Strain caused by Head Mounted Displays

Matthias Wille, Britta Grauel, & Lars Adolph
German Federal Institute for Occupational Safety and Health (BAuA)

Abstract

Head mounted displays (HMDs) have become smaller, lighter, brighter and less expensive during the last years. They are now on the threshold to mass market and a wider field of applications. During work they can be used as an adaptive assistant device showing context-sensitive information while hands are free for the main task. Within the project "head mounted displays – conditions of safe and strain-optimised use" of the German Federal Institute for Occupational Safety and Health (BAuA) a study was carried out: 41 subjects worked for four hours on a construction task and a parallel monitoring task. The content was either presented on a HMD or on a tablet PC. Besides performance within both tasks objective strain parameters (NASA-TLX, RSME) and a visual fatigue questionnaire were selected. Measurements at different timestamps gave insight into the strain trend over time comparing both display technologies. Results showed a higher strain and slower performance while working with a HMD compared to a tablet PC. Another outcome was a stronger increase over time in strain and visual fatigue while using a HMD, especially for elderly users. This implicates that working conditions, e. g. rest periods, have to be adjusted to fit to new technologies.

Introduction

In 1998, Keller and Colucci reported, that head mounted displays (HMDs) did not meet users´ expectations at that time. Nevertheless, they expected HMDs to be useful tools if design problems would be resolved. Now, 15 years later, the design of HMDs shows progress. HMDs are becoming lighter and smaller. Completely new solutions are at the prototype stage.

Monocular HMDs allow users to receive data from the real environment and additional information at the same time (Kawai et al., 2010). Thus, they can adopt the primary functions of job aids: provide information, prompt procedures of coach perspectives and decisions. The main intention of using job aids is to improve performance by reducing cognitive load (Rosset & Gautier Downes, 1991; Salvendy, 2012).

One advantage of HMDs over traditional job aids like manuals or new display technologies like handhelds is that they can provide the user with context sensitive information while working hands-free. Their usage is particularly suitable if tasks contain several steps, information can be provided in a graphical format, information is time critical and there is little tolerance for errors (for detailed information see

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Grauel, Kluge & Adolph, 2011). Nakanishi, Taguchi and Okada (2010) showed that visual on-task instructions using see-through HMDs are more suitable than auditory instructions for preventing careless skill-based errors. A comparison study (Glumm et al., 1998) showed that soldier performance of land navigating was better if using a helmet mounted display compared to using traditional navigating equipment. Soldiers also reported lower levels of mental workload.

However, HMDs can create problems which may not occur if traditional job aids or displays are used, because the viewing situation is not natural (Patterson, Winterbottom & Pierce, 2006). Kawai et al. (2010) stated that, in spite of technological progress, many questions regarding usability and usefulness remain unanswered.

To identify the conditions for strain-optimized use of HMDs to improve job performance a long term study which addresses the optimization of work and task load was conducted.

Method

To investigate strain and visual fatigue within longer sections of using head-mounted displays, a study was conducted in the laboratories of the BAuA. Participants worked in 3 sessions for 4 hours each with an HMD or comparatively with a tablet PC. Hereby the first session was always held with the HMD, while the repetition with the HMD and the comparison with the tablet PC were permutated as the 2^{nd} and 3^{rd} session among subjects. The repeated measurement with the HMD was arranged to estimate a first habituation to the new technology. Between the sessions were at least two and a maximum of 10 days.

Participants

A total of 41 subjects participated in the study with an age of 18-67 years. The sample was divided by median split into two age groups to investigate possible age effects. In the younger group were 21 subjects between 18-31 years (Mean = 23.48, SD = 3.341, 9 male /12 female). In the older group were 20 subjects aged 38-67 years (Mean = 50.55, SD = 9.214, 11 male / 9 female). None of the subjects had previously worked with an HMD.

Tasks

The subjects had to perform two tasks which were instructed as equal important (dual task paradigm): First, they should assemble model cars from building bricks, based on a famous construction toy. The assembly instructions were displayed on either the HMD or on the tablet PC. These image-based, step by step assembly instructions were chosen because it resembles industrial assembly instructions, such as those being used in maintenance. In a parallel presented monitoring task, subjects were asked to pay attention to three vertical bars on the outer edge of the screen and confirm certain state changes with a button press. The bars varied continuously, but very slowly in length and from time to time in its colour (blue / red). Colour change which includes a visual pop-out effect by the large change in presentation, should be confirmed by one button. On a second button each a change of position of the longest bar, caused by the continuous variation in length of the bar should be confirmed. The variation of the bars were programmed randomly, but on average a change in bar colour happened every 140 seconds and a change in the longest bar position every 95 seconds. As a further dependent variable in half of the blocks feedback was given on the last confirmation in form of a written notice on top of the bars (e.g. “centre - red”), while in the other blocks such feedback was not given.

Apparatus

The used HMD was a MAVUS-System from “Heitec” company, as it is currently used in industrial maintenance. It is a monocular look around display with a resolution of 800 x 600 pixels. The technique is fixed to a head carrier which includes a front camera and a headset. During the study the camera and headset had no function, while in industrial applications they are provided for communication with experts. The head carrier including all technology weighed 380 grams and was connected via a cable with a vest which included the radio technology for the transmission of data and the accumulator for power supply. As Tablet PC the CL900 by Motion was used. To ensure that the representation of the work content was comparable, only a window of 800 x 600 pixels was shown and the rest of the area was covered. All interactions - forward and backward switching the construction slides and buttons for confirming the bar tasks - took place over a converted number pad.

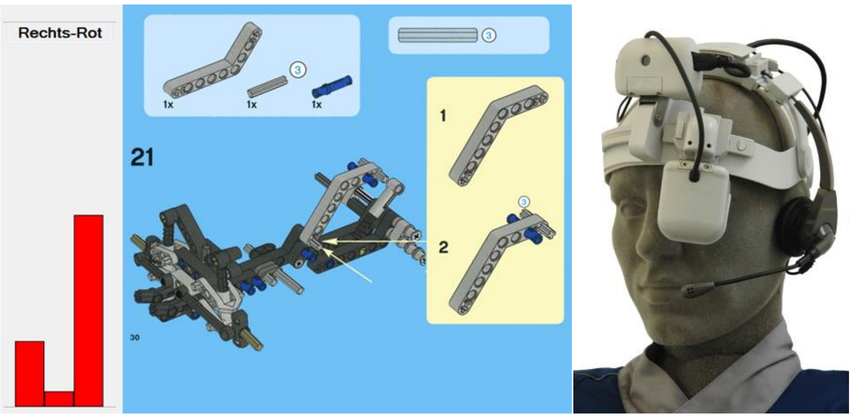

Figure 1. Left: work content as presented on the HMD. Right: HMD used in the study.

Procedure

The subjects were seated in a well-lit room in front of a height-adjustable table. At the beginning of the experiment, the table was set to the body dimensions of the participant. The experimenter sat behind a partition and monitored the session which was fully recorded on video. At the beginning of the first session participants filled out a questionnaire about demographical aspects and were informed about the test procedure in common and that they could cancel the experiment every time they want. After that, their experience with different kinds of technologies was asked and a questionnaire about technology affinity (TA-EG by Karrer et al., 2009) was completed. Then the RSME scale (Rating Scale of Mental Effort, Zijlstra, 1993) for monitoring mental strain was introduced. Subjects were asked to imagine different typical daily routine situations and to rate them on the RSME scale to get some practice in using the scale which was retrieved later at different points in time during the session. Afterwards, the instruction describing the tasks and the session procedure in common was given in a written form. If participants had any questions they were answered by the experimenter. Subsequently the HMD was shown to the subjects and all functions were explained. After that, the head carrier was adapted to the subjects and the HMD was put on. Hereby the monocular HMD was always mounted in front of the individual guiding eye of the subject and the bar-task always showed up in the outer region of the visual field (left side for left dominant eye, right side for right dominant eye). Then the combination of the two tasks was practiced for about 10 minutes in the first session. The following session was carried out in 4 blocks of 50 minutes each. There was a break of 5 minutes between the first and last two blocks where subjects keep the head carrier of the HMD on, but folded the display out of their sight. Between block 2 and 3 – in the middle of the session - was a major break from 15 minutes where the HMD was removed and the subjects were asked to stand up and move around a bit.

During the session, several subjective stress parameters were measured: The RSME, which reflects the perceived strain on a scale of 0-150, was administered in the middle and at the end of each block during the work. So in total eight measurement points are available to give insight in the development of strain over time. The NASA - TLX (Hart & Staveland, 1988) with its six subscales was not presented directly in the work situation, but right at the beginning of the big break and at the end of the experiment, so here are two measurement time points available. The pair comparison for the individual weight of the subscales was carried out at the end of the experiment. Furthermore, a questionnaire on visual fatigue (Visual Fatigue Questionnaire after Bangor, 2000) was administered at 5 time points: immediately before the experiment to determine the individual initial situation and directly after the end of each block. (As an objective strain parameter the heart rate was collected during the entire experiment, too. These data are still in analysis.) At the end of each test session, an interview was conducted with the subject.

Experimental design and variables

This study followed a repeated measurement plan with multiple dependent and independent variables. Independent variables are the experimental session - first measurement HMD (“HMD1”), second measurement HMD (“HMD2”) and tablet PC (“Tab”) - and the feedback in the monitoring tasks (blockwise change). The age and technology affinity of the subjects are between subjects factors (each median split into 2 groups). Dependent variable was the number of construction slides per session, as a performance indicator in the assembly task. This was possible because the complexity of each slide is similar across different car models and complex models primarily consist of more slides. Further dependent variables were the hit rate and reaction time in the monitoring tasks, RSME values at 8 times, NASA - TLX values at 2 times and Visual Fatigue Questionnaire at 5 times.

The data was analyzed in SPSS 20 using ANOVA with repeated measures. Here, two experimental sessions were always compared in pairs. A comparison over all 3 sessions at once would be not allowed with regard to contents because they do not represent levels of one factor, but in one case it is a repeated measurement to derive habituation effects and in the other case a comparison of different measurement technology. For purposes of clarity this article is primarily concerned with the comparison between the second HMD measurement and the Tablet-PC measurement, representing differences between display technologies. The second measurement with the HMD was chosen, because here a sufficient familiarization with the new technology already was given. The comparison between first and second measurement with the HMD which can describe habituation effects with the technology is also described if significant effects appeared.

Results

Construction task

As a performance indicator in the construction task, the number of construction slides processed per session was used. If comparing across both age groups with a HMD significantly less steps were done than when using a tablet PC (HMD1 = 130.9 , HMD2 = 128.8; Tab = 157.7). Taking the age factor into account the main effect of display stayed significant [$F(1, 37) = 24.285$, $p < .001$] and furthermore a highly significant main effect of age showed up [$F(1, 37) = 23.647$, $p < .001$]: Younger participants processed more constriction slides per session (see Figure 2). A habituation effect on the HMD by the comparison between the first and second measurement was not detected [$F(1, 37) = .208$, $p = .651$].

Monitoring task

The monitoring task showed a highly significant difference between the task types “bar colour” and “bar length”: changes in bar colour always had a higher hit rate [$F(1, 36) = 84.690$, $p < .001$] and lower response time [$F(1, 36) = 17.560$, $p < .001$] as the change of position of the longest bar (see tables 1 and 2). This effect illustrates the effect of a visual popout by the large variation in colour change and was expected. Similarly, the feedback on the last confirmed position led to a highly significant increase in the hit rate [$F(1, 37) = 58{,}257$, $p < .001$] in both bar tasks. But the reaction time showed no main effect of feedback [$F(1, 37) = .078$, $p = .782$]. The display type showed no significant effect on the hit rate [$F(1, 37) = 3.635$, $p = .065$] or reaction time [$F(1, 37) = 2.858$, $p = .100$]. However, the tendency is towards better values for the tablet PC. A habituation effect for the HMD by comparison of 1^{st} and 2^{nd} measurement showed up in terms of the hit rate [$F(1, 37) = 5.395$, $p = .026$], but not for the reaction time [$F(1, 37) = .196$, $p = .661$]. An effect of age was not found in the monitoring task.

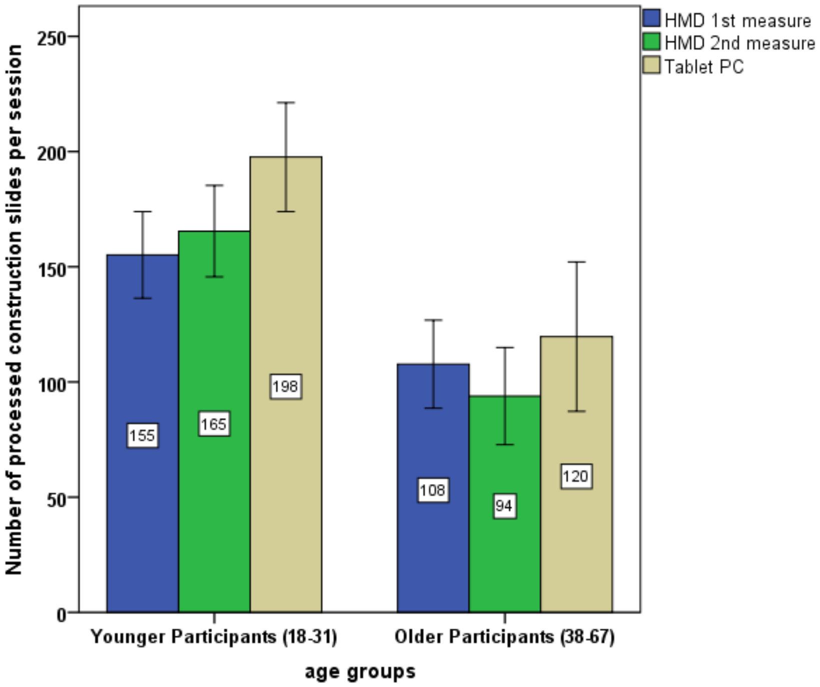

Figure 2. Number of processed construction slides per session by age groups. Error bars reflect the 95% confidence interval.

Table 1. Average hit rate in percent for the monitoring tasks by test session and feedback (standard deviation in parentheses).

	Feedback	*HMD 1st measurement*	*HMD 2nd measurement*	*Tablet PC*
Colour changes	no feedback	78.17 (15.91)	81.02 (14.99)	80.86 (15.12)
	given feedback	87.30 (15.27)	89.85 (13.25)	92.62 (11.18)
Length changes	no feedback	63.25 (23.43)	66.11 (21.96)	71.39 (18.45)
	given feedback	73.99 (21.58)	79.75 (18.07)	81.23 (14.07)

Table 2. Average reaction time in seconds for the monitoring tasks by test session and feedback (standard deviation in parentheses).

	Feedback	*HMD 1st measurement*	*HMD 2nd measurement*	*Tablet PC*
Colour changes	no feedback	9.60 (5.91)	11.50 (11.21)	11.86 (9.47)
	given feedback	13.68 (8.26)	12.85 (8.98)	11.50 (6.65)
Length changes	no feedback	20.29 (21.64)	20.93 (21.36)	17.66 (19.21)
	given feedback	19.42 (12.48)	19.99 (14.17)	16.14 (10.61)

Subjective Strain

The subjective strain measured by the NASA-TLX (see figure 3) showed a significantly higher score for the HMD compared to the tablet PC [$F(1, 37) = 26.952$, $p < .001$]. The increase of perceived stress over time is also significant [$F(1, 37) = 14.267$, $p =. 001$], but no interaction between display type and measurement time showed up [$F(1, 37) = 1.375$, $p =. 253$]. No main effect of age was found, but an interaction between display type and age reached statistical tendency [$F(1, 37) = 3.172$, $p =. 083$], based on a higher increase in strain over time on elderly participants. A habituation effect between 1st and 2nd measurement with the HMD can not be proved.

For the first time an influence of technology affinity showed up in interaction with the display type [$F(1, 37) = 5.000$, $p = .031$]: Participants with lower scores on technological affinity showed not only a higher stress score in general (HMD: 73.18, tablet PC: 66.17), but also less of the discharge in strain by the tablet PC compared to users with higher technology affinity (HMD: 68.96, tablet PC: 51.35).

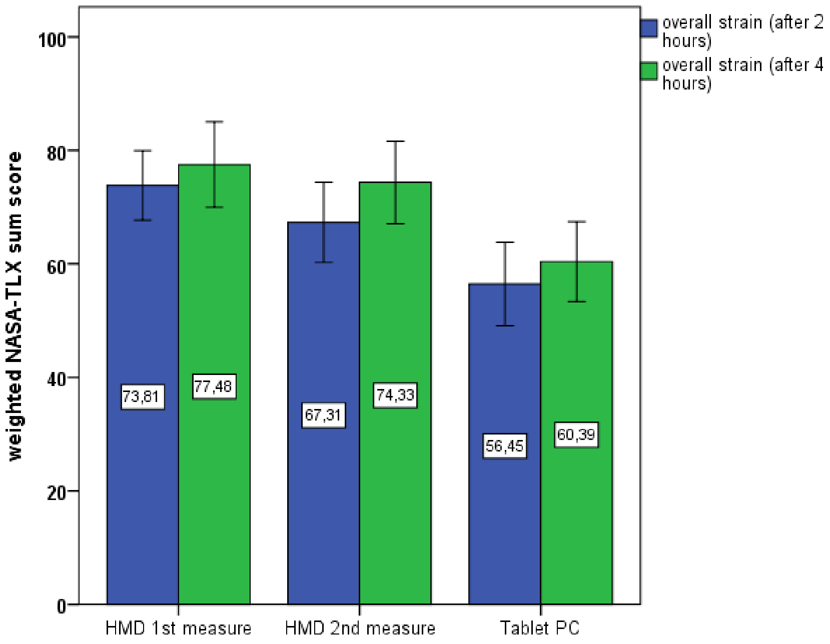

Figure 3. Weighted NASA-TLX sum score after 2 and 4 hours of working by session. Error bars reflect the 95% confidence interval.

The RSME value (see figure 4) also showed a highly significant increase in strain for the HMD compared to the Tablet PC [$F(1, 35) = 84.435$, $p < .001$], an increase over time [$F(1, 29) = 6.956$, $p < .001$] and a main effect of age [$F(1, 35) = 6.684$, $p = .014$], which again indicates that seniors are more stressed. The development of strain over time also showed the influence of breaks on the strain rating: The short breaks of about 5 minutes after 60 and 180 minutes – where the HMDs head carrier were kept on – seemed only to have little effects. However, the longer break after 120 minutes – where the HMDs head carrier was taken off – seemed to result in lower values on next strain request for younger participants, while older participants did not profit from that break.

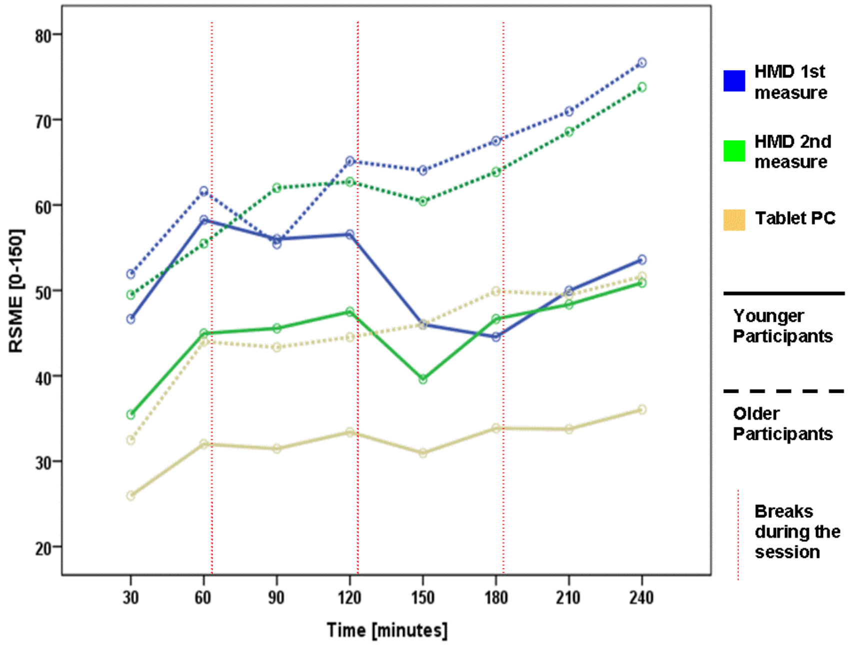

Figure 4. RSME score over time by session and age groups.

Visual fatigue

The Visual Fatigue Questionnaire (VFQ) collects various aspects of visual fatigue within 16 items. Each item is recorded on a ten-point scale, a sum score does not exist. The questionnaire was given before the start of the session and then hourly. Results showed a clear, highly significant change of almost all item scores, which however remain at a low average level of 1-3.5. All items showed an increase over time and higher strain levels on the HMD. For many items, the interdependency time x display was significant which reflects a greater increase over time during the HMD sessions.

As an example the item “difficulty in seeing sharp” showed a main effect of display [$F(1, 36) = 28.662$, $p < .001$] , a main effect of time [$F(1, 33) = 8.510$, $p < .001$] and an interdependency display x time [$F(1, 33) = 7.668$, $p < .001$]. But here also a reduction of the item scores in comparison between 1st and 2nd measurement with the HMD showed up [F(1, 37) = 9.319, p = .004] indicating that participants getting used to the display technology. Equally noteworthy are the items that are not directly attributable to visual fatigue: “headache” showed higher values for HMD [F(1, 37) = 21.435, p < .001], and an interdependency display x time [F(4, 34) = 5.546, p = .002] indicating a greater increase of headache over time while working with the HMD compared the tablet PC.

Effects on “neck pain” went into same direction with a main effect for display [F(1, 37) = 22.000, p < .001] and an interdependency display x time [F(4, 34) = 5.341, p = .002]. And also “mental fatigue” showed significant effects (display [F(1, 37) = 7.956, p = .008] , time [F(4, 34) = 16.665, p < .001], display x time [F(4, 34) = 3.688, p = .013]), but here also a habituation effect in the comparison between 1st and 2nd measurement of HMD showed up [F(1, 37) = 4.175 , p = .048].

In a final interview, many participants complained about the weight and discomfort of the HMDs head carrier. The monocular representation in front of only one eye was not a problem for most subjects after a short period of habituation. Only one of 41 subjects preferred the HMD for work, while the others found working with the Tablet PC more comfortable.

Discussion

The results generally show poorer performance and higher strain when working with a HMD compared to a tablet PC: participants work slower in the picture based construction task if presented on a HMD and the hit rate in a parallel monitoring task is also worse compared to presentation on a tablet PC. That last point illustrates that parallel monitoring is not necessarily easier, if the stimuli is presented always within the field of view. As attention needs to be focused on that area, too, even a presentation on a tablet PC, which is not always in direct sight, can result in a better reaction to stimuli.

One typical problem if investigating such new technologies like HMDs is that participants are often unfamiliar with this technology. Although they train tasks and handling of technology for a short moment before the measurement starts, their work performance and ratings will still be under the influence of getting used to the technology. Therefore in this study the repeated measurement with the HMD in a second session was done. However, habituation effects occurred only hardly in that second session with HMD. But habituation to such new technology may take longer than just 2 sessions of 4 hours each investigated during this study. Here a long term investigation will be useful that measures how people get used to that technology over weeks or months.

Some of the reported results are probably due to the hardware design of the used HMD: While most current HMDs still use a head carrier - similar to a bicycle helmet - to which the equipment is fixed, newer HMDs, which will be available during the next months or years, put all this technology in some slightly larger glasses. Head- and neck pain measured during this study could be based on the weight of the head carrier and be less dominant if using those upcoming HMDs. Also it is to mention that in this laboratory study, the benefits of HMDs - the mobility and the hands remain free - did not come into play, which could have a negative impact on the here reported mainly subjective parameters.

The approach of a long-term investigation and process-accompanying strain survey has proven itself and could often show a stronger increase over time in addition to higher strain with the HMD. This is especially important, if HMDs are used as a working assistance system. Based on these findings, recommendations for break times have to be reconsidered. For a more detailed presentation of the results and further discussion please also refer to the BAuA research report that will be published in 2014.

References

Bangor, A. W. (2000). *Display technology and ambient illumination influences on visual fatigue at VDT Workstations.* PhD thesis, Virginia Polytechnic Institute and State University, USA.

Glumm, M.M., Marshak, W.P., Branscome, T.A., Mc Wesler, M., Patton, D.J., Mullins, L.L. (1998). *A comparison of soldier performance using current land navigation equipment with information integrated on a helmet-mounted display.* Army Research Laboratory, Aberdeen Proving Ground.

Grauel, B., Kluge, A. & Adolph, L. (2011). Unterstützung makrokognitiver Prozesse in Teams durch Head Mounted Displays: Methoden der Eignungsermittlung und Gestaltung. In M. Grandt (Ed.): *Ergonomie im interdisziplinären Gestaltungsprozess: 53. Fachausschusssitzung Anthropotechnik der Deutschen Gesellschaft für Luft- und Raumfahrt Lilienthal-Oberth e.V. 27. und 28. Oktober 2011, Neu-Isenburg*. (pp. 71-79) Bonn: DGLR,

Hart, S.G., &Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.) *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009). Technikaffinität erfassen– Der Fragebogen TA-EG. In A. Lichtenstein, C. Stößel & C. Clemens (Eds.): *Der Mensch als Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme* (pp. 196-201). Düsseldorf, Germany: VDI Verlag GmbH. Online available: http://www.tu-berlin.de/fileadmin/f25/dokumente/8BWMMS/13.3-Karrer.pdf [August 2013].

Kawai, T., Häkkinen, J., Yamazoe, T., Saito, H., Kishi, S., Morikawa, H., Mustonen, T., Kaistinen, J. & Nyman, G. (2010). Ergonomic evaluation of ubiquitous computing with monocular head-mounted display. *Proc. SPIE 7542, Multimedia on Mobile Devices 2010*, 754202.

Keller, K. & Colucci, D. (1998). Perception in HMDs: what is it in head-mounted displays (HMDs) that really make them all so terrible? Proc. SPIE 3362, Helmet- and Head-Mounted Displays III, 46.

Nakanishi, M., Taguchi, K. & Okada, Y. (2010). Suggestions on the applicability of visual instructions with see-through head mounted displays depending on the task. *Applied Ergonomics, 41*, p. 146-155.

Rosset, A. & Gautier-Downes, J. (1991). *A Handbook of Job Aids.* San Francisco: Jossey-Bass Pfeiffer.

Salvendy, G. (2012). *Handbook of Human Factors and Ergonomics.* Hoboken: John Wiley and Sons, Inc.

Patterson, R., Winterbottom, M.D. & Pierce, B.J. (2006). Perceptual Issues in the Use of Head-Mounted Visual Displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 48, 3,* pp. 555-573.

Wille, M., Grauel, B. & Adolph, L. (2011). Methodische Vorgehensweise zur Beanspruchungsermittlung bei Head-Mounted Displays. In *Reflexionen und Visionen der Mensch-Maschine-Interaktion – Aus der Vergangenheit lernen, Zukunft gestalten. 9. Berliner Werkstatt Mensch-Maschine-Systeme*. (pp. 84-85) Düsseldorf, Germany: VDI Verlag, Fortschrittsberichte, Band 33.

Zijlstra, F.R.H. (1993). *Efficiency in work behavior: an approach for modern tools.* PhD thesis, University of Delft. The Netherlands. Online available: http://repository.tudelft.nl/view/ir/uuid:d97a028b-c3dc-4930-b2ab-a7877993a17f/ [August 2013].