.

TABLE IX. VERIFYING THE EFFICIENCY VARIABLE FOR THE FASHION ACCESSORY APPLICATION

Checklist	Severity degree	Expert comments
Is the loading time of virtual objects in the scene satisfactory?	1,0,1,0,0	Three seconds
Are the virtual objects merged correctly with the real world? (position, texture, scale)	4,4,4,4,3	The positions of the accessories in the scenes need to be improved. The bracelets remain stationary on the screen when a user moves.
Are actions/feedback standardized? (e.g., borders are added to the outside of the tracked object)	1,3,1,2,1	After selecting an accessory, the user must remain in a fixed position. The freedom to move is desirable.
Is it easy to stand the marker in an appropriate position and orientation to be detected by the camera/sensor?	2,1,3,2,1	The user must be in a fixed position to calibrate the system.
Is the user instructed about what to do during interactions? (e.g., show the marker to the camera, or is there a manual?)	1,2,1,0,1	The application provides a demo video, an online manual and a user silhouette to help position the user.
Are there specific requirements? (e.g., camera, marker, mobile, GPS, user position, lighting, print, calibration)	1,0,0,1,0	A webcam and a specific user position are required.
Is the tracker system stable?	4,4,4,4,3	It must be improved because it is slow. The tracking performance affected the interaction.
If the tracker system detects more than one object in the scene, does the application continue to function correctly?	3,4,4,3,4	The tracker loses the user if another user appears in the scene.

We did not aim to compare these applications, but we can highlight the following points:

- The main drawback of the automotive application was the lack of interaction options. It was very simple;
- The fashion application was markerless and the user had to remain in a pre-established position, which caused some discomfort for the user;
- The position of the virtual content was not accurate in the fashion application (e.g., earrings position), which caused some frustration for the users. The tracking system must be improved;
- The learning curve was low for both applications, although it was easier to interact with the fashion application.

Based on an analysis of the variable levels for efficiency, effectiveness, and satisfaction in both applications, we can conclude the following:

- The automobile application had good efficiency and effectiveness, but it received a poor satisfaction measurement. The main drawback of this application was the lack of interaction options, which could be improved with enhancement to functions such as car customization, opening the door, and driving the car;
- The fashion application obtained better marks in terms of satisfaction, but the others variables need to be improved. This can be achieved by enhancing the tracking system.

We find that reviews were consistent across the experts and across checklist questions, though some experts disagreed about the rate severity. Even though the results are preliminary, these experimental results proved useful in the validation of the checklist.

VI. CHECKLIST EVALUATION

Checklists reduce the chance of forgetting to verify something important and forces the evaluator to consider each relevant dimension of merit that becomes it a valuable evaluation device when carefully developed, validated, and applied. Although, when poorly designed or misapplied, checklists can results negative impacts, including checklist fatigue, low reliability by adding unnecessary complexity to the evaluation process and delays in the completion of the evaluation.

After the AR inspection (section 5), the five usability experts rated (a 5-point scale ranging from 1 as strongly disagree to 5 as strongly agree is used for the measurement) and commented the following statements (S1...S5) about the variables effectiveness, efficiency and satisfaction for measuring usability, and about the checklist cost-effectiveness:

- S1: The checklist measures the variable effectiveness.
- S2: The checklist measures the variable efficiency.
- S3: The checklist measures the variable satisfaction.
- S4: The checklist has a good cost-effectiveness.

Figures 5 shows the value of the mean for the responses of the expert to the four statements. Efficiency got the highest score. One expert commented: "Provides more details than other variables and the questions are more objective. Questions

are useful and answers would provide good feedback for developers to improve the application". Another expert emphasized the good scope of this variable. This was the variable with more questions. The variable Effectiveness got also a good grade. One expert commented "I suggest much more objective questions". The variable satisfaction reached the lowest grade. Three experts suggested more questions to evaluate satisfaction. All experts agreed that the checklist is very simple-to-use and it justifies the score of cost-effectiveness. Analyzing the results, more objective questions to satisfaction and effectiveness may improve the checklist. However, it is important to keep the checklist a process lightweight, cheap, fast, and easy to apply.

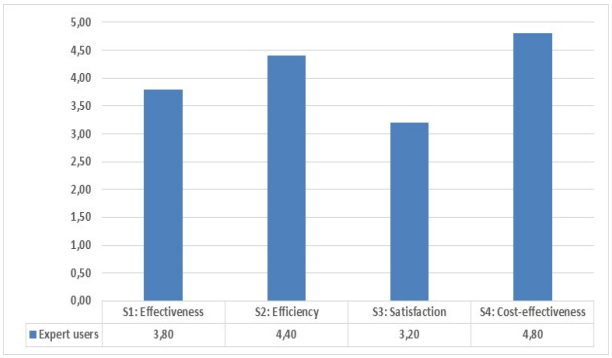

Fig. 5. Checklist evaluated by Expert users.

The main limitation of the study is the number of AR application evaluated and quantity of experts. With only two applications evaluated by five experts, we cannot be sure how well our checklist will generalize to our target of all AR applications. A thorough checklist evaluation will require additional testing, where more experts are involved, and where the checklist is used to evaluate different AR applications. However, we believe the results provide us with enough evidence to conclude that the checklist is well suited to uncovering important usability problems in AR context.

VII. CONCLUSIONS

This study developed a checklist based on ISO 9241-11 and Nielsen Heuristics for evaluating the usability of AR applications, which can be applied to finished products or during development by expert users. This ISO standard does not provide details about usability measures and it was implemented using an heuristic evaluation. Thus, we create a checklist to measure the usability of AR applications, and it was validated by five expert users. Checklists encourage experts to be more consistent about the way they review each item, and have been used to facilitate and speed up the evaluation task. It was applied to two case studies. The checklist presented in this paper provides a practical way to evaluate usability or AR applications.

The requirement for a formal and practical method to evaluate AR applications is clear from this study. The proposed checklist is a possible solution to this problem that it was approved by the experts in this study. Evaluation is expected to become part of the development process for AR applications, which will allow the users to perform tasks that achieve the AR application goals of effectiveness, efficiency and satisfaction.

To reach a usability evaluation of high quality, we must achieve consensus on what is high quality. One needs to agree on a conceptual definition of quality as well as a set of quality criteria, that can be implemented using a checklist. This checklist is not a universal solution for evaluate AR applications - such evaluation method does not exist even within traditional graphical user interfaces. However, this checklist can be recommended to evaluate similar applications.

In future research, we plan to improve the checklist applying it to additional AR applications to evaluate the merging of other virtual content and the use of other sensor types. We also plan to employ formal processes such as ISO 14598-1 and ISO 9126 (this ISO defines quality metrics for interactive systems), as well as other standards (e.g., ISO 9241-9).

REFERENCES

[1] R. Azuma, "A Survey of Augmented Reality Presence: Teleoperators and Virtual Environments", August, vol. 6, no. 4, 1997, pp. 355–385.

[2] A.Y.C. Neee, (editor). "Augmented Reality – Some Emerging Application Areas". ISBN 978-953-307-422-1, Romana Vukelic Publishing,Croatia, 2011. 266 pages.

[3] A. Klein and G.A. De Assis "A Markeless Augmented Reality Tracking for Enhancing the User Interaction during Virtual Rehabilitation", XV Symposium on Virtual and Augmented Reality (SVR), Cuiaba, Brazil, 2013, pp. 117–124.

[4] Y.Li, "Augmented Reality for remote education", 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), Chengdu, China, V 3, 2010, pp. 187–191.

[5] A. M. Mcnamara, "Enhancing art history education through mobile augmented reality". In: VRCAI '11: Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry. Hong Kong, China, 2011, pp. 507–512.

[6] V. Valjus, S. Jarvinen, and J. Pelota, "Web-based Augmented Reality Video Streaming for Marketing", In: ICMEW '12: Proceedings of the 2012 IEEE International Conference on Multimedia and Expo Workshops. Melbourne, Australia, 2012, pp 331–336.

[7] R. Honken and et al.,"Building a sustainable mobile device strategy to meet the needs of various stakeholder groups", In: SIGUCCS '12: Proceedings of the ACM SIGUCCS 40th annual conference on Special interest group on university and college computing services. Memphis, USA, 2012, pp. 41–48.

[8] A. Dünser and M. Billinghurst, "Evaluating Augmented Realisty Systems", Book Handbook of Augmented Reality, Chapter 13,Springer New York Publishing, 2011, pp. 289–308.

[9] N. N. Kostaras, and M.N. Xenos,"Assessing the usability of augmented reality system". In: 13th Panhellenic Conference on Informatics, Corfu, Greece, 2009, pp. 197–201.

[10] N. M. M Zainuddin,. H. B. Zaman, and A. Ahmad, "Heuristic Evaluation on Augmented Reality Courseware for the Deaf", In 2011 International Conference on User Science and Engineering, Shah Alam, Malaysia, 2011, pp. 183–188.

[11] A. Dunser, R. Grasset, and M. Billinghurst, "A survey of evaluation techniques used in augmented reality studies". Human Interface Technology Laboratory New Zealand, In Proceedings ACM SIGGRAPH ASIA 2008 course. New York, NY, USA, 2008. pp. 1–27.

[12] "ISO 9241-11". Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability, 1998.

[13] J. Nielsen, "Usability Engineering", Chestnut Hill, MA, Academic Press, 1st Edition, Morgan Kaufmann Publishing, 1993, 362 pages.

[14] J. Nielsen, "Designing Web Usability: The Practice of Simplicity", 1st Edition, USA: New Riders Publishing, 1999, 432 pages.

[15] J. Choi, and G. J. Kim, “Usability of one-handed interaction methods for handheld projection-based augmented reality”, Personal and Ubiquitous Computing, Volume 17, Issue 2, 2013, pp.399–409.

[16] M. Hartikainen, E. Salonen, and M. Turunen. “Subjective evaluation of spoken dialogue systems using SER VQUAL method”, In: INTERSPEECH-2004, Jeju Island, Korea, 2004, pp. 2273–2276.

[17] D. Pinelle, N. Wong, and T. Stach, “Heuristic evaluation for games: usability principles for video game design”. Proceeding CHI '08 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2008, pp. 1453–1462.

[18] W. Sawyerr, E. Brown, and M. Hobbs, “Using a Hybrid Method to Evaluate the Usability of a 3D Virtual World User Interface”, International Journal of Information Technology & Computer Science (IJITCS),Volume 8 : Issue No : 2 : Issue on March / April, 2013, pp. 66–74.

[19] J. Nielsen, and R. Lack, “Usability Inspection Methods”, 1st Edition, 1st Edition, John Wiley Publishing, New York, 199, 448 pages.

[20] S. Krug, “Don't Make Me Think: A Common Sense Approach to Web Usability”, 2nd Edition, New Riders: Berkeley, CA, 2006, 216 pages.

[21] D. J. Mayhew,”The Usability Engineering Lifecycle: A Practitioner's Handbook for User Interface Design (Interactive Technologies)”, 1st Edition, Morgan Kaufmann Publishing, 1999, 560 pages.

[22] D. Norman, “The Design of Everyday Things”, New Your: Doubleday, Basic Books Publishing, 2002, 288 pages.

[23] B. Tognazzini, “Tog on interface. First principles of interaction design”, 1992, . Available at : http://www.asktog.com/basics/firstPrinciples.html, Accessed January 6, 2014.

[24] W. Cybis, A. H. Betiol, and R. Faust, “Ergonomia e usabilidade: conhecimentos, métodos e aplicações”. Editora Novatec, 344 pages, 2007,pp. 187–189.

[25] R. O. Lessa, V.R.N. Schuhmacher, M.I Castineira, and A.S. Sousa, “GLIST : checklist automatizado para usabilidade”. Seminário de Informática (SEMINFO), Brazil, November,2006.

[26] E. V. Veenendaal, “Low Cost Usability Testing. Software Quality and Software Testing in Internet Time”, 1999, pp. 153–165.

[27] J.Nielsen, “Why you only need to test with 5 users”. Jakob Nielsen´s Alertbox, March, 2000, Accessed at: November, 10, 2013, Available at: http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/.

[28] W. Hwang, and G. Gavriel Salendy, “Number of People Required for Usability Evaluation: the 10±2 rule”,Communications of the ACM, Vol. 53 No. 5, 2010, pp. 130–133.

[29] M. Schemettow, “Sample size in usability studies”, Magazine Communications of the ACM, Volume 55 Issue 4, April, 2012, pp. 64–70.

View publication stats