

Comparison Between the Strain Indicator *HRV* of a Head-Based Virtual Retinal Display and LC-Head Mounted Displays for Augmented Reality

Olaf Oehme
Ludger Schmidt
Holger Luczak

Chair and Institute of Industrial Engineering
and Ergonomics, RWTH Aachen University,
Aachen, Germany

Augmented Reality uses Head Mounted Displays (HMD) to overlay the real world with additional virtual information. Virtual Retinal Displays (VRD), a new display technology, no longer requires Liquid Crystal Displays (LCD). VRD technology addresses the retina directly with a single laser stream of pixels. There are no studies on the user's informational strain in this new VRD technology. Various papers have shown that Heart Rate Variability (HRV) is a valid indicator for the user's informational strain. An empirical test revealed no difference in the user's HRV between VRD technology and LCD technology. Consequently, there seems to be a comparable user informational strain regarding the display types.

virtual retinal display heart rate variability workload
strain augmented reality

The research project "Augmented Reality in Design, Production and Service" (ARVIKA) is funded by the German Federal Ministry for Education and Research (BMB+F) within the research under grant 01 IL 903 R 4. Dr. Krahl from the Informationstechnik department of DLR deserves additional special gratitude for his kind support of the research projects. Furthermore, we want to thank all partners in the ARVIKA project who supported this research, especially Wolfgang Wohlgemuth from Siemens AG Automation & Drives.

Correspondence and requests for offprints should be sent to Olaf Oehme, Chair and Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Bergdriesch 27, D-52065 Aachen, Germany. E-mail: <o.oehme@iaw.rwth-aachen.de>.

1. INTRODUCTION

In Augmented Reality (AR) the real world is superimposed with additional virtual information. Thus, for example, repair instructions for a machine tool or important installation advice can be directly superimposed in the worker's field of view (Luczak, Wiedenmaier, Oehme, & Schlick, 2000). To realize this enrichment technically, Head Mounted Displays (HMD) are necessary. Nowadays, Liquid Crystal Display (LCD) technology is the most common technology. When using an LCD an additional half-silvered mirror is needed to see both the real world and the virtual information. A reduction in illuminance of the real world for the human eye (an effect similar to looking through sunglasses) is the disadvantage of this method.

Virtual Retinal Displays (VRD) are another new display technology. They project the objects with a single laser stream of pixels directly onto the retina. This results in a very clear and sharp projection of different information. Furthermore, a high light transmission of the real worlds pictures can be ensured (Microvision, n.d.). First investigations on VRD show a quicker identification of virtual information, especially of graphical symbols (Oehme, Wiedenmaier, Schmidt, & Luczak, 2001). There are no empirical studies on the user's informational strain with this new VRD technology. However, such studies are urgently necessary to ensure a permission for using VRD technology in an industrial environment because of health and safety reasons for the working personnel.

2. METHODOLOGY

2.1. The Independent Variable, Display

In a counterbalanced treatment order, four different display types were tested. The first display was the Retinal Scanning Display, a monocular prototype of a Virtual Retinal Display (VRD) from Microvision Inc. (USA), which can be used either on the right or on the left eye. It has a vertical field of view (FOV_V) of 21.37° . Superimposed information appears in monochrome red ($\lambda = 635 \text{ nm}$).

The other three displays were LC-HMDs. The Clip On Display from Microoptical Inc. (USA) was also a prototype. It is a monocular display, which can only be used on the left eye. This Microoptical Display is a so-called Look-around-Display, which means that a small monitor is now

placed in front of the eyes. Having a very small *FOV* ($FOV_V = 11.62^\circ$) with this type of display one can only see the virtual information displayed. The possibility of seeing through to “real” reality is not given in this case. Only by looking over the display, can the user look at the real world. Following an accurate definition, this display type is not a real AR display, but rather a Wearable Display, by which at least a video-based AR can be realized.

The Glasstron from Sony (Japan) was the third display ($FOV_V = 15.48^\circ$, binocular) and binocular i-glasses ProTech from i-O Display Systems (USA, $FOV_V = 17.24^\circ$) were the fourth display used.

The displayed information was monochrome red in order to avoid controlling color perception. All displays had a screen resolution of 640×480 pixels and an image frequency of 60 Hz.

2.2. Dependent Variables

There are various methods to measure the user’s informational strain. Due to a relatively easy applicability subjective measurement methods are often used, for example, NASA TLX-Task Load Index (Hart & Staveland, 1988) or rating scales like the category-division technique from Heller (Heller, 1981). However, subjective methods can be influenced by the user’s prejudice against such high tech equipment (Wierwille & Eggemeier, 1993). Compared to a regular product, attitudes towards prototypes can be that they are neither user-friendly nor do they have the same optical and haptic characteristics as a regular product. Statements such as “I will never wear such a monster in my work” can lead to influenced results and rejection of a technically excellent technology. Another disadvantage of subjective measurement methods are the large inter-individual differences among the users’ judgements, which sometimes makes a clear interpretation of the data impossible. (Pfundler & Schweingruber, 2000).

2.2.1. *The normalized arrhythmia quotient ARQ_N*

To measure the user’s informational strain there are also various psychophysiological indicators, such as the heart rate (*HF*) or the heart rate variability (*HRV*). Measuring these indicators requires a great deal of equipment, specialized knowledge, and a large-scale evaluation. Normally, this effort only pays for research studies (Baggen & Hemmerling, 2000). Due to the fact that ARVIKA is a research project we decided to use the *HRV*. Furthermore, a comparison between prototypical and regular displays is necessary. By inves-

tigating literature one can see that there exist various different definitions for *HRV*. For this study the *ARQ* (arrhythmia quotient) as described and evaluated in Luczak and Laurig (1973) and Laurig, Luczak, and Philipp (1971) was used:

$$ARQ = \frac{\sum_{i=1}^{n-1} (HF_i - HF_{i+1}) \nabla (HF_i - HF_{i+1}) > 0}{H_n \left\{ \left[(HF_i > HF_{i+1}) \wedge (HF_i > HF_{i-1}) \right] \vee \left[(HF_i < HF_{i+1}) \wedge (HF_i < HF_{i-1}) \right] \right\}}$$

A higher arrhythmia results in a higher *ARQ* value and can be interpreted (by constant low energetic effectoric load) as smaller informational strain (Luczak, 1975).

Because of a usually very large deviation of the *ARQ* values among the participants, the normalized *ARQ* (*ARQ_N*) was additionally calculated for a better comparability:

$$ARQ_N = \frac{ARQ - ARQ_{\min}}{ARQ_{\max} - ARQ_{\min}}$$

In this case, a normalization relating to a resting value was not practicable. First, because mental resting conditions cannot be controlled (Luczak, 1975; Luczak, Philipp, & Rohmert, 1980) and second, because the participants had only a limited amount of time available from their daily working time.

2.2.2. *The normalized 0.1-Hz component of the HRV, HRV0.1_N*

Besides examining heart frequencies in time periods, the data can additionally be transformed from a time into a frequency range and be analyzed by a spectral analysis, for example, by Fourier Analysis. Depending on the spectrum different physiological phenomena are allocated to this power spectrum (power density), which describes the Heart Rate Variability in its frequency range. From those three spectra considered, the changes of the second one in the middle (LF: 0.05–0.15 Hz)—the so-called 0.1-Hz component—are connected with a short blood pressure regulation, the baroreceptor effect (Luczak & Laurig, 1973; Luczak & Raschke, 1975). This baroreceptor activity is deemed to be an indicator of mental strain during the process of information processing. It allows to estimate the course of the attention modulation in connection with the cognitive effort necessary for task coping (e.g., Luczak,

1998; Manzey, 1998; G. Mulder, 1979; L.J.M. Mulder, 1992; G. Mulder, L.J.M. Mulder, Meijman, Veldman, & van Roon, 2000; G. Mulder & L.J.M. Mulder, 1981; Richter & Hacker, 1998). In the same way as with as with the *ARQ*, with the 0.1-Hz component an increase in cognitive effort is linked with an amplitude decrease in the 0.1-Hz spectrum.

2.2.3. Notes on the measurement of indicators of physiological strain

Neither the currently discussed subject matter, whether psychological stress and strain are generally measurable or not, nor the claim of a commonly accepted and practical theoretical model about what is understood by psychological stress and strain, what one really wants to measure with it or believes to be measuring (Nachreiner & Schütte, 2002), do keep us from operationalizing variables as already done in this article. Generally, different assumptions are stated. The surveys used to support Schmidtke's (2002) and Nachreiner's (2002) doubts on the measurability of psychological stress and strain mainly refer to an absolute significance of the calculated stress level of completely different work activities with the conclusion that degrees of heart frequency and arrhythmia

- are only applicable for comparative studies that aim at determining the relation between different forms of stress (Schütte & Nickel, 2002) and
- are only applicable for studies that are solely interested in detecting a (partial) component of psychological strain, namely a task specific excitation or activation or emotional strain (Nickel, Eilers, Seehase, & Nachreiner, 2002).

Due to the fact that these conditions were given in our experiment (comparison of different HMDs, i.e., forms of strain), the dependent variables can be operationalized as before, regardless of a discussion whether the aforementioned critique can be justified for a subject, on which research has already been done for many years or whether the methods used can prove a counter evidence (e.g., Oesterreich, 2003).

2.2.4. Reaction time and false response percentage

In addition to the *HRV* indicators the performance of information processing, reaction time, and false response percentage were measured. Reaction time was thereby operationalized by the time difference between the presentation of the signal and the reaction of the test person. The analysis only included the mean value of those reaction times in which the tasks were done correctly (Green & Sweets, 1988).

2.3. Participants

Twenty-five participants were recruited from the ARVIKA project partner Siemens AG (Germany), the department for Automation and Drives. All participants were service technicians or service engineers, aged between 21 and 50 years with a mean of 36.9 years ($SD = 8.3$). In earlier tests it was found that subjective measures of vision problems and tiredness differed significantly between male and female participants. Because of an almost exclusively male target group of an AR system in the service area it was decided to investigate male participants only. Participation was voluntary and the participants were allowed to finish the experiment at any time.

2.4. Procedure

Before beginning the experiment, the test person's sight was tested (farsightedness, shortsightedness, stereoscopic sightedness) and the dominant eye was determined. After they had been introduced to the technical features, all participants gave their consent to have their heart rate measured during the test.

The participants had to work on four different tasks with all displays that were selected from tasks defined in the requirements phase of the ARVIKA project (the participants had to find the displayed virtual information—forms, letters, and graphical symbols—on paper in the real world; they had to confirm whether the paper contained the displayed information or not). During the whole time of the experiment the participants wore a portable computer (Xybernaut MA IV, USA) around their hips fixed on a belt. To interact—to confirm or to negate—the participants had to push the left or the right mouse button on the Xybernaut.

After each display the participants had a short break before solving the tasks with the next display. The display sequence was counterbalanced among the test persons.

The heart rate (HF_i) was measured with a mobile digital electrocardiogram monitor (BHL-6000 of the Med-Natic Company, Germany) with the ability to set marks on characteristic moments. These marks were used to differentiate between each time period during data preparation.

The experiment was executed in a darkened room with lightning from a nonflickering lamp with constant illumination. The participants were seated on a chair (low energetic effectoric load) in a fixed position in front of a black background.

3. RESULTS

From the 25 participants, the data of 24 were taken for the evaluation; one test person's data were excluded from the analysis due to technical defects during the execution of the experiment, which caused unequal test conditions that would have lead to an inaccurate analysis. Both the ARQ_N and $HRV0.1_N$ show significant differences between the mean value of all working time periods and the mean value of all break periods ($p < .01$).

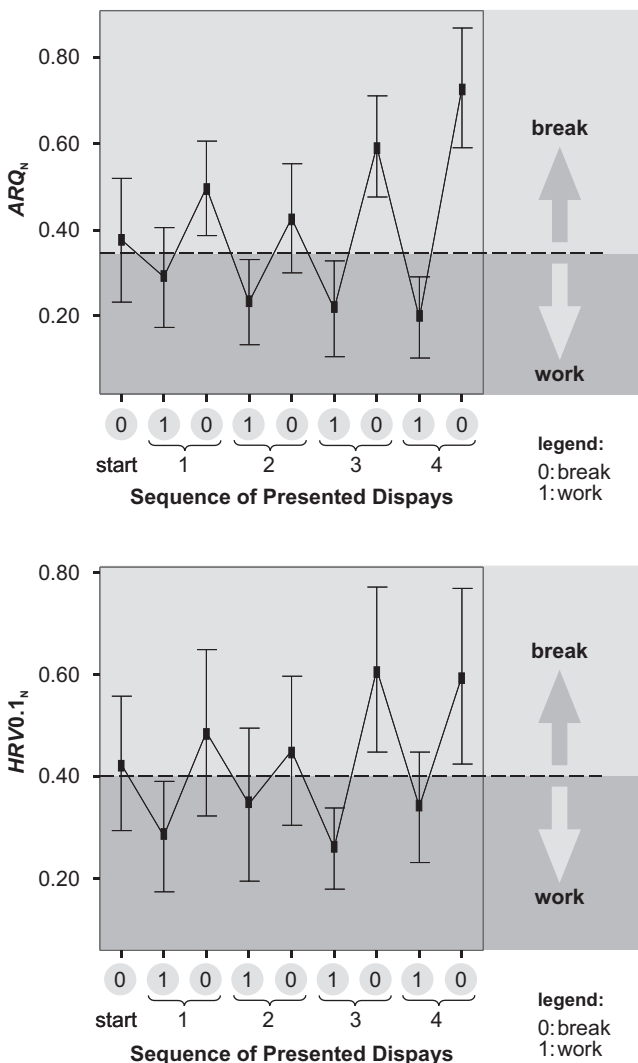


Figure 1. ARQ_N and $HRV0.1_N$ as well as 95% confidential interval dependent on time period (work, break). *Notes.* ARQ_N —normalized arrhythmia quotient, $HRV0.1_N$ —normalized 0.1-Hz component of Heart Rate Variability.

However, the Tukey-HSD Post-Hoc-Test showed that there were no significant differences between the four working time periods; there were only significant differences between the working periods and break periods ($p > .05$). Figure 1 shows smaller ARQ_N and $HRV0.1_N$ values during the working periods than during the breaks and this can be interpreted as higher informational strain in the working periods than in the break periods (expected as natural). So, the ARQ_N and $HRV0.1_N$ appear to be a valid indicator in our experimental design (manipulation check).

Concerning the independent variable display, the ANOVA shows that none of the null hypothesis could be denied, $F_{ARQ_N}(3, 69) = 0.858, p = .466$; $F_{HRV0.1_N}(3, 69) = 2.286, p = .084$. The very improbable probability of error of about 8.4 % for the 0.1-Hz component, however, makes a closer examination of the results necessary (Figure 2, Table 1).

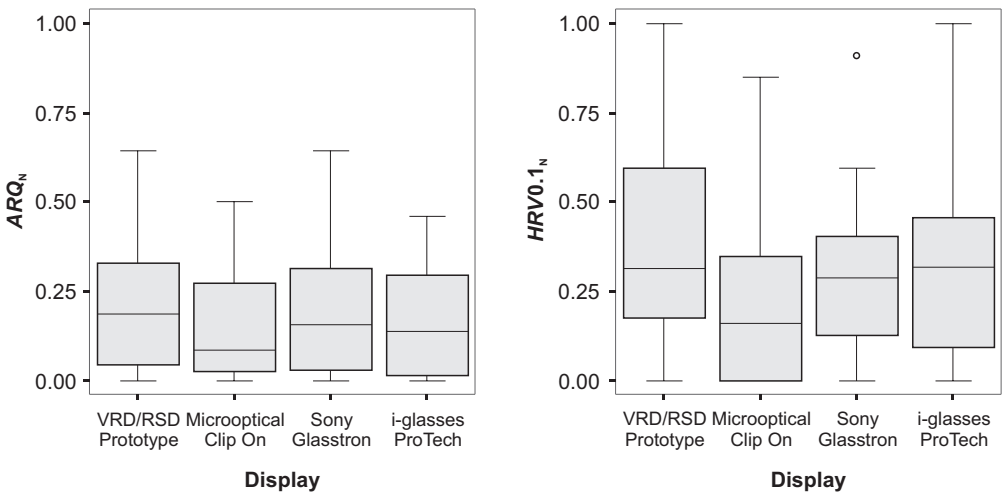


Figure 2. ARQ_N and $HRV0.1_N$ boxplots (median, quartile, extremum) dependent on display type. Notes. ARQ_N —normalized arrhythmia quotient, $HRV0.1_N$ — normalized 0.1-Hz component of Heart Rate Variability.

As can be seen, among the person subgroup examined, Virtual Retinal Displays generated the highest mean value for the ARQ_N as well as for the $HRV0.1_N$. This is an indication for less strain when using this kind of display. Due to the fact that the present differences between the mean values of the strain indicators are not statistically secured by the chosen level of significance ($p = .05$), the corresponding null hypothesis cannot be denied. Thus, neither higher nor lesser strain can be concluded when using a certain display.

TABLE 1. ARQ and HRV0.1_N Mean Values Dependent on Display Type

Display	ARQ _N		HRV0.1 _N	
	M	SD	M	SD
VRD Prototype	0.227	0.205	0.388	0.291
Microoptical	0.151	0.155	0.201	0.227
Sony Glasstron	0.193	0.184	0.290	0.229
i-glasses ProTech	0.165	0.151	0.348	0.298
Total	0.184	0.175	0.307	0.269

Notes. ARQ—arrhythmia quotient, HRV—Heart Rate Variability, VRD—Virtual Retinal Displays.

This also means that higher strain of the examined Laser Retinal Display compared to the other displays that were examined and are available on the market is not verifiable. On the contrary, these descriptive data lead to the exact opposite. However, this could only be proved by an eased level of significance, $p = .1$.

Concerning reaction time as well as false response percentage the ANOVA showed that the difference between the mean values were statistically significant, $F_{rt}(3, 69) = 15.809, p < .01$; $F_{frp}(3, 69) = 22.640, p < .01$. Results from the paired comparison concerning statistical significance are shown in Figure 3 (Tukey-HSD Post-Hoc-Test).

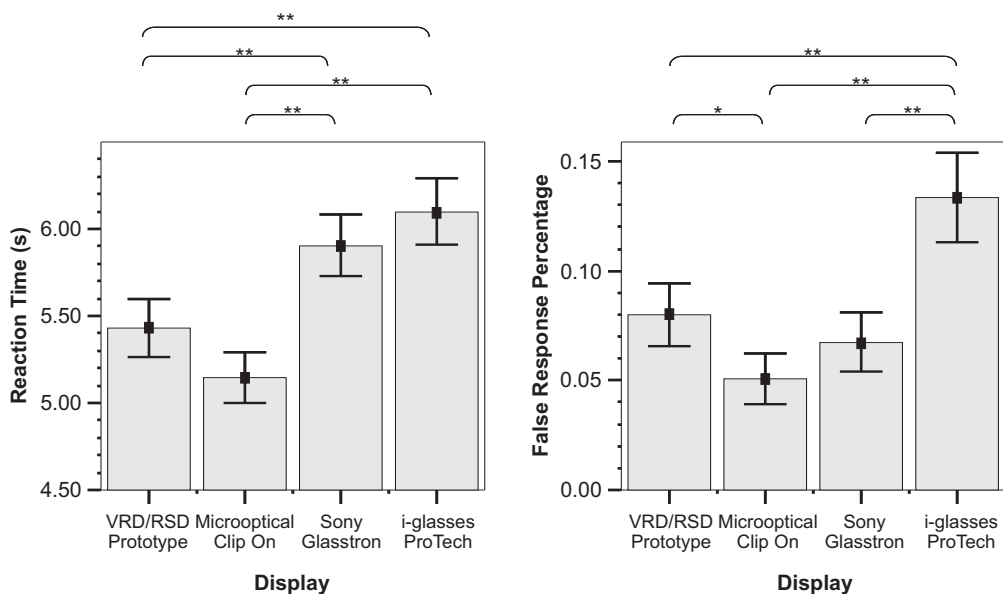


Figure 3. Reaction time (rt), false response percentage (frp), and 95% confidential interval dependent on display type. Notes. * $p < .05$, ** $p < .01$.

When using both monocular display types (VRD and Microoptical, Germany) the processing rates were significantly higher compared to the other two binocular displays, the Sony Glasstron Display (8.6 and 13.7%, respectively) and i-glasses (10.9 and 15.9%, respectively). However, a difference between the two monocular as well as between the binocular displays could not be proved ($p > .05$).

Regarding the false responses it could be shown that when using one of the first three displays (VRD, Microoptical, Sony) 39.8, 68.4, and 49.6%, respectively, fewer false responses were registered than when using the display that came off the worst (i-glasses).

4. DISCUSSION

With all displays, the same degree of difficulty, as well as the same number of tasks could be managed. Thus, we can conclude that the use of different displays can lead to different perceptions of strain. In literature the *HRV* is described as a valid indicator for informational strain (in our case the ARQ_N and $HRV0.1_N$). Also in the case of our experimental design the $ARQ_N/HRV0.1_N$ is appropriate for distinguishing the working periods and breaks, but it was not appropriate for verifying a significant difference between the displays.

This fact made an additional consideration of the dependent variables, reaction time and false response percentage, necessary. Compared to the other two optical HMDs (Sony Glasstron and i-glasses), the VRD permitted faster information processing and compared to the i-glasses an even lower false response percentage. This means that with the same stress higher visual performance and thus higher effectiveness could be achieved.

Compared to the VRD, only the Microoptical HMD had a lower false response percentage. Concerning reaction time, no differences could be proved. In this connection, the different technology has to be pointed out explicitly. It allows no optical overlay, which leads to the fact that no optical AR can be provided.

At this point we can conclude: Concerning informational strain it can be said that it seems that the VRD prototype has comparable qualities to LC-HMDs. For an effective AR use, which obligatorily needs an optical overlay, it also shows far better values concerning visual performance and should thus be preferred. Insofar as we solely considered a wearable presentation of information, the Clip-On display from Microoptical comes off best concerning false response percentage. However, regarding the very small *FOV*, it has other disadvantages concerning the amount of information to be shown in the display.

REFERENCES

- Baggen, R., & Hemmerling, S. (2000). Evaluation von Benutzbarkeit in Mensch-Maschine-Systemen [Evaluation of the usability in man-machine-systems]. In K.-P. Timpe, T. Jürgensohn, & H. Kolrep (Eds.), *Mensch-Maschine-Systemtechnik* (pp. 233–284). Düsseldorf, Germany: Symposium Publishing.
- Green, D.M., & Sweets, J.A. (1988). *Signal detection theory and psychophysics*. New York, NY, USA: Wiley.
- 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, The Netherlands: Elsevier.
- Heller, O. (1981). *Theorie und Praxis des Verfahrens der Kategorienunterteilung (KU)* [Theory and practice of the category division]. Würzburg, Germany: Würzburger Psychologisches Institut, Lehrstuhl für Allgemeine Psychologie, Forschungsbericht.
- Laurig, W., Luczak, H., & Philipp, U. (1971). Ermittlung der Pulsfrequenzarrhythmie bei körperlicher Arbeit [Determination of the work pulse rate arrhythmia]. *Internationale Zeitschrift für angewandte Physiologie*, 30(1), 40–51.
- Luczak, H. (1975). *Untersuchungen informatorischer Belastung und Beanspruchung*. Düsseldorf, Germany: VDI-Verlag.
- Luczak, H. (1998). *Arbeitswissenschaft* [Industrial engineering and ergonomics] (2nd ed.). Berlin, Germany: Springer.
- Luczak, H., & Laurig, W. (1973). An analysis of heart rate variability. *Ergonomics*, 16(1), 85–97.
- Luczak, H., Philipp, U., & Rohmert, W. (1980). Decomposition of heart-rate variability under the ergonomic aspects of stressor analysis (pp. 123–177). In R.I. Kitney, & O. Rompelman (Eds.), *The study of heart-rate variability*. Oxford, UK: Clarendon Press.
- Luczak, H., & Raschke, F. (1975). Regelungstheoretisches Kreislaufmodell zur Interpretation arbeitsphysiologischer und rhythmologischer Einflüsse auf die Momentanherzfrequenz: Arrhythmie [Theoretical control engineering circulation model for the interpretation of the influence of occupational physiology and rhythmology on the heart rate: arrhythmia]. *Biological Cybernetics*, 18, 1–13.
- Luczak, H., Wiedenmaier, S., Oehme, O., & Schlick, C. (2000). *Augmented reality in design, production and service—Requirements and approach*. In T. Marek & W. Karwowski (Eds.), *Manufacturing agility and hybrid automation III, proceedings of the HAAMAHA 2000* (pp. 15–20). Kraków, Poland: Institute of Management, Jagiellonian University.
- Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung [Psychophysiology of mental workload]. In F. Rösler (Ed.), *Ergebnisse und Anwendungen der Psychophysiologie. Enzyklopädie der Psychologie. Band C/I/5* (pp. 799–864). Göttingen, Germany: Hogrefe.
- Microvision (n.d.). *Information in the form of light* [Homepage]. Retrieved September 15, 2003, from <http://www.mvis.com/>
- Mulder, G. (1979). Sinus arrhythmia and mental workload. In N. Moray (Ed.), *Mental workload—Its theory and measurement* (pp. 327–343). New York, NY, USA: Plenum Press.
- Mulder, G. & Mulder, L.J.M. (1981). Information processing and cardiovascular control. *Psychophysiology*, 18, 392–402.

- Mulder, G., Mulder, L.J.M., Meijman, T.F., Veldman, J.B.P., & van Roon, A.M. (2000). A psycho-physiological approach to working conditions. In R.W. Backs & W. Boucsein (Eds.), *Engineering psychophysiology. Issues and applications* (pp. 139–159). Mahwah, NJ, USA: Erlbaum.
- Mulder, L.J.M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*, *34*, 205–236.
- Nachreiner, F. (2002). Über einige aktuelle Probleme der Erfassung, Messung und Beurteilung der psychischen Belastung und Beanspruchung [On some problems of assessing, measuring and evaluating mental workload]. *Zeitschrift für Arbeitswissenschaft*, *56*(1–2), 11–21.
- Nachreiner, F., & Schütte, M. (2002). Zur Messung psychischer Belastung und Beanspruchung [To the measurement of mental strain and workload]. *Zeitschrift für Arbeitswissenschaft*, *56*(1–2), 1–3.
- Nickel, P., Eilers, K., Seehase, L., & Nachreiner, F. (2002). Zur Reliabilität, Validität, Sensitivität und Diagnostizität von Herzfrequenz- und Herzfrequenzvariabilitätsmaßen als Indikatoren psychischer Beanspruchung [On the reliability, validity, sensitivity and diagnosticity of heart rate and heart rate variability indices for measuring mental work strain]. *Zeitschrift für Arbeitswissenschaft*, *56*(1–2), 22–36.
- Oehme, O., Wiedenmaier, S., Schmidt, L., & Luczak, H. (2001). Empirical studies on an augmented reality user interface for a head based virtual retinal display. In J.M. Smith & G. Salvendy (Eds.), *Systems, social and internationalization design aspects of human-computer interaction. Volume 1 of the Proceedings of the Human Computer Interaction International (HCII) 2001, New Orleans, August 5–10* (pp. 1026–1030). Mahwah, NJ, USA: Erlbaum.
- Oesterreich, R. (2003). Methoden der bedingungsbezogenen psychologischen Arbeitsanalyse [Methods for condition related physiological work analysis]. In H. Strasser, K. Kluth, H. Rausch, & H. Bubb, H. (Eds.), *Quality of work and products in enterprises of the future/Qualität von Arbeit und Produkt im Unternehmen der Zukunft. Proceedings of the 50th Annual Spring Conference of the GfA, Munich, Germany, May 07th–09th, 2003* (pp. 731–732). Stuttgart, Germany: Ergonomia Verlag.
- Pfendler, C., & Schweingruber, J. (2000). Subjektive Beanspruchungsermittlung mit SWORD [Subjective workload measurement with SWORD]. *Zeitschrift für Arbeitswissenschaft*, *54*(1), 11–18.
- Richter, P., & Hacker, W. (1998). *Belastung und Beanspruchung—Stress, Ermüdung und Burnout im Arbeitsleben*. Heidelberg, Germany: Roland Asanger Verlag.
- Schmidtke, H. (2002). Vom Sinn und Unsinn der Messung psychischer Belastung und Beanspruchung [Sense and nonsense of the measurement of mental strain and workload]. *Zeitschrift für Arbeitswissenschaft*, *56*(1–2), 4–9.
- Schütte, M., & Nickel, P. (2002). Multivariate Analyse der bedingungsbezogenen Messgenauigkeit von Beanspruchungsindikatoren [Multivariate analysis of the condition related reliability of strain indicators]. *Zeitschrift für Arbeitswissenschaft*, *56*(1–2), 55–66.
- Wierwille, W.W., & Eggemeier, F.Th. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors*, *35*(2), 263–268.