In D. de Waard, K.A. Brookhuis, S.M. Sommer, & W.B. Verwey (eds.), *Human Factors in the Age of Virtual Reality* (pp. 117-128). Maastricht, the Netherlands: Shaker Publishing

Evaluation of a virtual reality-based ergonomics tutorial

Dietmar Gude, Eike Branahl, Philipp Kawalek, Andreas Prions & Wolfgang Laurig

Institut für Arbeitsphysiologie an der Universität Dortmund, Germany

Abstract

The interest in distributing education via the World Wide Web, in the form of web-based training (WBT), has increased greatly during recent years. Shortcomings of traditional hypertext WBTs have let to the proposal to merge them with virtual reality (VR) scenarios. To test the validity of this proposal, the usability of such a VR-enhanced WBT on the ergonomic design of computer workstations was investigated, from two perspectives: the utilisation of the content, in terms of page views and view times, and task performance, using sensitivity and bias indices of signal detection theory, and comparing them to data recorded in an equivalent physical laboratory setting. Results for sensitivity indicate that VR scenarios are suitable to represent the general spatial layout, but less suitable to represent object details, due to current technical limitations. Results for response bias, page views, and time data provide evidence for a cursory information processing, suggesting a high segmentation of the content.

Introduction

Recently much effort has been dedicated to developing e-learning applications for the World Wide Web (WWW). The potential benefits of such web-based trainings (WBTs) are manifold, like the use of standard software, rapid and frequent updating of the courseware, and the option to choose location, time, and speed of the study (Daniel, 1998). Despite widespread enthusiasm, the currently available results on the utilisation of WBTs are rather disillusioning, as illustrated by findings of a joint study of the American Society of Training and Development and the MASIE Center (ASTD & MASIE Center, 2001). They determined start rates of such courses at 16 companies in the U.S., the percentage of employees who actually started of those who were offered a course, which was 32% for voluntary programmes, on the average. Completion rates – specified in terms of the percentage of employees who actually finished of those who started a course – are not reported quantitatively but rather described as ‘notoriously low’.

Some authors (e.g., Wegerif, 1998) attribute low start and completion rates to the lack of social interaction and teamwork opportunities. Others (e.g., Nielsen, 1999) doubt that the social dimension is the main problem but rather focus on the usability of the corresponding applications. A common strategy in developing WBTs is repurposing already existing courseware for use on the WWW, resulting in applications that are mainly text-based, eventually supplemented by graphics, animations, or digital videos. However, there are research results that render such a simple porting and enriching of already existing material as problematic, due to the characteristics of how WWW content is processed. Most importantly, content is rarely studied word-by-word but rather scanned for keywords, examining corresponding sections in a selective way. This is illustrated by results from NetRatings (2002), which found mean times per page view (view times) of 37 seconds (Germany), 48 seconds (UK), and 54 seconds (U.S.), for example. These view time values are negatively correlated with the Internet access costs in these countries (OECD, 2001), implying that the cursory information processing of WWW content is, at least in part, due to a cost-induced time pressure.

Reviewing the research on traditional hypertext applications, Vora and Helander (1997) propose to merge them with virtual reality (VR), which embodies several characteristics of an ideal training medium (e.g., Stone, 2001; Wilson, 1999). First evidence for the utility of this concept comes from laboratory experiments by Howes *et al.* (2001) on e-commerce, which demonstrated that three-dimensional virtual shops allow better navigation and memory performance than virtual shops based on hypertext or hypertext with pictures in the hierarchy.

Several authors, however, warn to expect high levels of usability simply because information is represented in a three-dimensional format, especially if precise judgements of relative positions and distances are required, which are difficult because of the line-of-sight ambiguity of a perspective projection (e.g., St. John *et al.*, 2001; Wickens *et al.*, 2000). Beyond these content-related limitations, there are several technical aspects constraining the use of VR in the context of WBTs. In laboratory-based applications additional and still expensive gadgetry is used, like shutter glasses for stereoscopic vision, to which relatively few WWW users have access. Consequently, only some features of VR can be incorporated, in the form of a two-dimensional (2D) display to give the user the visual impression of a three-dimensional (3D) space, resulting in comparatively low levels of immersion. Furthermore, Howes *et al.* (2001) mention the high bandwidth requirements of their 3D shops, as another potential problem. Jacko *et al.* (2000) found that users may prefer highly graphical web sites, but are rather unwilling to tolerate substantial delays. Consequently, the size of VR scenarios needs to be kept to a minimum, by sacrificing their realism to some extent, using simple geometries and avoiding non-essential textures. This raises the question whether the utility of such 'minimalist' VR will actually resemble that of laboratory applications or real-world training.

In sum, the integration of VR scenarios is viewed as a promising way to overcome well-established deficiencies of more traditional hypertext WBTs, but also associated with qualified concerns about the utility of such applications. To contribute to the clarification of this question, an empirical evaluation of a VR-enhanced WBT on the ergonomic design of computer workstations was conducted. The application allowed the users to practice the evaluation of a workplace scenario, assisted by a checklist specifying the minimum ergonomic requirements.

Of the spectrum of usability dimensions relevant for evaluating WWW applications (see Vora & Helander, 1997, for an overview), the study focussed on two aspects, content utilisation and task performance, which were regarded as the most essential, in the present context. One aim was to collect practical data on the way in which the content was received and, especially, the start and completion rates, in a format compatible with the results provided by NetRatings (2002) and ASTD and MASIE Center (2001). This excludes a pure laboratory approach, but rather suggests the application of the online research paradigm (Batinic *et al.,* 2002). Correspondingly, the access to the WBT was unrestricted – except for a registration procedure providing general personal data – and the web server recorded which elements were visited for what period of time.

The second aim was to evaluate task performance, in relation to that in a real-world training situation. This was accomplished by analysing the users' responses to the checklist items, indicating to what extent they were able to identify ergonomic deficiencies of the scenario (hits) and avoided negative responses when there was actually no deficiency (false alarms). Hits and false alarms were analysed by referring to non-parametric indices of signal detection theory. The index chosen for sensitivity was A' (Grier, 1971), indicating the accuracy of the responses, the index for response bias was B''d (See *et al.* 1997), indicating the users' tendency to report an ergonomic deficiency, regardless of their sensitivity. These data were compared to those recorded in a laboratory setting, in which the participants evaluated an equivalent physical computer workstation. Concerning sensitivity, the WBT condition was expected to be generally inferior to the laboratory condition, due to a more cursory information processing, resulting in a speed-accuracy trade-off. This difference was expected to be moderate for the assessment of the general spatial layout of the workstation, but more profound for the evaluation of object details, as a consequence of the limited fidelity of the VR scenario. Concerning bias, no systematic differences between the WBT and the laboratory condition were expected.

Method

Participants

The analysis of *content utilisation* was based on 6825 visits of the WBT, recorded during 18 months. The *task performance* analysis was based on two samples, the WBT group and the laboratory group. The *WBT group* consisted of a subset of 211 visitors completing the task for the first time, 68 women and 143 men (68%), with a mean age of 31.6 years and a S.D. of 11.0 years. Their occupations covered a broad spectrum: pupils (9%), students (18%), teachers (9%), health professionals (10%), IT specialists (18%), other technical professionals (8%), clerks (19%), and other occupations (10%). The *laboratory group* consisted of 31 occupational health professionals of two insurance companies, 12 women and 19 men (61%), with a mean age of 41.2 years and a S.D. of 9.5 years. For the participants in this condition, the task served as a final practice at the end of a one-day seminar on the ergonomic design of computer workstations, organized by the research group.

Task

The participants were requested to identify ergonomic deficiencies of a computer workstation. In the laboratory condition a physical workplace was evaluated, in the WBT condition it was presented as a VRML97-model (Figure 1).

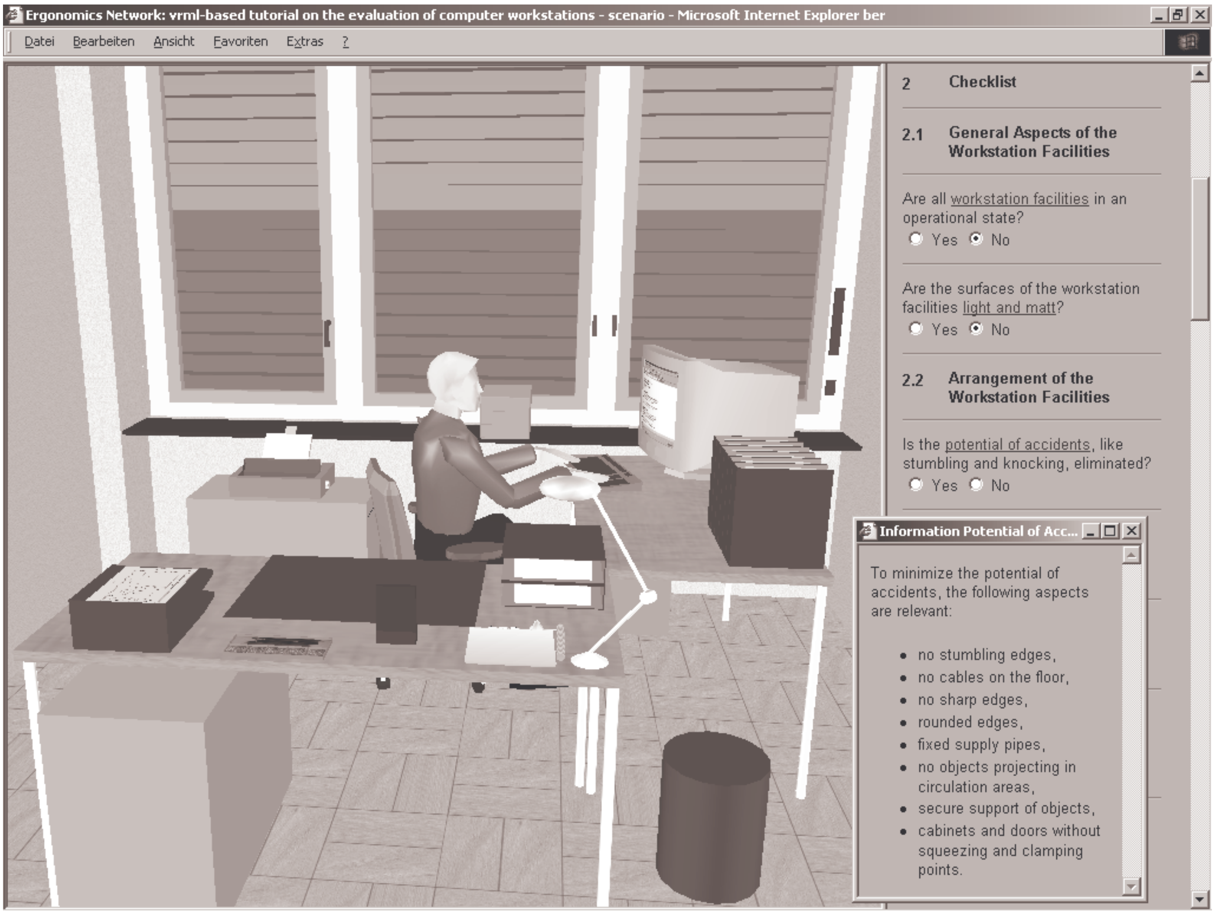

Figure 1. View of the computer workstation scenario from the entrance door (LHS) and the checklist with one of the pop-up windows (RHS).

The model included interactive elements mimicking those of real-world offices, allowing to open the door, the filing cabinet, and the drawers, moving the louvers up and down, switching the luminaries on and off, etc. To keep the size of the scenario within acceptable limits, some information was displayed in detail in seven separate popup windows (model pop-ups). An example are reflections on the screen, which were presented in a digital photo. The access of the model pop-ups was realised by defining hotspots, which changed the cursor to a pointing hand. The resulting size of the geometry file was 114 Kb; the size of the subsequently loaded textures was 187 Kb, in total.

In both the WBT and the laboratory condition the participants performed the evaluation by using a computerised version of Seidel-Fabian's (1998) checklist for screening computer workstations according to the German labour protection laws. The checklist consisted of 38 items of two types, 13 items were concerned with the *general spatial arrangement* (e.g., 'Is the monitor completely supported by the desk and does not project beyond its surface?'), 25 items were concerned with *object details* (e.g., 'Is the screen without any glare and reflections?'). For each item an additional explanation was available, which was displayed in a separate popup window (checklist popup), see Figure 1. The checklist pop-ups were retrieved by clicking on hyperlinks that were integrated in the text of the items.

With each item the response alternatives 'yes' and 'no' were presented as radio buttons. A 'yes'-response implied that the corresponding aspect of the computer workstation complied with the criterion specified in the item, a ‘no’-response signalled an ergonomic deficiency. The deficiencies represented typical problems at computer workstations, e.g., the screen surface was not at a right angle with the window front. Based on these deficiencies, 20 of the 38 items were to be answered in a negative way.

Procedure

In the laboratory setting the participants first filled a questionnaire assessing general personal data. Subsequently, the experimenter provided standardised instructions about the workstation, the evaluation task, and how to handle the checklist. The checklist was run on a Personal Computer with Windows NT 4.0 in a Navigator 4 browser. The computer recorded the answers to the checklist items and the time required to complete the evaluation. After finishing the task, the participants received an oral feedback about the actually present ergonomic deficiencies and the corresponding correct responses.

The WBT was run on a Personal Computer with Windows NT 4.0, Internet Information Server 4.0, and a T2 connection. The web server recorded the visits in standard log files. The visitors entered the application via an index page, the only element of the WBT that was linked with other, external pages or websites. The index page presented a short description of the service, together with the statement that users have to register, for methodological reasons. Via hyperlinks four other pages providing background information could be accessed, (a) the preview page with information about the content of the WBT, (b) the requirements page with technical information about the required hardware, software, and browser configuration, (c) the about page with information about the project background, and (d) the contact page with information about how to get in touch with the authors of the WBT in case of problems or questions.

A fifth hyperlink initiated the registration procedure. First-time users had to fill a form that recorded general personal data, e.g., age, gender, and occupation. Finally, the users selected a user name and password, for future logins.

After registration the scenario frameset with the model of the computer workplace in the left frame and the checklist in the right frame was presented. Having completed the checklist, the users pushed a button that submitted their responses to the web server, together with their user name and task completion time, and initiated the presentation of a feedback page.

The feedback page was identical to the submitted checklist, with the exception that the correct responses were marked by hyperlinks (the ‘yes’ or ‘no’ text of the radio buttons). Clicking on such a hyperlink, a feedback pop-up appeared, with an explanation why the marked alternative was the correct one. Moreover, at the end of the page the performance was summarised in a table, presenting the percentages of hits and false alarms, together with an explanation of these terms.

Results

Content utilisation

Based on the log files of the web server, the number of page views and the mean view times of the different elements of the WBT were computed, which are presented in Figure 2. According to this, the index page had the largest number of views, because all visitors entered the application via this page. The view time was about twofold of the average view time of German WWW users.

Of those pages that could be directly accessed via hyperlinks, the preview had the highest number of views, more than half of that of the index page. Compared to this, the visitors were much less interested in the technical requirements page, the about page, and the contact page. This implies that before registering most visitors first wanted to get more information about the content, but did not check in detail the identity of the source of this service and whether the WBT would actually work with their hard- and software. The view times of the four pages correlate with the amount of content – the description of the technical requirements was the most complex text, the contact information was the most elementary.

The start rate, the relation between the number of page views for the scenario and the index page, and the completion rate, the relation between the number of page views for the feedback and the scenario page, were both quite low. In about one-fifth of the cases the scenario was accessed, and in less than one-third of these cases the entire checklist was filled and submitted to get feedback. On the other hand, the view times of the scenario and the feedback page were both about sevenfold of the average value for German WWW users.

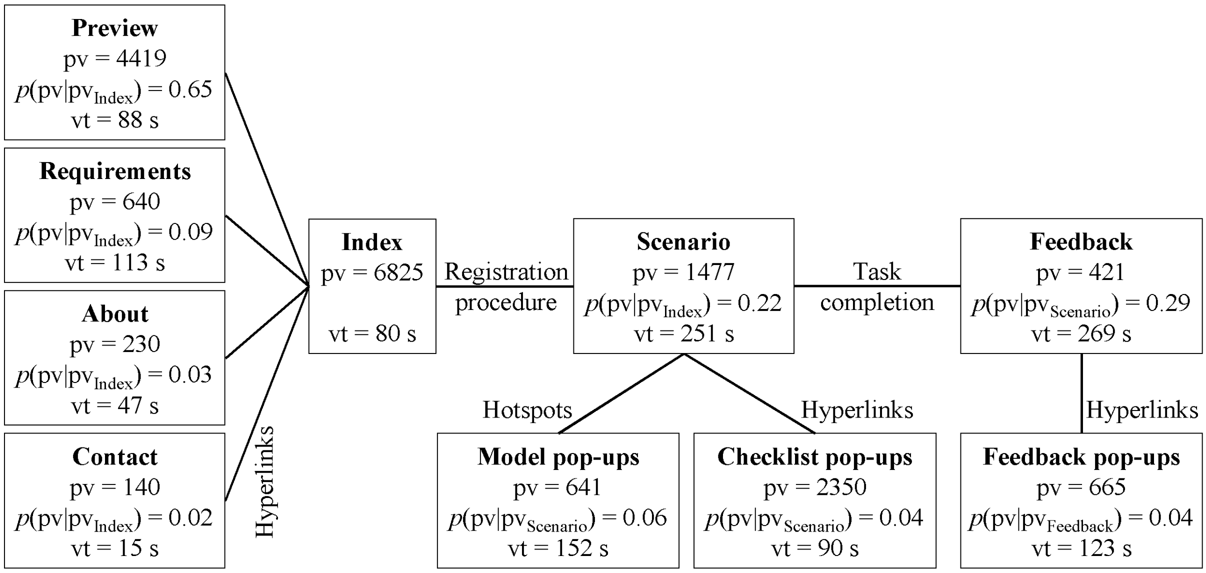

Figure 2. Elements of the WBT and their relationships, with the number of page views (pv), the conditional probability of a page view [$p(pv|pv_x)$], and the mean view time (vt). The conditional probabilities for the model, checklist, and feedback pop-ups were computed by dividing the number of page views by the number of pop-ups (7, 39, and 42, respectively) and by the number of page views of the superordinate page.

Scenario and feedback page contained links to three sources of additional information, the model, checklist, and feedback pop-ups. Generally, the conditional probabilities for the retrieval of the pop-ups were low, especially those for the checklist and feedback pop-ups. That is, while all sources of optional information were accessed rather infrequently, scenario related information was of more interest than the other explanatory texts.

Task performance

The mean task completion time in the WBT condition (M = 18.23 min, S.D. = 25.57 min) was significantly shorter than that in the laboratory condition (M = 47.03 min, S.D. = 19.22 min), $t(240)$ = 36.28 with $p < 0.001$. That is, the WBT users completed the task in less than half the time compared to those evaluating a corresponding real-world workstation, providing evidence for the expected more cursory information processing. On the other hand, they invested almost thirty fold of the average view time of German WWW users.

The sensitivity index A’ and the response bias B’’d were analysed in two separate 2 x 2 ANOVAs with the between-participants variable presentation (laboratory vs. WBT) and the repeated measurement variable item type (object details vs. general spatial arrangement). In the analysis of A’, the main effect of presentation was significant, $F(1, 240) = 9.19$ with $p < 0.01$, but there was also a marginally significant main effect of item type, $F(1, 240) = 3.81$ with $p = 0.05$, and a marginally significant interaction between presentation and item type, $F(1, 240) = 3.31$ with $p = 0.07$. In the analysis of B’’d, all effects were significant, the main effect of presentation, $F(1, 240) = 17.18$ with $p < 0.001$, the main effect of item type, $F(1, 240) = 60.81$ with $p = 0.001$, and the interaction between presentation and item type, $F(1, 240) = 22.56$ with $p = 0.001$. The effect patterns of the two variables are presented in Figure 3.

The sensitivity index A’ can vary between 0.5 (chance level) and 1.0 (perfect, 100% hits and 0% false alarms). In the laboratory condition object details and aspects of the general spatial arrangement were judged with intermediate and identical accuracy (M = 0.74, S.D. = 0.09 and M = 0.74, S.D. = 0.17, respectively). Compared to this, the WBT condition had a generally lower sensitivity (M = 0.64, S.D. = 0.12 and M = 0.70, S.D. = 0.16, respectively). However, the difference between the laboratory and the WBT condition was significant only for the object detail items, $F(1, 240) = 17.39$ with $p <$

0.001, not for the general spatial arrangement items, $F(1, 240) = 1.25$ with $p > 0.20$. The finding that the general spatial arrangement was evaluated more adequately than the object details is in accord with the initial expectation, that the representation of the latter would be problematic in the VR scenario. But the found accuracy in evaluating the general spatial arrangement, which was equivalent to that in the laboratory condition, was better than expected and underlines the efficiency of the VR scenario, especially if one takes into account that the WBT condition was considerably faster than the laboratory condition.

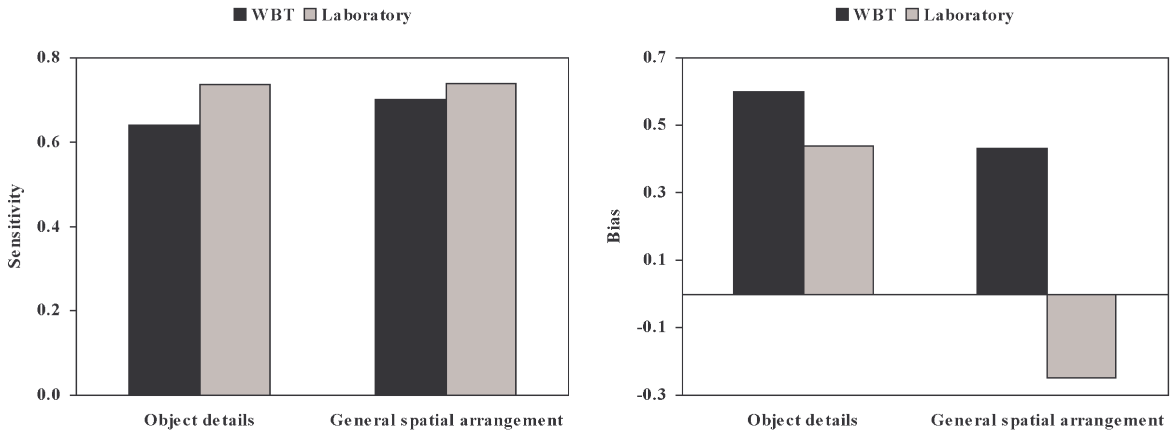

*Figure 3. Effects of presentation and item type on sensitivity A' (LHS) and response bias B''d (RHS). (A' = 0.5 + [(h-fa)(1+h-fa)]/[4h(1-fa)], B''d = [(1-h)(1-fa)-h*fa]/[(1-h)(1-fa)+h*fa], with h = hits and fa = false alarms.)*

The bias index B''d varies between –1 (by always responding 'no', 100% hits and 100% false alarms) and 1 (by always responding 'yes', 0% hits and 0% false alarms). In the laboratory condition there was a significant difference between the item types, $F(1, 30) = 39.26$ with $p = 0.001$, the bias for the object detail items was larger than that for the general spatial arrangement items ($M = 0.44$, S.D. = 0.40 and $M = -0.25$, S.D. = 0.70, respectively). That is, the general spatial arrangement of the computer workplace was analysed in a more sceptical way than the object details. The same difference was found for the WBT condition, $F(1, 210) = 18.53$ with $p = 0.001$, but here it was less pronounced ($M = 0.60$, S.D. = 0.52 and $M = 0.43$, S.D. = 0.67, respectively). Moreover, in the WBT condition the bias values were generally larger, indicating that the users were less sceptical about the ergonomic design of the workstation. This finding implies that the more cursory information processing did not affect the sensitivity, as expected initially, but rather the response bias.

Discussion

Summarising the results, the VR-enhanced WBT was of interest for a large number of people from a wide spectrum of occupations, who invested much more time than usual for WWW applications, and most importantly, showed a performance which was comparable to that of participants which completed the task under equivalent laboratory conditions. But the results also suggest some important qualifications, that currently VR scenarios are not necessarily the best choice for all types of content and that the task structure of the WBT was not optimally tuned to the specific way people are processing information on the WWW.

The sensitivity in judging the general spatial arrangement was not significantly different from that in a physical laboratory setting, supporting the notion that VR scenarios can adequately simulate aspects of a real-world situation (Stone, 2001; Wilson, 1999), which was even the case for the 'minimalist' application investigated here. The judgement of object details, on the other hand, turned out to be more difficult than in the laboratory condition. This result is to be attributed to the limited fidelity of the scenario, to allow its use on conventional computer equipment and to keep download time to an acceptable value, as WWW users are known to be rather unwilling to tolerate substantial delays (Jacko *et al.*, 2000).

The limited fidelity could not be compensated for by providing information on details in separate pop-up windows, as this optional information was retrieved rather infrequently. This could have been due to a low saliency of the corresponding hotspots, as the users did not know their position in advance, but rather needed to scan the scenario with the mouse, to detect elements of the model where the cursor changed to a pointing hand. On the other hand, for the checklist and feedback pop-ups as two other sources of optional information, which were accessed via hyperlinks, even lower retrieval rates were found. That is, the users had a pronounced tendency to generally ignore optional information, which implies that rather all performance relevant information need to be presented in a direct way.

Returning to the fidelity problem, the results make clear that presently VR scenarios are not the 'one best way' to display content – in case of subjects in which details are essential other media would be more adequate. But this problem is expected to be less grave in the near future, with the more widespread availability of high bandwidth and sophisticated computer hardware the quality of VR scenarios for the WWW will approach that of laboratory-based applications.

Not only the results on sensitivity, but also those concerning response bias indicated that the WBT simulated the conditions of real-world performance. As in the laboratory setting, the response criterion for the evaluation of object details was more liberal than that for judging the general spatial arrangement. But also a generally more liberal criterion was found. This result, together with the much shorter task completion time, provides evidence for a rather cursory than analytical information processing style, which is in line with the generally short times per page view of WWW users (NetRatings, 2002). One might expect that this problem would be mitigated, at least in part, by lower Internet access costs and especially a higher diffusion of time-independent cost models.

Finally, the low start and completion rates need to be addressed. This finding could be put into perspective, as there have been, undoubtedly, a number of visitors whose primary motivation was just to take a glimpse. But it nevertheless raises the questions what barriers existed and how they could be lowered or eliminated, to tap the user potential to a larger extent. Regarding the start rate, which was of similar magnitude as that found by ASTD and MASIE Center (2001) for voluntary courses, the general attractiveness of the topic and the content does not seem to have been a problem, as implied by the frequent access of the preview section. Rather, a considerable number of potential users might have been discouraged by the registration procedure, which was included for methodological reasons, to get information about the user population. Consequently, in applications without an online research background such a procedure should not be included. On the other hand, an identification mechanism is also essential for advanced, individual centred applications, e.g., for tracking progress and providing session-spanning feedback. If this is required, the registration procedure needs to be unobtrusive as possible.

Concerning the completion rate, the WBT obviously overstrained the endurance of many users, by simulating a complete evaluation of a computer workplace before providing performance feedback, which resulted in a mean task completion time that was almost thirty fold of the average view time of German WWW users. On this background, a more adequate solution would have been to break down the task into smaller units. That is, rather than asking for 38 responses at once, the task could have been structured according to different aspects (e.g., furniture, technical equipment, arrangement of the workstation facilities), or broken down to single items, each with a separate feedback loop. While it is evident that a 'bit-wise' presentation of WBT courseware is essential, to match the cursory information processing style of WWW users, general guidelines for the proper specification of content units are currently not at hand, which is one of the most urgent research questions in this field, from our point of view.

Acknowledgements

This investigation was conducted in the context of the joint research projects INTEGRAL and INTEGRAL II (Methodological integration of web-based multimedia and interactive tools to assist interdisciplinary learning in ergonomics programmes), which were supported by grants from the Joint University Competence Network for Multimedia and the German Ministry of Education and Research (Förderkennzeichen 08NM133A), respectively. We wish to thank our colleagues in these projects for their conceptual contributions, especially Ralph Bruder and Matthias Rötting.

References

ASTD & MASIE Center (2001). *If we build it, will they come?* Retrieved October 14, 2002, from http://www.masie.com/masie/researchreports/ASTD_Exec_Summ.pdf

Batinic, B., Reips, U.-D., & Bosnjak, M. (Eds., 2002). *Online Social Sciences*. Göttingen: Hogrefe.

Daniel, Sir J. (1998). Can you get my nose in focus? Universities, mass education and appropriate technology. In M. Eisenstadt and T. Vincent (Eds.), *The Knowledge Web* (pp. 21-29). London: Kogan Page.

Grier, J.B. (1971). Nonparametric indexes for sensitivity and bias: computing formulas. *Psychological Bulletin*, *75*, 424-429.

Howes, A., Miles, G.E., Payne, S.J., Mitchell, C.D., & Davies, A.J. (2001). Incidental memory and navigation in panoramic virtual reality for electronic commerce. *Human Factors*, *43*, 239-254.

Jacko, J.A., Sears, A., & Borella, M.S. (2000). The effect of network delay and media on user perceptions of web resources. *Behaviour & Information Technology*, *19*, 427-439.

NetRatings (2002). *Global Internet usage.* Retrieved October 14, 2002, from http://www.nielsen-netratings.com/corporate/corp_hot_off_the_net_i.jsp

Nielsen, J. (1999). *Designing web usability: the practice of simplicity.* Indianapolis: New Riders Publishing.

OECD (2001). OECD science, technology and industry scoreboard: towards a knowledge-based economy. Paris: OECD.

See, J.E., Warm, J.S., Dember, W.N., & Howe, S.R. (1997). Vigilance and signal detection theory: an empirical evaluation of five measures of response bias. *Human Factors*, *39*, 14-29.

Seidel-Fabian, B. (1998). Checkliste für Bildschirmarbeitsplätze [checklist for computer workstations]. In H.-D. Kempf (Ed.), *Fit am Bildschirm* (pp. 167-182). Reinbek: Rowohlt.

St. John, M., Cowen, M.B., Smallman, H.S., & Oonk, H.M. (2001). The use of 2D and 3D displays for shape-understanding versus relative-position tasks. *Human Factors*, *43*, 79-98.

Stone, R.J. (2001). Virtual reality: virtual and synthetic environments – technologies and applications. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (pp. 772-782). London: Taylor & Francis.

Vora, P.R. & Helander, M.G. (1997). Hypertext and its implications for the Internet. In M.G. Helander, T.K. Landauer, and P.V. Prabhu (Eds.), *Handbook of human-computer interaction* (pp. 877-914). Amsterdam: Elsevier.

Wegerif, R. (1998). The social dimension of asynchronous learning networks. *Journal of Asynchronous Learning Networks*, *2*. Retrieved October 14, 2002, from http://www.aln.org/alnweb/journal/vol2_issue1/wegerif.htm

Wickens, C.D., Thomas, L.C., & Young, R. (2000). Frames of reference for the display of battlefield information: judgement-display dependencies. *Human Factors*, *42*, 660-675.

Wilson, J.R. (1999). Virtual environments applications and applied ergonomics. *Applied Ergonomics*, *30*, 3-9.