Evaluation of Virtual Reality for Usability Studies in Occupational Safety and Health

Peter Nickel, Andy Lungfiel, Georg Nischalke-Fehn, Preethy Pappachan, Michael Huelke, Michael Schaefer
Institute for Occupational Safety and Health of the German Social Accident Insurance (IFA), Department "Accident Prevention and Product Safety", Alte Heerstr. 111, D-53757 Sankt Augustin, FRGermany,
Tel. +49 (0)2241 2312832, Fax. +49 (0)2241 2312234, E-mail peter.nickel@dguv.de, http://www.dguv.de/ifa

Virtual Reality, Usability, Evaluation, Human-Robot-Interaction, Human Performance

ABSTRACT

Virtual reality (VR) has grown into a simulation tool for humans to interact with dynamic virtual environments and into a methodology for applied research in human-machine system design and evaluation. Applications of VR nowadays often concern human-machine interface design and evaluation. The VR of the IFA is already in use for e.g. usability investigations of safety devices. In the present study, however, the VR system itself is evaluated by addressing usability in terms of ergonomic basics and of human information processing for appropriate human-robot interaction. An application of a checklist and heuristic evaluations on ergonomics resulted in continuous improvements of the VR system. In a pilot experimental study the effects of intensity of human-robot interaction on human information processing has been investigated by means of task performance measures, psychophysiology and questionnaires. Also included were tests for simulator sickness and the level of immersion and presence under human-robot interaction in VR. The results of the pilot study suggest a high quality of the VR system with simulator sickness not being an issue and immersion and presence experienced at medium to high levels. More intensive human-robot interaction in VR increased the experience of presence, accompanied by shifts in human performance indicating an increase in human information processing demands. Although the empirical basis is too small to draw general conclusions, results are promising for future VR evaluation studies that will include comparisons of human-robot interaction in VR with reality. Applied usability research with VR on accident prevention and product safety as already initiated by accident insurance institutions can therefore be performed under fortunate circumstances.

1 INTRODUCTION

Virtual reality (VR) has grown into a simulation tool to enable humans to interact with dynamic representations of real or imaginary worlds and into a methodology for applied research in human-machine system design and evaluation. Applications in different industry and services sectors such as manufacturing, process industries, transportation or medical engineering suggest suitability of VR in general. VR has also been matured to facilitate efficient investigations with the potential to substitute more traditional research at shop floor level while being careful with human, material and financial resources /12/. Improvements in VR technology with appropriate adjustments for human information processing in human-machine interaction made VR also attractive for training and product design and evaluation in occupational safety and health /2/7/.

Substantial research efforts in accident prevention and product safety using VR could already highlight several advantages but also demonstrate some limitations of relevance also for the occupational safety and health domain /8/7/. Disadvantages of VR are often seen in limited sensational or feedback support (e.g. tactile interaction) closely linked to information processing requirements for human-machine interaction. Advantageous is the use of VR across the product life cycle. VR allows investigating the usability of products not yet available, because they are still under development. Already in this stage empirical testing and comparison of design alternatives is recommended. VR also enables investigation of products no more available, e.g. relevant when accidents destroyed some functionality but cause-effect analysis is required. Products already available may be of concern in VR when safety devices require improvements or were avoided /11/. Thereby, VR more comprehensively covers product design and has the potential to also widen the scope for usability investigations. Product usability concerns the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (/4/ p. 2). Usability aims at work system design /3/ in that interaction between humans and system components (e.g. work environment, place, equipment) should serve appropriate task performance and consider ergonomic design strategies and principles. While usability will be of relevance for the evaluation of product design in VR applications, it is also important to investigate the usability of the VR system itself to find out about whether and to what extent it will be possible to generalise VR results for the real world.

Usability of a VR system is concerns simulator sickness, e.g. caused by perceptual mismatches between VR and reality such as optical distortion, flicker or position-tracking errors /13/. Furthermore, human behaviour can only be realistic and of relevance for real world applications if an operator immerses into VR and mentally experiences her/his presence in VR /15/. Before a VR system is usable for research into the design of human-machine interfaces it is necessary to find out about whether working in VR appropriately affects human performance and whether human performance measures are suitable to tap human information processing demands in VR. The preliminary and formative evaluation study therefore aims at assessments of ergonomic basics and a pilot study to prepare for comprehensive usability evaluations of the VR system.

2 METHODS

2.1 Participants

The number of participants in the pilot experimental study was limited to four (mean age = 41.5, SD = 9.2, all male), because this preliminary study aimed at gaining experience with the quality of VR system design and with human performance assessment in human-robot interaction. All participants were right-handed and all reported normal or corrected-to-normal vision.

2.2 Testing Environment and Tasks

The study was scheduled in the VR laboratory of the IFA /6/ equipped with a 7 m² operating space in front of a curved presentation wall of 24 m² (3 x 8 m). The wall represents a 164° circle segment of 2.8 m arc radius. The physical dimensions allow the workspace and projection area big enough to fully cover the human field of vision for stereoscopic depth perception when facing the projection wall. Dimensioning facilitates interaction of human operators with large stationary machinery. Interference filter technology (Infitec® GmbH, Germany) was applied to 3D rear projection with three pairs of high luminance beamers and to operator glasses for depth cue perception in 3D space. Four infrared cameras for motion capturing (VICON®, OMG plc, UK) mounted on top of the projection served the match between operator head movement and dynamic adaptations for VR visualisation.

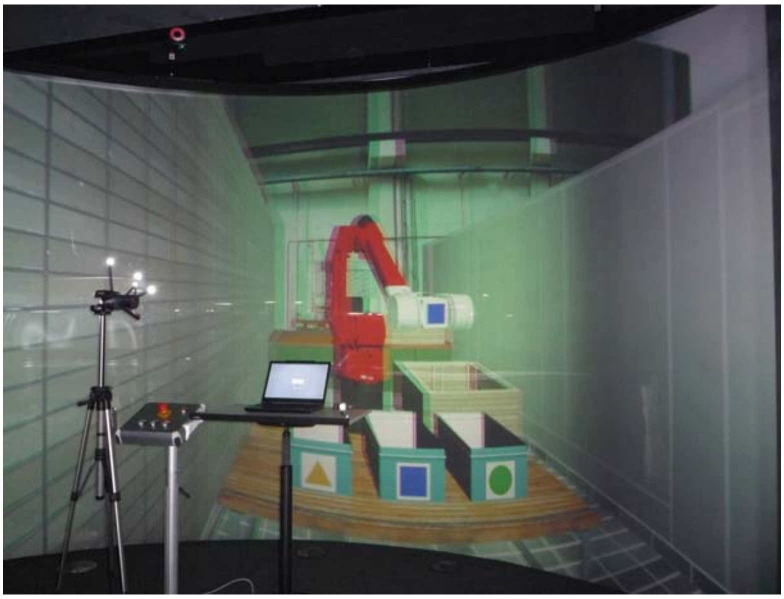

Figure 1. *VR work place for human-robot interaction (observers' point of view without 3D glasses).*

A workplace for human-robot interaction was applied as VR scenario (Fig. 1), similar to a workplace in another laboratory of the IFA. The operator was facing the robot and had direct control for robot movements by a purpose-built panel placed outside robot reach and left to a high desk. On the high desk the operator performed additional tasks using paper and pencil or a notebook. The scenario used augmented reality, because for the operator physical and digital objects co-existed in real-time and the human-robot interaction was merged by real and virtual worlds; e.g. robot control panel and dynamic projection of the robot workplace.

The operator primary task was a human-robot interaction task (RIT). It required the operator to work within the dynamic augmented environment and directly interact with the robot in VR, if necessary. In case of direct interaction the robot took packages with different labels from a buffer linked to an invisible a conveyor belt and the operator was asked to assign packages to boxes located in front of the operator within limited time intervals. Operator decision making demands were represented by the number of correct assignments and were taken as a measure of primary task performance. As a secondary task the operator performed the Grammatical Reasoning Task (GRT /1/) in concurrence with the primary task (RIT). The GRT was presented on the notebook at the high desk. This task imposed high cognitive demands on working memory processes and other information processing resources /10/. It was used in the pilot study to maintain operator cognitive demands at a high level and to represent cognitive processes such as sustained attention and reasoning with resemblance to processes in construction work. Task performance in the secondary task referred to speed and accuracy of operator responses. Measures taken from primary and secondary tasks were recorded during task performance for off-line analysis. The measures served an assessment of effectiveness and efficiency of task performance.

In addition to operator task performance psychophysiological parameters were taken into account for assessments of efficiency of task performance. Inter-beat intervals of the electrocardiogram were recorded with Polar® RS800 (Polar, Finland) and saved to file for off-line analysis. Artefact correction was done by visual inspection of detected inter-beat intervals and manually corrected, if necessary. Heart rate variability measures were calculated using procedures of a LabVIEW™ (National Instruments, USA) virtual instrument as described in /10/. Measures were used as a marker of effort investment during task performance.

2.3 Experimental Design and Procedure

The sessions of the pilot experimental study lasted about 45 minutes each, subdivided into consecutive sections for preparation, experimentation and debriefing. During the experimental section operators performed primary and secondary tasks in VR while wearing Infitec® glasses for 3D perception. The intensity of human-robot interaction differed between three periods within the experimental section. Throughout all three periods the operator performed the GRT in the virtual human-robot environment. More intensive and direct human-robot interaction was required only in the middle period, when the operator was asked to additionally perform the RIT. Interaction periods were interrupted only by filling in questionnaires for task load and presence assessments.

An experimental session started with participants fill in a demographic questionnaire and a check for detrimental effects on psychophysiological measures. The Polar® device was attached for continuous recording of inter-beat intervals and participants were presented questionnaires for initial operator state assessment. Baselines were taken for the Simulator Sickness Questionnaire (SSQ /13/), the Presence Self-Assessment Manikin (PSAM /14/), the Immersive Tendency Questionnaire (ITQ /15/) and the NASA Task Load indeX (TLX /5/). Participants were instructed for the GRT and the RIT before they started the experimental section. In the debriefing section the SSQ, PSAM, TLX and the Presence Questionnaire (PQ /15/) were presented for final operator state assessment.

Repeated measures analyses of variance (ANOVA; GLM: PASW Statistics Inc., USA) referred to pre-processed data (Office Excel™, Mircrosoft Corp., USA) integrated across intervals for the three consecutive experimental periods. Linear and quadratic contrast analyses were applied for post-hoc analysis and reported if significant.

3 RESULTS

3.1 VR Ergonomics

The growing numbers of applications of dynamic virtual and augmented environments for work systems design warrant ergonomic analysis, particularly as these environments become a common tool in occupational safety and health. Basic factors to be considered in ergonomic VR system design are compiled in an ergonomics checklist /9/, referring to potential ergonomic concerns such as requirements for operator posture, space and grasp or for the prevention of hazardous sensors or obstacles. In addition heuristic evaluations on ergonomics have been used for an assessment of the VR system including the set up for the pilot study. The screening instrument could not highlight noticeable problems.

An auditory representation in the VR system by means of a sound system is currently under construction. Heuristic evaluations during the development of the VR laboratory identified echo effects alongside the large and curved presentation wall. This effect could already be significantly reduced by installation of sound absorption on the ceiling above the operating space. The VR laboratory is built up without windows because exposure to direct sunlight or airborne dust would negatively affect projection qualities. Climatic conditions in the VR laboratory were therefore given special emphasis by installation of an air condition. As regards illumination, slightly similar luminance contrasts between different points of measurement were found when comparing measures taken within the VR scenario with those taken within the realistic human-robot workplace in another laboratory. Relatively lower luminance levels in general, however, have been identified in VR. Luminance in VR is technically limited by beamer specifications, rear projection and mirroring and can therefore not provide luminance at sunlight levels as may be available at realistic industrial work places with roof-lights or windows. In addition, illumination in VR is restricted to the projection wall to avoid optical reflexion and glare, whereas in reality reflexion may serve as an additional means to improve illumination. Illumination design in the VR laboratory will therefore remain an issue in future VR evaluation; especially when direct comparisons will be made between a workplace for human-robot interaction and its simulation in VR.

3.2 Simulator Sickness, Immersion and Presence in VR

The number of motion sickness reports tends to increase with increasing use of VR or other simulation techniques /13/. Results obtained from the pilot study yielded low level ratings for the simulator sickness total score as well as for the symptom subscales nausea, oculomotor and disorientation (see Fig. 2, left). According to a standardised six level classification for pre-post differences ranging from 'no symptoms' (SSQ = 0) to 'bad simulator' (SSQ > 20) by /13/ negligible symptoms could be identified in the pilot study only. In addition, the results for presented self-reports were much lower than reported on average across several other VR studies /13/.

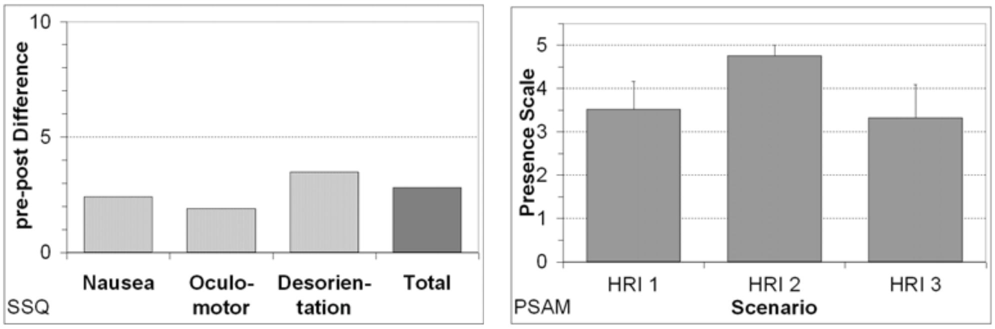

Figure 2*. Results for the Simulator-Sickness-Questionnaire (pre-post-differences) (left) and for the Presence Self-Assessment Manikin (right) for three consecutive experimental periods.*

Immersion in the context of VR is seen as a mental state during interaction experience that allows an operator the perception of evolvement and of being mentally and emotionally united with the virtual environment /15/. The Immersion Tendency Questionnaire /15/ discloses operator capabilities or tendencies to become involved or immersed, i.e. a pre-requisite to become united with VR and to experience presence. The ITQ ratings resulted in levels for the total score and the subscales (involvement, focus, games) not being significantly different from average levels obtained in other studies /15/. The analysis of the Presence Questionnaire /15/ showed the same effect. The operator experience of presence in the VR was at average level and therefore similar to most other simulation environments.

The Presence Self-Assessment Manikin /14/ is a screening instrument and therefore has also been administered to investigate differences in human-robot interaction across the three experimental periods. With a range between 1 and 5 the results indicated a high level for all periods (see Fig. 2, right). Statistically significant differences could be identified for experimental period ($F(2,6) = 4.986$, $p = 0.05$), with the middle experimental period in quadratic contrast to the initial and last period ($F(1,3) = 5.864$, $p = 0.09$). The results suggest that the experience of presence in the VR increased with the intensity of active human-robot interaction.

3.3 Operator Performance during Human-Robot Interaction in VR

Differences between experimental periods were also reflected in operator task performance by primary and secondary performance measures, psychophysiological responses and self-assessments of mental workload. Operators achieved RIT primary task performance at a very high level of above 95% correct responses under direct human-robot interaction. This result suggested allocation of attentional resources to primary task performance. Although the GRT remained the same across the three experimental periods, GRT secondary task performance was affected by differences in intensity of human-robot interaction. GRT response times were significantly different for experimental period ($F(2,6) = 7.253$, $p = 0.03$) (Fig. 3, left). Quadratic contrast analysis identified a performance decrement for the more intensive level of human-robot interaction ($F(1,3) = 12.285$, $p = 0.04$). The number of correct GRT (Fig. 3, right) could be maintained at a level above 95 % with statistically no differences between experimental periods ($F(2,6) = 2.533$, $p > 0.05$). Results for GRT performance indicated a substantial decrease in memory and attentional resources and suggested competition for spare cognitive capacity for task performance during more intensive human-robot interaction.

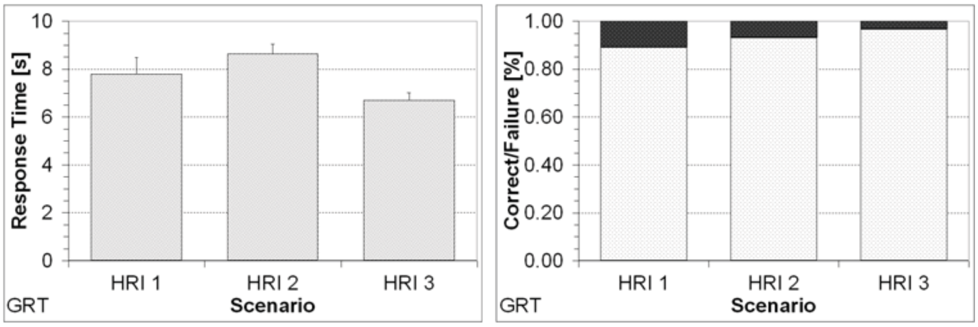

***Figure 3**. Results for GRT response times (left) and percentage of correct responses (right) for the GRT for three consecutive experimental periods.*

The competition among spare cognitive resources during the more intensive human-robot interaction period was also reflected in self-reports on mental workload. The results for the NASA TLX (Fig. 4, left) indicated significant differences for experimental period ($F(2,6) = 72.905$, $p < 0.01$) with a significant quadratic contrast for the middle as compared to the initial and final periods ($F(1,3) = 155.987$, $p < 0.01$). Analysis of the 0.1 Hz component of heart rate variability as a measure of effort investment revealed no significant differences for experimental period ($F(2,6) = 1.266$, $p > 0.05$). Figure 4 (right) suggests inter-individual differences having been too high to substantiate an effect on effort investment for the more intensive human-robot interaction period.

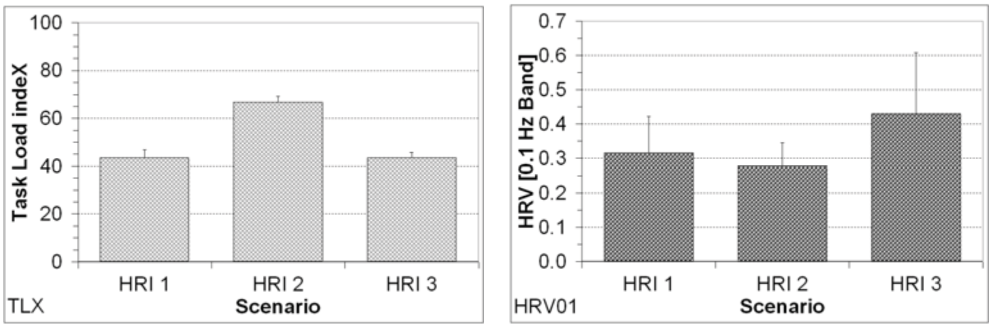

***Figure 4**. Results for NASA TLX self-report (left) and the 0.1 Hz component of HRV (right) for three consecutive experimental periods.*

4 DISCUSSION AND CONCLUSIONS

The aim of the preliminary and formative usability evaluation study was to investigate the design of the VR system to trigger VR design improvements and to prepare for more comprehensive usability evaluation studies in order to facilitate generalisation of research into VR applications for real world scenarios to the greatest possible extent. Basics in ergonomic design refer to usability of the VR system more in general. The usability related to human information processing has been addressed more specifically in a pilot study. Assessments of basics in ergonomics using a standardised screening instrument and heuristic evaluations resulted in efforts for re-design (e.g. hazardous obstacles, echo effect, climatic conditions) while requirements e.g. for illumination will remain an issue in future evaluations.

In the pilot study the VR system was used by a small group of users to achieve a high level of human performance with effectiveness, efficiency and satisfaction in the context of human-robot interaction /4/. In this context simulator sickness could be assumed a detrimental factor to affect effectiveness of human performance; however, for the present study the level for symptoms remained negligible. According to /13/ some symptoms for simulator sickness already take effect after short periods of time, but in general become more severe with increasing exposure to simulation environments. Although the experimental periods in the pilot study were rather short, effects could therefore have immerged if relevant. Despite these results, simulator sickness symptoms should be considered an issue for longer exposure to the VR system. This study also provided a specific context for testing the level of immersion and presence and yielded immersion and presence experienced at medium to high levels; consistently indicated by the ITQ, PQ and PSAM questionnaires. The experience of presence in the VR increased with more intensive human-robot interaction. A similar relationship has also been suggested by /15/. In the present study, in addition, this was accompanied by shifts in performance measures indicating an increase in human information processing demands. The measures taken for the investigation of immersion and presence seem to be suitable for future studies with the VR system and may serve as a marker to ensure effectiveness of human performance in terms of operator involvement in human-machine interaction in VR.

Information about the impact of the VR on human performance has also been based on results for performance measures, psychophysiology and questionnaires across periods different in intensity of human-robot interaction. The results indicate active and continuous human performance throughout the human-robot interaction scenarios. During the period with direct control for robot movement competition among attentional resources resulted in costs indicated by a decrease in secondary task performance and an increase in reported workload; suggesting a decrement in efficiency. Heart rate variability seems to be rather insensitive to variations in interaction intensity; probably due to inter-individual differences and the fact that cognitive rather than emotional processes were relevant for differences in interaction intensity /10/. In conclusion, these results suggest suitability of the measures during VR scenarios and provide evidence that an assessment of human information processing demands is feasible in the VR of the IFA. Under a perspective of accident prevention and product safety the results also suggest that care must be taken to avoid strong reductions of attentional resources in human-robot interaction when allocating multiple operator tasks. Even superficially self-evident information processing for active and direct human-robot-interaction may require extensive cognitive resources.

Although the empirical basis is too small to draw general conclusions the results provide evidence for a sound and an appropriate quality of VR system suitable for applied research in human-machine system design and evaluation. This is a promising basis for future VR evaluation studies that will include comparisons of human-robot interaction scenarios in VR and in reality. Applied usability research on accident prevention and product safety as the projects already initiated by accident insurance institutions can therefore be performed under fortunate circumstances.

5 REFERENCES

1. AGARD: Human Performance Assessment Methods, AGARDograph-308, Neuilly sur Seine, NATO, 1989.
2. Cobb S., Neale H., Crosier J., Wilson J.R.: Development and evaluation of vortual environments for education. In K.M. Stanney (Ed.), Handbook of virtual environments, Mahwah, LEA, 2002, 922-936.
3. EN ISO 6385: Ergonomic principles in the design of work systems, Brussels, CEN, 2004.
4. EN ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability, Brussels, CEN, 1998.
5. Hart S.G., Staveland L.E., Development of the NASA task load index (TLX). In P.A. Hancock, N. Meshkati (Eds.), Human mental workload, Amsterdam, North-Holland, 1988, 139-183.
6. Huelke M., Nickel P., Lungfiel A., Nischalke-Fehn G., Schaefer M.: Cave automatic virtual environments for research into occupational safety and health – practical recommendations and solutions for the construction. (this volume)
7. Määttä T.J.: Virtual environments in machinery safety analysis and participatory ergonomics. Human Factors and Ergonomics in Manufacturing, 17(2007) 5, 435-443.
8. Marc J., Belkacem N., Marsot J.: Virtual reality: A design tool for enhanced consideration of usability 'validation elements'. Safety Science, 45(2007), 589-601.
9. McCauley Bell P.: Ergonomics in virtual environments. In K.M. Stanney (Ed.), Handbook of virtual environments, Mahwah, LEA, 2002, 807-826.
10. Nickel P., Nachreiner F.: Sensitivity and diagnosticity of the 0.1 Hz component of heart rate variability as an indicator of mental workload. Human Factors, 45(2003), 575-590.
11. Schaefer M., Lüken, K.: Reasons for the manipulation (tampering) of protective devices, Proceedings of the 4th International Conference Safety of Industrial Automated Systems, SIAS 2005, Chicago, 2005, S 8, 7 p..
12. Stanney K.M., Cohn J.: Virtual environments. In G. Salvendy (Ed.), Handbook of human factors and ergonomics, Hoboken, Wiley, 2006, 1079-1096.
13. Stanney K.M., Kennedy R.S., Drexler J.M.: Cybersickness is not simulator sickness, Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting, HFES 1997, 1138-1142.
14. Wissmath B., Weibel D., Mast F.W.: Measuring presence with verbal versus pictorial scales: A comparison between online- and ex post-ratings. Virtual Reality, 14(2010) 1, 43-53.
15. Witmer B.G., Singer, M.J.: Measuring presence in virtual environments: A presence questionnaire. Presence, 7(1998) 3, 225-240.