Exploitation of Heuristics for Virtual Environments

Ebba Thora Hvannberg
University of Iceland
Dunhaga 5, 107 Reykjavik
Iceland
ebba@hi.is
+354 525 4702

Gyda Halldórsdóttir
University of Iceland
Dunhaga 5, 107 Reykjavik
Iceland
gyda@hi.is

Jan Rudinsky
University of Iceland
Dunhaga 5, 107 Reykjavik
Iceland
janr@hi.is

ABSTRACT
Although generic usability heuristics lists have been popular with researchers and practitioners, emerging new technologies have called for more specific heuristics. One of these heuristics was proposed by Sutcliffe and Gault in 2004 [37]. This paper examines research which has cited these heuristics with the aim to see how it has been exploited. The results showed that a fifth of the papers citing the heuristics have used the heuristics fully or partly, and that researchers have adapted it to their current needs. Following this result we proposed that a patchwork of heuristics might be more useful than a single list. We evaluated a crisis management training simulator using the virtual reality heuristics and discussed how the outcome of the evaluation fitted the patchwork.

Author Keywords
Heuristics evaluation, Virtual Reality, Crisis Management Training

ACM Classification Keywords
H5.2 [**Information interfaces and presentation**]User Interfaces –Evaluation/methodology

INTRODUCTION
Although Nielsen's [24] heuristics have been tremendously popular, with the emerging of technologies and domains, new heuristics have been proposed. Examples are heuristic evaluation (HE) for virtual environments [37], games [29, 30], learning [34], and collaborative work [30, 39]. One motivation for designing specific heuristics is that traditional HCI evaluation methods fail to support the technologies or domains in the narrower area, such as desktop Virtual Reality (VR) [18].

A practitioner may ask how widely used such a list is and how effective it is. Indeed, the latter question has been answered in several research studies, e.g. comparing the effectiveness of a HE to other methods [17]. Conversely, designers of heuristics list can ask themselves how likely it is to be adopted. Previous research by Glass et al. [13] has shown that Computer Science researchers are more likely to invent new methods than to evaluate them, which should warn us not to be overly creative. Example benchmarks, which can predict adoption, can be how effective a heuristic list is compared to other heuristics or evaluation methods, or how easy it is to use.

In preparing for an evaluation of crisis management training in a virtual environment, we pondered this question. Crisis management, as discussed in this paper, is an immediate response to a mass-casualty incident. The response must be well organized in order to protect human lives, minimize personal injury and limit damage to physical assets or environment. The success of crisis management is highly dependent on the training of response and command personnel, e.g. rescue, security, medical and firefighting personnel, whereby they obtain skills in information exchange, decision making, cause analysis and prediction of situation development. Current training methods (e.g. large-scale, multiagency, real-life exercises) for crisis management bring a considerable amount of personnel and organizational complexity that implies significant financial and temporal demands. The latest virtual environments (VE) can provide high fidelity, require less preparation time and resources, and provide an advantage of testing hypothetical situations without a disruption in a community.

Despite the widespread development of VE, there seems to be remarkably few heuristics or design guidelines suggested for their design or evaluation. Our investigation of evaluation methods for virtual environments revealed that the proposed heuristics list by Sutcliffe and Gault [37] was the one that was most cited. Another method by Kalawsky [19] who proposed a special questionnaire to evaluate the usability of a VR has been cited heavily. Furthermore, Gabbard [10] published a detailed VR taxonomy of characteristics of VE. Although they are no doubt useful, the two last mentioned are particularly extensive and not compatible to Nielsen's heuristics [24].

NordiCHI '12, October 14-17, 2012 Copenhagen, Denmark

The aim of this paper is to research how a VR Heuristics proposed by Sutcliffe and Gault [37] (hereafter termed SG2004) has been exploited. We looked at papers referencing the heuristics and researched if they reported using the heuristics and if so how useful they found it. One of the results of this analysis is a proposal of a heuristics patchwork which we present. Thereafter, we describe a heuristics evaluation using SG2004 on a training simulator for crisis management and evaluate the results with respect to the heuristics.

Sutcliffe and Gault proposed [37] 12 heuristics which address usability and presence issues for the evaluation of virtual environment. It is based on Nielsen's [24] heuristics and Sutcliffe's and Kaur's work from 2000 [38]. The twelve heuristics are listed in Table 1. The heuristics on presence are influenced by Witmer & Singer [40] and Slater et al. [33]. A heuristic on task fit was used from Johnson [18] and Gabbard's and Hix's [10] taxonomy influences heuristics on close coordination, realistic feedback, viewpoints and navigation support. Sutcliffe and Gault [37] used the VE heuristics in two cases and found that it was able to uncover serious errors.

To the best of our knowledge, an analysis of the exploitation of VR heuristics has not been performed. There have been survey papers on VR methods [6], and practitioners have conducted surveys to try to find out how usability methods are used [2, 14, 15], but not specifically in VR.

COLLECTION OF REFERENCES AND ANALYSIS METHOD

To find citations to SG2004, the research portals Google scholar (scholar.google.com), Web of knowledge (www.webofknowledge.com) and Scopus (scopus.com)

Nr.	Heuristic
H1	Natural engagement
H2	Compatibility with the user's tasks and domain
H3	Natural expression of action
H4	Close coordination of action and representation
H5	Realistic feedback
H6	Faithful viewpoints
H7	Navigation and orientation support
H8	Clear entry and exit points
H9	Consistent departures
H10	Support for learning
H11	Clear turn-taking
H12	Sense of presence

Table 1. Sutcliffe and Gault heuristics

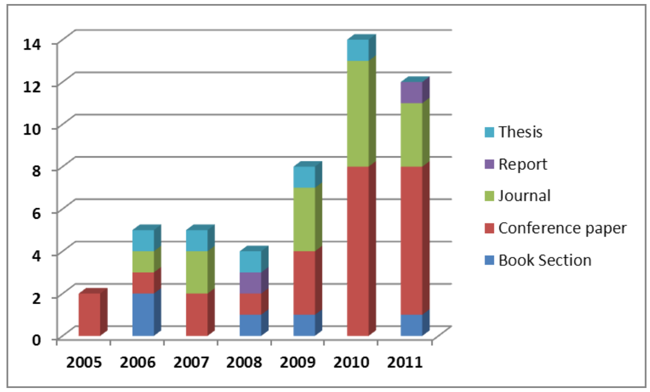

Figure 1. Citations per year and publication type

were searched. After we had verified that SG2004 was cited by these papers, the surrounding text was extracted. Then, each paper was divided into groups according to categories emerging from the data, here the citing papers. Finally, papers which described the use of SG2004 were summarised.

RESULTS

The number of cited papers found by search of the three research portal and verified to cite Sutcliffe and Gault [37] were 56. Two papers were not available to us and four were written in foreign languages (Dutch, French and Chinese), leaving 50 papers for analysis. Figure 1 illustrates the number of papers per publication source shows the citations per year. The total number of journal papers is 14, the number of conference papers is 24, and there are five book sections, five theses and two reports.

Ten papers used to some extent Sutcliffe's and Gault's [37] heuristic method giving occasions for further exploration. Furthermore, seven references discussed SG2004 and 31 used Sutcliffe's and Gault's article for background citations to their literature reviews. The first and the second group are discussed in the next two sections, but the final group was not explored further. A list of those references is available at www.hi.is/~ebba/HvannbergNordiCHI2012Data.xlsx

Papers Using VR Heuristics

This section summarises ten papers which used at least partly the Sutcliffe & Gault heuristics. Bordegoni et al. [5] were interested in testing the usability of a haptic-based system for a virtual manual assembly, in particular how users reached the targeted objectives of effectiveness, efficiency and satisfaction. They decided, in the first phase, to run a preliminary evaluation of the development and setup of the system. Belluco et al. [4] described the evaluation in a more advanced state. An expert user evaluated the system using nine of twelve SG2004 heuristics. The expert performed two tasks and then listed the problems, related them to the heuristics and assigned a rate from 1 (very poor) to 4 (very good). The uncovered problems were classified according to features (interaction, action feedback, haptic feedback and affordances) and ratings of importance. Using the heuristics, Bordegoni et al. [5] managed to identify key usability problems of the application and modify the design of some features, by using one single expert user. The authors planned to involve several evaluators for more formal and comprehensive test of the system. Furthermore, they were planning to use the data about assembly tasks, performed by users for playing back the user's task by a virtual human, with the aim of performing a more accurate ergonomic analysis without real users and using the same data for purposes of training novel users. Admitting that more users would have been preferred, the authors stated that by using this heuristic evaluation method with a single expert user, they got useful information.

Opiyo and Horvath wrote two papers published in 2010 which report on the use of SG2004 for 3D visualisation system for product models. In the first article, discussed here, Opiyo and Horvath [27] applied an expert evaluation of a holographic display (i.e. the HoloVizio 128WD display) to uncover problems in product models visualisation that designers and engineers using holographic displays might encounter. The evaluation method applied in this study followed both Nielsen's [25] recommendations for expert evaluation and Sutcliffe's and Gault's [37] approach for heuristic evaluation of VR applications, with some differences. It was largely the same as Nielsen's expert evaluation method with the main exception of closely matching Sutcliffe's and Gault's [37] heuristics dedicated for evaluation of 3D visualisation technologies. Opiyo and Horvath [27] developed seventeen heuristics, thereof seven closely matched heuristics from Sutcliffe's and Gault's [37]. Independent familiarisation of each evaluator with the holographic display and evaluation was emphasised for ensuring that they would make and independent and an unbiased evaluation. Examples of holographic display specific heuristics are Free-viewing: there should be no need of using 3D visualisation gears such as 3D glasses and head-mounted displays; geometrically volumetric images; and Image quality: The display should generate high-resolution images. The study was conducted six months after the installation of the system, and, hence, the seven experts participating had enough time for familiarising themselves with the display. Altogether, the evaluators together identified 18 problems. Some problems seemed to be rather easy to identify, but others were only identified by one or two of the evaluators, which was concluded to justify the necessity of involving multiple evaluators in expert evaluations. The results showed that barrier free, free-viewing, compatibility and space occupancy criteria were met. Customisation and scale heuristics, on the other hand, appeared to be a nonissue for all expert evaluators. Although this particular holographic display met key benchmarks, its limitations in e.g. image quality, image manipulation, stereoscopic experiences, natural manipulation, interaction and user interface could clearly be seen where expert evaluators uncover the problems. By dividing the problems into two categories, Display related problems and Interactivity problems Opyio and Horvath [27] were able to identify research directions in the area of 3D displays, including display hardware problems.

Opiyo and Horvath [28], in a second journal article, drafted several quality characteristics that could be used as a basis for comparison and selection of features for 3D product visualisation. The criteria presence, perceptual expectations, natural interaction, compatibility and collaboration criteria closely matched the quality characteristics widely used in the evaluation of virtual reality applications by referring to Sutcliffe & Gault [37]; Mengoni & Germani [23], and Witmer & Singer [40]. Of twelve characteristic features which are listed, three closely matched Sutcliffe & Gault's [37] heuristics. Noteworthy is that one of the characteristic features new to this is list is termed collaboration / multiple user support. The article further evolves the classification presented in [27] and describes sets of requirements for a spatial product visualisation system, including fourteen visualisation requirements, nine computational, eight interaction requirements and seven other requirements. Prior to writing these two papers, Opyio and Horvath [26] published a conference paper, also on heuristics evaluation on holographic displays. In that paper, they use the problem rating of Sutcliffe and Gault [37], without using the heuristics themselves.

We now proceed to discuss two theses using SG2004 partly or planning to use them. Ma [21] described a plan of evaluating a collaborative virtual environment for molecular modeling (called AMMP-EXTN) based on an improved cognitive walkthrough, heuristic evaluation and quantitative data collected automatically. The intention was to adopt the form of questionnaire given by Sutcliffe and Gault [37] for VR evaluation covering many aspects of VR specific issues, such as user presence, natural expression of action, navigation support, and collaboration consistency. The evaluation difficulty was considered to lie in the absence of analysis tools for 3D user interface interaction and measurement standards. A comprehensive evaluation, consisting of both a subjective questionnaire and objective user data validating collaborative effectiveness, was expected after the completion of the AMMP-EXTN system. However, this paper did not report on its evaluation and thus could not contribute to our analysis of the usefulness or ease of use of the heuristic.

In his thesis, Seidel [32] created an agent-based 3D VR e-Tourism environment. As a part of the development, he evaluated the acceptance of the system with 20 test participants. Prior to the user evaluations, Seidel conducted an expert evaluation using the SG2004 heuristics and addressed the problems uncovered. Hence, we conclude that it was useful to the system's development, although the results of that evaluation are not described explicitly.

Manesh and Schaefer [22] presented preliminary results of students' experiments by a usability evaluation of VR system using application-related questions based on Sutcliffe and Gault [37] heuristic. Results of selected questions gave pertinent information about the basic-to-advanced level training on the intelligent manufacturing systems.

Galimberti et al. [12] who studied the usability of VR environments for psychotherapeutic applications, such as for Eating Disorder Treatments, presented an ergonomic evaluation and its results. They found their results keen to what Sutcliffe and Gault [37] had discovered, and made a large use of their formulation of Nielsen heuristics [25]. The heuristics used attempted to consider areas of VR not yet explored from the usability point of view, i.e. spatiality (the use of spatial dimensions), virtuality (the sense of telepresence), representation (the construction of meaning) and four indicators: navigability, expected utility, communicative efficacy, and graphical appeal. Furthermore, they say that it is high time and conditions are right to develop a psychosocial model of usability for VR.

A second paper by Galimberti et al. [11] illustrated the development of psychological model for analysing the usability and ergonomics of virtual environments, which is used as a support in cognitive behavioral psychotherapy. Referring to Sutcliffe and Gault's [37] they said that, although field studies by VR designers demonstrated the need for HCI knowledge and methods, only few evaluation methods have been proposed for assessing the usability of VEs criteria. Generally speaking, most studies reviewed by them have followed observation and expert interpretation of users' errors or experimental studies reporting performance data and problems in a range of VR technology. The authors cite Sutcliffe and Gault [37] saying that the investigation of the real context of use was the fundamental step that can effectively contribute to the optimisation of the whole design processes together with the effective integration of existing methods and the improvement of usability evaluation tools that are still too vague for VR applications.

Papers Discussing VR Heuristics

Seven papers summarized in this section discuss Sutcliffe's and Gault's [37] work, but do not use their heuristics for evaluation.

Flavian et al. [9] developed an e-store at a website and studied key factors regarding web design influencing online success. The authors observe that the original Nielsen's heuristics is not applicable for web design and use SG2004 heuristics as an example of a heuristic list which has been developed for a new area. Having noted that an existing heuristic does not meet their demands, Flavian et al. developed a heuristic test according by using the process of Sutcliffe's and Gault's [37] [36]. The authors mentioned that attracting users' attention, with aesthetic elements playing a pivotal role, was the main goal of such a cost-effective method [37].

Hornbæk [16] described seven dogmas in recent works on Usability Evaluation Methods (UEMs) and criticised them for low validity and limited reliability. The aim of the study was to synthesise the dogmas and go further than existing reviews in suggesting how to move beyond. While discussing evaluators satisfaction with individual heuristic sets, Hornbæk cited Sutcliffe and Gault [37] who asked evaluators to assess the applicability of 12 heuristics for evaluation of virtual reality applications on a scale from 1 to 7 and suggested that as all ad hoc satisfaction measures, such data hardly could stand by themselves but might capture orthogonal dimensions to the effectiveness of a UEM.

Khan et al. [20] discuss usability evaluation for haptic systems, noticing that SG2004 addresses haptics partially. They performed a literature survey to find out the patterns related to different evaluation methods and haptic devices used in various domains. The authors' suggestions from the study findings were that the development of generic evaluation methods, applicable to different haptic systems in various domains, was needed. The work is reported to be in progress, but the paper proposes a framework for the usability of haptic systems.

Robinson et al. [31] describe the design and evaluation of an immersive design system for cable harness design. A practical user evaluation with ten participants was made. The goal was to identify in general terms the relative distribution of user activity between specific purposes during practical system operation. The findings of the evaluation showed that Navigation accounted for 41%, Design 27%, System Operation 23% and looking at Task Instructions 9% of all user activities. The authors mentioned that other techniques such as Kalawsky's [19] usability questionnaire, Witmer and Singer's [40] presence questionnaire or Sutcliffe and Gault's [37] heuristic approach were not used because many of the generalised questions were not relevant and the approaches did not achieve the detail required for the testing of specific system features.

Dang et al. [7] presented an initial framework for studying issues in the design and evaluation of Collaborative Virtual Environments (CVE) based on immersive projection technology systems. The objective was to gain insights on both technological aspects regarding the development and human factors on collaborative work within VR, besides the interrelation between technological and human factors aspects for building usable CVE. The authors pointed out that existing heuristics [37] and ergonomic criteria [1] for design and evaluation of the VE interface would be useful for the design and evaluation of individual mode in CVE.

Barrett and Blackledge [3] pointed out Sutcliffe's and Gault's [37] suggestions regarding motivation as a significant factor that influences learning and thus better-motivated if users can learn more effectively. The results of their study agreed with this suggestion in that they found that motivation was an influential psychological factor that is positively related to the VR measurement outcomes.

Dunser and Billinghurst [8] give an overview of usability evaluation of augmented reality (AR) systems. They discuss the SG2004 heuristics, saying that some of the twelve heuristics apply to AR systems with slight adaptations, but others do not. The ones applicable are natural engagement, compatibility with the user's task and domain, realistic feedback, support for learning, clear turn-taking and consistent departures. Those which are thought to be less applicable for AR are Natural expression of action, Close coordination of action and representation, faithful viewpoints, navigation and orientation support, clear entry and exit points, and sense of presence. What these heuristics have in common is that they deal with users' representation or navigation in the VR.

DISCUSSIONS ON LITERATURE REVIEW

From the analysis of the papers, we see that they represent many different application domains, such as manufacturing, psychology treatment and tourism. The way of working, e.g. collaboration, is fairly apparent. Different technologies, such as 3D displays and haptic peripherals can be seen.

Categorizing papers according to how much they used the SG2004 heuristics, we found that only one paper used them fully, eight papers used them partly, and one planned to use them.

Of the papers, we described in the previous sections, Robinson et al. [31] rejected the use of the heuristics as a questionnaire, claiming that they were too generic for their application. This was contradicted by Khan et al. [20] who suggested that a generic evaluation method needed to be designed for haptic systems in various domains.

Our results confirm the conclusion of Woolrych et al. [42] who say that usability evaluation methods are collections of resources, which practitioners adapt to their specific project contexts. Woolrych et al. [41] identified seven categories of distributed cognitive resource (DCR), so termed because they were distributed across the context of an evaluation. The categories are knowledge on: Users, Tasks, Application domain, Product/System, Implementation platform, Users' interaction with computers, and Design options and their consequences. By looking at citations to Sutcliffe and Gault, one realizes that many of the papers use SG2004 partly, and, hence, suggesting that HCI specialists pick, choose and adapt what is applicable to them.

We recognize that in the same way as DCRs are proposed, not a single heuristics list may be useful to help evaluators uncover problems, but a patchwork of heuristics. Influenced by the kinds of specialized heuristic lists that have been derived in the past and the characteristics of the crisis management training simulator, we propose that the colors in that patchwork may be industry sectors, application domain, social organization, interaction technology platform, or they may be others according to the genre of the system. Figure 2 shows an example of such a framework and Figure 3 shows an instantiation for heuristics evaluation of a training simulator for crisis management for the transport sector, which is set in a virtual environment and uses the concept of serious games. How the patchwork is sewed, depends partly on which factors designers want to emphasize.

Acknowledging that the number of citations may not be the only metric to predict the popularity of a method, it certainly is an indication. It should be further noted, that some use Nielsen's heuristics for the evaluation of VR, such as [35], but it is not the topic of this paper to explore to what extent. However, it may be an attractive topic for further work.

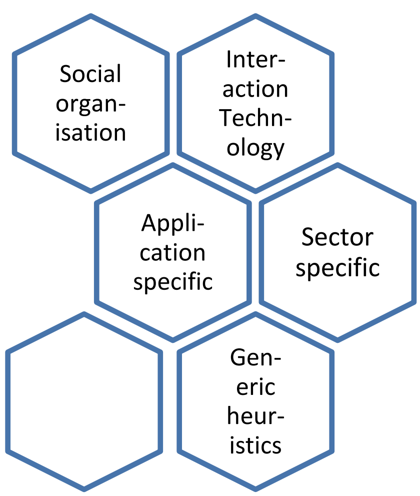

Figure 2. Patchwork of heuristics

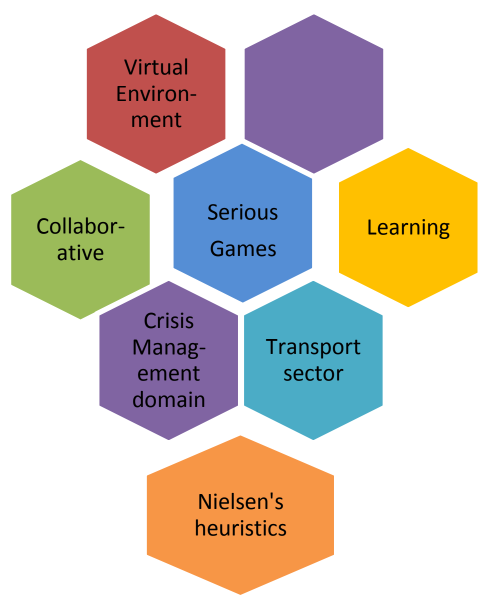

Figure 3. Patchwork of HE for Crisis Management Training

HEURISTICS EVALUATION OF A TRAINING SIMULATOR

Heuristics Evaluation Protocol

A first version of a crisis management training simulator developed in the CRISIS project, http://idc.mdx.ac.uk/projects/crisis/, (a video accompanies this paper) was used in a heuristics evaluation using SG2004. The evaluation took place in a laboratory. Ten participants were recruited, six students, all with basic knowledge on HCI evaluation and VR, and four experts with more knowledge on HCI evaluation, VR and extensive knowledge on the application domain, crisis management training. None of the participants was a trainee of crisis management training. Participants played one of two roles, a medical coordinator or a fire commander, assigned by the evaluation conductor. These roles had defined five and four tasks respectively, which participants were asked to carry out. The tasks are listed in Table 2.

Although the training simulator is multi-player, participants evaluated the system individually, without co-players. Prior to the evaluation, participants were asked to familiarize themselves with the evaluation protocol, which contained a description of the heuristics, conformance questions, added by us, and problem report forms. Evaluators were asked to report on Brief description, Likely difficulties, Specific context, Assumed Causes, Severity (3 different ratings), and Relevant heuristics. The heuristics were presented to them both in an overview list, much the same as Sutcliffe and Gault did, but also in a fuller version, listing a Conformance Question, Evidence of Conformance and Motivation, written by us. We did this to ensure that evaluators understood the heuristics. An example description for H1 is shown in Table 3.

Medical commander	**Fire commander**
T1) Go to the scene – Commander arrives to RVP (Rendezvous Point) and walks from here to the scene.	T1) Go to the scene – Commander arrives to RVP (Rendezvous Point) and walks from here to the scene.
T3) Primary triage – Commander performs primary triage on the casualties himself (commander can use a triage diagram) or s/he can task an assistant to do primary triage at the scene.	T2.1) First assessment – Commander makes first assessment of the situation on the scene (what can be seen or heard from a safe distance). Results are written on a piece of paper (provided by the evaluator).
T5) Select location for casualty placement – Commander selects a location for the placement of discovered casualties at Casualty Assembly Point (CAP). Commander picks up a flag from an ambulance vehicle to mark each area (green, yellow, red and black) by placing the appropriate flag.	T2.2) Commander explores the scene, discovering as many casualties as there are and logs them.
T7) Casualty transport – Commander asks ambulance vehicle to move the casualty from the scene to CAP.	T3) Primary triage – Commander performs primary triage on the casualties himself (commander can use a triage diagram) or s/he can task an assistant to do primary triage at the scene.
T8) Secondary triage – Commander performs secondary triage on casualties at yellow CAP (commander can use Glasgow Coma Score diagram).	

Table 2. Tasks of two roles

Heuristic name	H1: Natural engagement
Conformance question	Was the navigation around the scene and the manipulation with objects and their response similar to the real world?
Evidence of Conformance	Interaction in the virtual world resembles the interaction in the real world. The user also feels immersed.
Motivation	Interaction should approach the user's expectation of interaction in the real world as far as possible. Ideally, the user should be unaware that the reality is virtual.

Table 3. A detailed description of an example heuristic

Heuristics Evaluation Results

The evaluation resulted in 69 problem instances. After consolidating the problem instances, we identified 25 individual problems. On average, a participant uncovered AVG=5.5 problems, SD=3.3, Min=2, and Max=13, N=10. Figure 4 shows how effective individual participants were in uncovering problems and Figure 5 shows the frequency of the problems, i.e. how many participants uncovered a particular problem.

Average frequency of problems (N=25) was 2.8, STD=2.3, MIN=1, MAX=10. 28 problem instances were rated by the participants as severe, 26 as moderate and 13 as minor.

Some of the problem instances have more than one heuristics listed, up to 4 heuristics per problem instance. However, most frequently, the problem instances only had one heuristics associated with them. The average number of heuristics listed per problem instance was AVG=1.4, STD=0.78, Max=4 and Min=0, N=69.

Figure 6 shows the distribution of heuristic instances mentioned by participants. The total number of such instances is 96 since some of the 69 problem instances have more than one heuristic associated with them. Participants used H1: Natural engagement most frequently. The problem

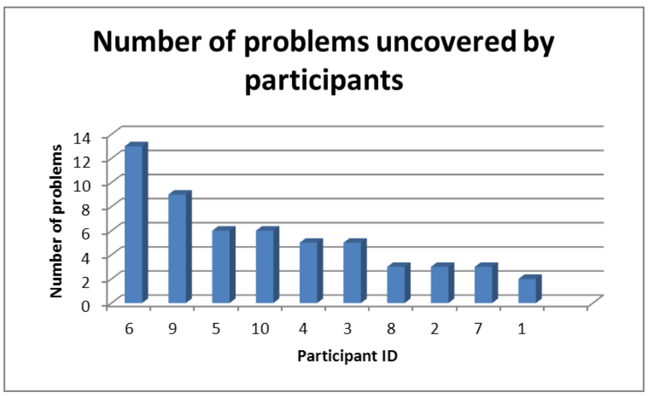

Figure 4. Number of problems uncovered by participants

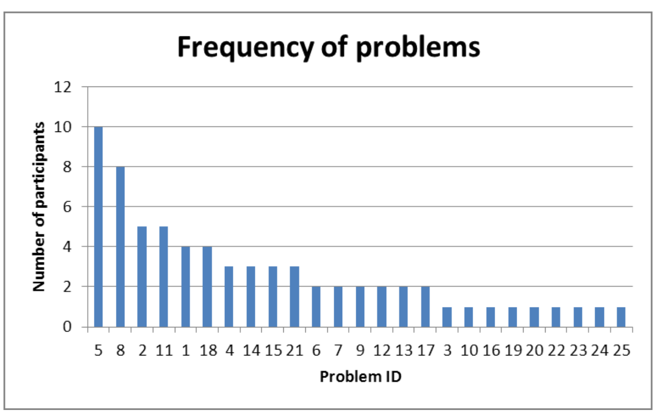

Figure 5. Frequency of problems

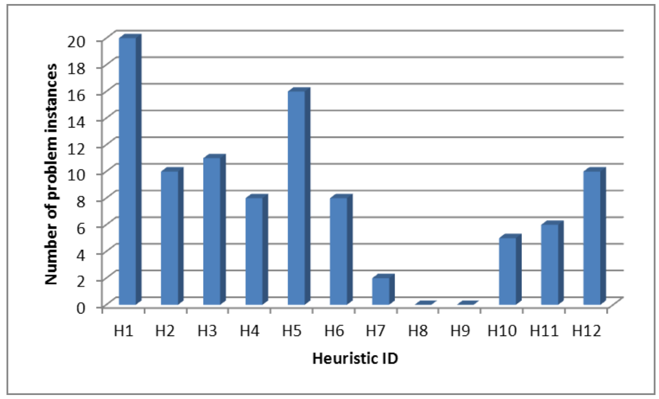

Figure 6. Heuristics instances

instances covered all twelve heuristics except two, H8: Clear exit and entry points, and H9: Consistent departures.

Comparing these results to the findings of Sutcliffe and Gault [37] on evaluators' view on the applicability of the heuristics to VR measured as Net Positive Values (NPV) (data from seven evaluators, the scale was -3, not useful and +3, very useful for interpreting usability), we see that there is some resemblance. The correlation is 0.65, significant at the 0.05 level, and the data is shown in Figure 7.

Preliminary Findings of the HE Results with Respect to a Patchwork

This section discusses to what extent the HE covers the different aspects of existing problems of the crisis training simulator. If we assume that SG2004 covers VR and presence related problems, they should have been uncovered. However, recalling the patchwork (see Figure 3), there may be problems of other aspects that have not been uncovered, since evaluators were not prompted to look for them, or they did not categorize the uncovered problems to those categories. Below, we will systematically analyze the aspects (patches):

Generic heuristics (Nielsen): The SG2004 heuristics H7-H11 map directly to Nielsen. We have not examined if any of the problems could be attributed to Nielsen's other five categories.

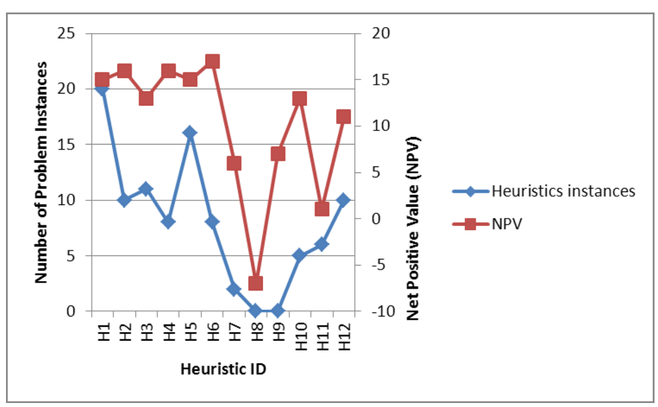

Figure 7. Comparison of heuristics used and the assessed utility of heuristics

Collaboration: One problem instance mentioned that the player was alone in the world, except for the assistant, which is a non-playing character (NPC). Several problem instances were on collaboration and interaction with the NPC. Heuristic *H11: Clear turn taking* can be classified into this category. It should be noted that the tasks were not tuned to collaboration.

Training: Quite a few problems were uncovered related to training. They have been classified into *H10: Support for learning*, which is strictly speaking not correct, since the description of that category is for learning to use the VR, but none on training per se.

Crisis management: A number of problems fall into this category, and are mostly classified into *H2: Compatibility with user's task and domain.*

Transport sector: No problem could be directly attributed to this aspect. A reference to the transport sector can be seen when the objects, such as ducking under the plane is mentioned when reporting a problem.

Serious games: No problem could be attributed to this aspect.

The above analysis shows that a number of problems could be attributed to the training aspect. This preliminary review of the different aspects shows that problems, for the most part, could not be attributed to the aspects. It is difficult to conclude whether this was because there were no such problems, or if they evaluators did not uncover them because they were not probed to do so. Also, the design of the tasks, obviously, will influence the type of problems. In particular, one should note that collaborative tasks in a multi-player setting were not included. Nonetheless, we conclude that the results can justify further research on the patchwork concept.

A caveat of such a patchwork is that a list of heuristics may become too long. One of the advantages of the current heuristics and probably one of the reasons for their popularity is that they usually contain only ten or twelve items.

CONCLUSIONS AND FUTURE WORK

This paper has conducted an extensive analysis on the exploitation of the heuristics by Sutcliffe and Gault [37] on virtual reality and presence. The conclusion is that the heuristics are being cited increasingly, but they have only been used, either fully or partly by 10 out of 50 citations. What is more noteworthy, are the lessons we learned from the analysis, namely, that, in most cases, researchers adopt and adapt the heuristics to suit their own needs. It should be mentioned that researchers do not seem to agree on how generic or specific such heuristics should be. This has prompted us to present a patchwork of heuristics aspects, which can be sewed for a particular system, according to its most relevant genres. This paper has proposed several genres which could make up such a patchwork, but future work can include a formal framework on which patches to select for a particular system. Such research could include a thorough analysis of current heuristics accompanied by a process for instantiating the patchwork for a particular system. The final contribution of this paper is the evaluation of the validity of the patchwork approach. For that purpose, we conducted a heuristics evaluation on a crisis management training simulator. The results showed that problems were attributed to ten of the twelve heuristics. A preliminary analysis showed that while most of the problems fitted quite well into the SG2004 heuristics, a number of problem instances on training were uncovered and fitted not well into the SG2004 heuristics. This warrants further research on the patchwork approach, including an evaluation of the effectiveness of a patchwork guided heuristics list.

ACKNOWLEDGMENTS

We gratefully acknowledge that the research reported in this paper was carried out as part of the FP7 project, CRISIS, funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. [FP7-242474].

REFERENCES

1. Bach, C., Scapin, D. Ergonomic criteria adapted to human virtual environment interaction. In: Baudel, T. (ed.): *IHM 2003. The 15th French-Speaking Conference on Human-Computer interaction on 15eme Conference Francophone Sur L'interaction Homme-Machine.*, Vol. 51. ACM: New York, NY (2003) 24-31
2. Bak, J.O., Nguyen, K., Risgaard, P., Stage, J. Obstacles to usability evaluation in practice: a survey of software development organizations. *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*. ACM, Lund, Sweden (2008)
3. Barrett, M., Blackledge, J. Evaluation of a Prototype Desktop Virtual Reality Model developed to Enhance Electrical Safety and Design in the Built Environment *ISAST Transactions on Computers and Intelligent Systems* 3, 3 (2011), 1-10
4. Belluco, P., Bordegoni, M., Polistina, S. Multimodal navigation for a haptic-based virtual assembly application. *ASME 2010 World Conference on Innovative Virtual Reality, WINVR 2010* (2010) 295-301
5. Bordegoni, M., Cugini, U., Belluco, P., Aliverti, M. Evaluation of a Haptic-Based Interaction System for Virtual Manual Assembly Virtual and Mixed Reality. In: Shumaker, R. (ed.), Vol. 5622. Springer Berlin / Heidelberg (2009), 303-312
6. Bowman, D.A., Gabbard, J.L., Hix, D. A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence: Teleoperators & Virtual Environments* 4, 11 (2002), 404-424

7. Dang, N.-T., Chatelain, C., Pergandi, J.-M., Mestre, D. A Framework for Design and Evaluation of Collaborative Virtual Environments. (2008)
8. Dunser, A., Billinghurst, M. Evaluating Augmented Reality Systems. In: Furht, B. (ed.): *Handbook of Augmented Reality*. Springer, New York (2011), 289-308
9. Flavian, C., Gurrea, R., Orus, C. A heuristic evaluation of websites design for achieving the web success. *International Journal of Services and Standards* 1, 5 (2009), 17-41
10. Gabbard, J.L. A Taxonomy of Usability Characteristics in Virtual Environments. Vol. Master of Science in Computer Science and Applications. Virginia Polytecnic Institute and State University, Blacksburg, Virginia (1997) 195
11. Galimberti, C., Belloni, G., Cantamesse, M., Cattaneo, A., Gatti, F., Grassi, M., Menti, L. The development of an integrated psychosocial approach to effective usability of 3D Virtual Environments for Cybertherapy. *PsychNology Journal. The Other Side of Technology* 2, 4 (2006), 161-180
12. Galimberti, C., Belloni, G., Cattaneo, A., Grassi, M., Manias, V., Menti, L. An Integrated Approach to the Ergonomic Analysis of VR in Psychotherapy: Panic Disorders, Agoraphobia and Eating Disorders. In: Riva, G., Botella, C., Légeron, P., Optale, G. (eds.): *Cybertherapy. Internet and Virtual Reality as Assessment and Rehabilitation Tools for Clinical Psychology and Neuroscience*. IOS Press,© 2004, 2005, 2006, Amsterdam, (2006)
13. Glass, R.L., Ramesh, V., Vessey, I. An analysis of research in computing disciplines. *Commun. ACM* 6, 47 (2004), 89-94
14. Gulliksen, J., Boivie, I., Göransson, B. Usability professionals" current practices and future development. *Interacting with Computers* 4, 18 (2006), 568-600
15. Gulliksen, J., Boivie, I., Persson, J., Hektor, A., Herulf, L. Making a difference: a survey of the usability profession in Sweden. *Proceedings of the third Nordic conference on Human-computer interaction*. ACM, Tampere, Finland (2004)
16. Hornbæk, K. Dogmas in the assessment of usability evaluation methods. *Behaviour & Information Technology* 1, 29 (2009), 97-111
17. Hvannberg, E.T., Law, E.L.-C., Larusdottir, M.K. Heuristic evaluation: Comparing ways of finding and reporting usability problems. *Interacting with Computers* 2, 19 (2007), 225-240
18. Johnson, C. Why `traditional' HCI techniques fail to support desktop VR. *The 3D Interface for the Information Worker (Digest No. 1998/437), IEE Colloquium on* (1998) 1/1-1/5
19. Kalawsky, R.S. VRUSE—a computerised diagnostic tool: for usability evaluation of virtual/synthetic environment systems. *Applied Ergonomics* 1, 30 (1999), 11-25
20. Khan, M., Sulaiman, S., Said, A.M., Tahir, M. Development of usability evaluation framework for haptic systems. *VR Innovation (ISVRI), 2011 IEEE International Symposium on* (2011) 343-344
21. Ma, W. AMMP-EXTN: A User Privacy and Collaboration Control Framework for a Multi-User Collaboratory Virtual Reality System. *Department of Computer Science*, Vol. Computer Science Theses. Georgia State University (2007)
22. Manesh, H.F., Schaefer, D. A Virtual Factory Approach for Design and Implementation of Agile Manufacturing Systems. *2010 ASEE Annual Conference &* Vol. AC 2010-24. American Society for Engineering Education (2010)
23. Mengoni, M., Germani, M. Integration of virtual reality technologies in industrial design processes: a structured approach. *Proc. of Virtual Concept*, Playa, Mexico (2006) 1-12
24. Nielsen, J. Enhancing the explanatory power of usability heuristics. *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*. ACM, Boston, Massachusetts, United States (1994)
25. Nielsen, J. Heuristic evaluation. In: Nielsen, J., Mack, R.L. (eds.): *Usability Inspection Methods*. Wiley & Sons New York City, NY (1994)
26. Opiyo, E.Z., Horvath, I. Using hybrid heuristic evaluation method to uncover the conceptual design tasks supported by a holographic display based truly 3d virtual design environment. *2008 Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, DETC 2008*, Vol. 3 (2009) 1649-1659
27. Opiyo, E.Z., Horvath, I. Exploring the viability of holographic displays for product visualisation. *Journal of Design Research* 3, 8 (2010), 169-188
28. Opiyo, E.Z., Horvath, I. Towards an interactive spatial product visualisation: a comparative analysis of prevailing 3D visualisation paradigms. *International Journal of Product Development* 1, 11 (2010), 6-24
29. Pinelle, D., Wong, N., Stach, T. Heuristic Evaluation for Games: Usability Principles for Video Game Design. *CHI '08, CHI Conference on Human Factors in Computing Systems* ACM New York, NY, USA, Florence, Italy (2008) 1453-1462
30. Pinelle, D., Wong, N., Stach, T., Gutwin, C. Usability heuristics for networked multiplayer games. *Proceedings of the ACM 2009 international conference on Supporting group work*. ACM, Sanibel Island, Florida, USA (2009)
31. Robinson, G., Ritchie, J.M., Day, P.N., Dewar, R.G. System design and user evaluation of Co-Star: An immersive stereoscopic system for cable harness design. *Computer-Aided Design* 4, 39 (2007), 245-257
32. Seidel, I. Engineering 3D Virtual World Applications. Design, Realization and Evaluation of a 3D e-Tourism

Environment. Vol. Doctors der technischen Wissenschaften. Vienna University of Technology (2010) 196

33. Slater, M., Linakis, V., Usoh, M., Kooper, R. Immersion, presence and performance in virtual environments: an experiment using tri-dimensional chess *ACM Virtual Reality Systems and Technology (VRST '96)*,M. Green ed., Hong Kong, July l-4. (1996)
34. Squires, D., Preece, J. Predicting quality in educational software:: Evaluating for learning, usability and the synergy between them. *Interacting with Computers* 5, 11 (1999), 467-483
35. Stromberg, H., Vaatanen, A., Raty, V.-P. A group game played in interactive virtual space: design and evaluation. *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*. ACM, London, England (2002)
36. Sutcliffe, A. Assessing the Reliability of Heuristic Evaluation for Website Attractiveness and Usability. *Proceedings of the Annual Hawaii International Conference on System Sciences* Conf 35 (2002), 137
37. Sutcliffe, A., Gault, B. Heuristic evaluation of virtual reality applications. *Interacting with Computers* 4, 16 (2004), 831-849
38. Sutcliffe, A.G., Kaur, K.D. Evaluating the usability of virtual reality user interfaces. *Behaviour & Information Technology* 6, 19 (2000), 415-426
39. Tromp, J.G., Steed, A., Wilson, J.R. Systematic Usability Evaluation and Design Issues for Collaborative Virtual Environments. *Presence: Teleoperators & Virtual Environments* 3, 12 (2003), 241-267
40. Witmer, B.G., Singer, M.J. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence* 3, 7 (1998), 225-240
41. Woolrych, A., Cockton, G., Hindmarch, M. Knowledge Resource in Usability Inspection. *HCI 2005, The 19th British HCI Group Annual Conference*. BCS (2005)
42. Woolrych, A., Hornbaek, K., Frokjaer, E., Cockton, G. Ingredients and Meals Rather Than Recipes: A Proposal for Research That Does Not Treat Usability Evaluation Methods as Indivisible Wholes. *International Journal of Human-Computer Interaction* 10, 27 (2011), 940-970