

PETER NICKEL, KARIN EILERS, LIANE SEEHASE UND FRIEDHELM NACHREINER
ARBEITS- UND ORGANISATIONSPSYCHOLOGIE, CARL VON OSSIETZKY UNIVERSITÄT OLDENBURG

ZUR RELIABILITÄT, VALIDITÄT, SENSITIVITÄT UND DIAGNOSTIZITÄT VON HERZFREQUENZ- UND HERZFREQUENZVARIABILITÄTSMASSEN ALS INDIKATOREN PSYCHISCHER BEANSPRUCHUNG

· *Herzfrequenzvariabilität* · *psychische Beanspruchung* · *Reliabilität* · *Validität* · *Sensitivität* ·
Diagnostizität

ZUSAMMENFASSUNG

Maße der Herzschlagfrequenz oder ihrer Variabilität werden mittlerweile als valide Standard-Indikatoren psychischer Beanspruchung empfohlen und in Feld- und Laborstudien eingesetzt. Dem widersprechen Ergebnisse verschiedener systematischer Evaluationsstudien, wonach einerseits mit den Indikatoren eher unterschiedliche Personen als unterschiedliche Belastungsbedingungen differenziert werden können. Andererseits weisen die Ergebnisse nicht etwa auf aufgabenspezifische Effekte, sondern auf eine Indikation aufgabenunspezifischer Aktivierung (z.B. durch Tageszeiteffekte, Pacing-Effekte) hin. Eine vermeintliche Indikation psychischer oder mentaler Beanspruchung ist auf Konfundierungen mit emotionalen Beanspruchungskomponenten (z.B. durch Zeitdruck, Bedrohung des Scheiterns, Frustration) rückführbar.

PRAKTISCHE RELEVANZ

Zur Untersuchung von Zusammenhängen psychischer Belastung und Beanspruchung in Arbeitssystemen mit dem Ziel der Verhältnisprävention sind solche Indikatoren unzureichend. Es besteht die Gefahr, stark psychisch und insbesondere mental beanspruchende Situationen, in denen keine emotionalen oder Stress-Reaktionen auftreten, als akzeptabel zu beurteilen, während es tatsächlich bereits in normalen Situationen zu psychischer Ermüdung und Fehlern durch mentale Beanspruchung gekommen sein kann, die nicht indiziert wird.

FIABILITÉ, VALIDITÉ, SENSITIVITÉ ET DIAGNOSTIQUE DES MESURES DE LA FRÉQUENCE CARDIAQUE ET DE LA VARIABILITÉ DE LA FRÉQUENCE CARDIAQUE COMME INDICATEUR D'ASTREINTE PSYCHIQUE (CHARGE MENTALE).

· *Fréquence cardiaque* · *variabilité de la fréquence cardiaque*
· *astreinte psychique (charge mentale)* · *fiabilité* · *validité* ·
diagnostique

RESUME

La mesure de la fréquence cardiaque et sa variabilité sont maintenant devenues des indicateurs standard pour mesurer l'astreinte psychique aussi bien sur le terrain qu'en laboratoire. Cependant cette thèse est contredite par les résultats obtenus dans différentes études systématiques : d'un côté avec ces indicateurs différentes personnes ont été différenciées comme étant différentes conditions, et d'un autre côté les résultats n'ont pas démontrés des effets spécifiques à la tâche, mais au contraire une indication une non spécificité par rapport à la tâche (par exemple : influence du moment de la journée). Cette supposée indication d'astreinte psychique ou mentale est liée à une confusion avec les composantes émotionnelles de l'astreinte.

IMPORTANCE PRATIQUE

Pour évaluer le lien entre la contrainte et l'astreinte psychique dans le système homme-machine, avec pour objectif la prévention, de tels indicateurs ne sont pas suffisants. Il y a le danger que des situations engendrant de fortes astreintes psychiques et mentales, dans lesquelles aucune réaction émotionnelle ou stressante ne se présentent, soient jugées acceptables alors qu'en fait, une fatigue psychique ou des erreurs peuvent apparaître dans des situations normales, bien que non détectés.



ON THE RELIABILITY, VALIDITY, SENSITIVITY AND DIAGNOSTICITY OF HEART RATE AND HEART RATE VARIABILITY INDICES FOR MEASURING MENTAL WORK STRAIN

· *heart rate variability* · *mental workload* · *reliability* · *validity* · *sensitivity* · *diagnosticity*

SUMMARY

Mental workload is directly connected with impairing effects of work on the operator and with the development of the operator's capabilities and skills. Mental workload therefore is a relevant criterion to evaluate the quality of work system design. Consequently, legal regulations in the EU require such an evaluation of mental workload which calls for suitable methods for this purpose (Nachreiner 2002).

Due to their practical utility and relative simplicity in both application and subsequent interpretation, when compared with alternative psychophysiological techniques, heart rate (HR) and its derivatives are currently the most often used methods for the assessment of mental workload. HR on the other hand is a parameter, among others, in the complex homeostatic cardiovascular system reflecting energetic, thermic, respiratory, emotional and hypothetically even cognitive processes. HRV measures are thus used in order to be able to decompose these processes which affect the cardiovascular regulation and to separate effects of mental workload. Currently, the 0.1 Hz component of HRV is considered an attractive, valid, and therefore nearly standard indicator of mental strain (Boucsein & Backs 2000; Mulder et al. 2000). However, based on a closer inspection of the relevant literature some severe doubts about the psychometric properties of this and other indicators can be raised (Wilson et al. 1995).

Therefore, the psychometric properties of different HR and HRV measures as indicators of mental strain and/or an indirect measure of mental stress imposed upon the operator by different tasks have been investigated systematically by four experimental studies, with this paper giving an overview over the results of these studies.

Using a generalizability theory approach various HR and HRV measures of ten participants during rest periods and subsequent vigilance task performance for about two hours with two different types of discrimination, three levels of discriminability at six different times of the day, were analyzed in a study by Eilers (1999). According to the results of the analysis of the rest periods these measures showed an acceptable reliability as indicators of person parameters. However, none of these measures showed an effect for type of task (and thus no diagnosticity) or level of difficulty (and thus no sensitivity), but significant effects for work vs. rest, time on task, and time of day only. These results are not in agreement with theoretical assumptions and do not correspond to results of performance or perceived effort data.

Thus there is no support for any validity of these measures for task related mental workload effects.

In the study by Lasner (1997) two signal detection tasks at two different levels of difficulty were presented for 40 minutes each to ten participants. The results for the 0.1 Hz component of HRV did not show effects for time on task, type of task (consistent vs. varied mapping), or level of difficulty, whereas these variations were reflected

in performance measures. Furthermore, an ANCOVA including a base line task as a covariate revealed a complete disappearance of variance for 'level' of workload. Variance in the HRV measure was completely due to interindividual, but not to intersituational (workload or mental stress) differences, thus there was again no evidence for sensitivity or diagnosticity of HRV for mental workload.

14 tasks (from the STRES battery; AGARD 1989) differing in type and level of workload with subsequent rest periods (5 minutes each) were performed by 14 participants in a repeated measurement design in the experiment by Nickel (2001). The ANOVA for the 0.1 component of HRV supported sensitivity for the discrimination between task vs. rest periods, but again showed no support for a more fine grained discrimination between different levels of difficulty within the same type of work task. Furthermore there was no support for any diagnosticity of this measure to discriminate different tasks and thus types of workload. These results were again inconsistent with theoretical assumptions, performance or and perceived difficulty data.

Since it appeared that most of the tasks of this experiment - with one exception - might have been influenced by „pacing“ effects the choice reaction task (machine-paced), the grammatical reasoning task in a self-paced and in a reconstructed machine-paced version (in order to induce different levels of pacing) were presented to ten participants in another experiment. ANOVA for the 0.1 Hz component showed an effect for pacing only but not for type of task, with the two tasks with comparable cognitive demands but different pacing conditions leading to different effects in the HRV measure. However, analyses of the performance data and the perceived difficulties showed the expected effect for different types of tasks.

Therefore, HRV might indicate emotional strain (or stress reactions) or general activation (through time pressure) rather than mental or especially cognitive strain. It is thus concluded that HR and HRV measures are neither valid, nor sensitive, nor diagnostic for the assessment of mental strain. Since these measures do not meet conventional psychometric requirements they should not be used in mental and especially cognitive workload evaluation.

PRACTICAL RELEVANCE

For practical purposes, HR and HRV measures may indicate dysfunctional emotional strain or stress reactions. However, using these measures for evaluating mental workload of a work system may lead to aversive consequences when highly demanding tasks - without producing emotional or stress reactions - are evaluated as acceptable, whereas in fact they may lead to mental fatigue and thus to errors due to cognitive strain. An evaluation of the emotional strain only is thus definitely not sufficient for an ergonomic evaluation of work systems; in fact it might be misleading and dangerous, since errors with severe consequences might otherwise result.



1 EINLEITUNG

Die Beurteilung der psychischen Belastung am Arbeitsplatz ist unter arbeitswissenschaftlichen Gesichtspunkten notwendig, um beeinträchtigende oder schädigende Folgen für Mensch und Sachmittel zu vermeiden, die insbesondere in komplexen Arbeitssystemen durch eine Gefährdung der Systemsicherheit zu noch weiter reichenden Konsequenzen führen können. Konsequenterweise wird daher auch im Rahmen gesetzlicher Regelungen und Verordnungen eine Beurteilung der Belastung eingefordert, ohne dabei jedoch zu definieren, wie diese Forderung operational zu erfüllen wäre (vgl. dazu auch Nachreiner 2002).

Tatsächlich existiert bislang keine Dimension, auf der die in einem Arbeitssystem auftretende psychische Belastung einheitlich abgebildet werden könnte (Schmidtke 1993) - und es erscheint fraglich, ob es eine solche einheitliche Dimension jemals geben kann. Andererseits sollte es möglich sein, wenn die psychische Beanspruchung eine Funktion der psychischen Belastung ist, von der psychischen Beanspruchung auf die psychische Belastung rückzuschließen, sofern sich hierfür geeignete Indikatoren finden ließen. (Auf damit zusammenhängende Defizite vorhandener theoretischer Konzepte zur psychischen Belastung und Beanspruchung oder verwandter Konstrukte kann an dieser Stelle nicht eingegangen werden (vgl. dazu z.B. Gopher & Donchin 1986; Manzey 1998; Nachreiner 2002; Nickel 2001)). Darüber hinaus erscheint aus arbeitswissenschaftlicher Sicht natürlich die Erfassung der psychischen Beanspruchung, also der Auswirkung der Belastung im Menschen, von besonderer Bedeutung. Die Suche nach Indikatoren konnte allerdings bis heute keine Methode zur *direkten* Beanspruchungsmessung hervorbringen, da ein *valides* Außenkriterium für den Grad der psychischen Beanspruchung nicht bekannt ist (Schmidtke 1993). Möglicherweise gibt es aber zumindest Indikatoren, mit deren Hilfe valide und „belastbare“ Aussagen zur psychischen Beanspruchung relativ zu variablen bzw. variierten Belastungsbedingungen ableitbar sind. Bevor solche Indikatoren jedoch im Rahmen einer arbeitswissenschaftlichen Beurteilung der Arbeitsbedingungen eingesetzt werden können - ggf. sogar im Rahmen gesetzlicher oder tariflicher Auseinandersetzungen - ist deren Überprüfung auf Brauchbarkeit, und zwar zunächst im Labor und dann unverzichtbar auch im Feld, zwingend erforderlich. Das scheint allerdings in weiten Bereichen nicht der Stand der Technik zu sein.

In der Literatur werden neben Verhaltensindikatoren, z.B. dem Leistungsverhalten bei ermüdender Tätigkeit, und Erlebensindikatoren, hier vor allem dem subjektiven Beanspruchungsempfinden, insbesondere psychophysiologische Indikatoren diskutiert (z.B. Boucsein & Backs 2000; Caldwell et al. 1994; Manzey 1998; Moray 1988; Hacker & Richter 1984). Letztere vor allem, weil apparativen oder sogenannten ‚naturwissenschaftlichen‘ Messungen von vornherein eine vermeintlich höhere ‚Zuverlässigkeit‘ unterstellt wird und sie eine lange, z.T. allerdings unreflektierte (arbeits-)wissenschaftliche Tradition haben. Ebenso ist davon auszugehen, dass nahezu alle existierenden psychophysiologischen Maße und Messverfahren einschließlich der nahezu täglich neuen Verfahren(svarianten) bereits vielfach - und jetzt verstärkt auch kommerziell - eingesetzt wurden, um damit psychische Beanspruchung - meist unvollständig oder auf der Grundlage äußerst zweifelhafter Konzepte - über den psychophysiologischen Reaktionsteil zu erfassen. Die Aussagen zur psychometrischen Güte solcher Indikatoren sind, falls überhaupt welche mitgeteilt werden, in der Vielzahl der publizierten Untersuchungen ausgesprochen widersprüchlich; plausible Erklärungen sind zahlreich und man gewinnt bei genauerer Betrachtung der einschlägigen Literatur den Eindruck, dass die Zuverlässigkeit und Gültigkeit psychophysiologischer Indikatoren zur Erfassung psychischer Beanspruchung weder durch Labor- noch durch Feldexperimente empirisch eindeutig nachgewiesen ist (vgl. Luczak 1987; Manzey 1998) - insbesondere wenn man die psychometrischen Kriterien und deren Überprüfung streng auslegt.

2 PROBLEMSTELLUNG

Die aus dem Herz-Kreislauf-System ableitbaren Maße werden als die am einfachsten einsetzbaren sowie am häufigsten in Untersuchungen verwendeten beschrieben (vgl. Klimmer 1988; Meshkati et al. 1999) und stehen mittlerweile seit mehr als 100 Jahren an erster Stelle unter den psychophysiologischen Indikatoren psychischer Beanspruchung oder verwandter Konzepte (vgl. z.B. bereits Mosso 1892). Diese Präferenz beruht im wesentlichen auf der Annahme, dass der menschliche Organismus nicht nur in physischen, sondern auch in psychischen Belastungssituationen eine Leistungsanpassung der unterschiedlichen Kreislaufgrößen vornimmt. Es weisen zwar bereits Untersuchungen des 19. Jahrhunderts mit

psychisch beanspruchenden Tätigkeiten auf Veränderungen der Herzschlagfrequenz (HF) hin, sie ist jedoch nur eine einzige Stellgröße in dem aus komplexen homöostatischen Funktionen bestehenden kardiovaskulären System, das auf energetische, thermische, respiratorische, emotionale und/oder affektive Einflüsse u.U. stärker reagiert als auf mental beanspruchende Tätigkeiten. In immer feineren Auflösungen und Modifizierungen wurden Arrhythmie-Maße im Zeit- und Frequenzbereich (z.B. Boucsein & Backs 2000; Jenkins et al. 1983; Opmeer 1973; van Dellen et al. 1985) mit dem Ziel entwickelt, zu einer Dekomposition jener die kardiovaskuläre Regulation beeinflussenden Effekte und damit zur besseren Abbildung kürzer oder länger andauernder psychischer oder mentaler Prozesse zu gelangen (vgl. Eilers 1999; Nickel 2001).

So wird z.B. nach experimentellen Studien von Mulder (1979; zusf. Mulder et al. 2000) davon ausgegangen, dass es begleitend zu Prozessen gesteuerter Informationsverarbeitung einen spezifischen kardiovaskulären Zustand gibt, der sich in der spektralen Dichte im 0,1 Hz Band der HFV manifestiert. Die (interagierende) Wirkungsweise der multiplen Steuerungsmechanismen, die die allgemein zu beobachtenden Veränderungen z.B. im Blutdruck und der Herzschlagrhythmik hervorrufen, ist allerdings bis heute nicht eindeutig nachweisbar (vgl. Obrist 1981; Schmidt & Thews 2000; van Dellen et al. 1985). Dennoch wird aus den verschiedenen HF- und HFV-Maßen aktuell meist die sogenannte 0,1 Hz Komponente der HFV als Standardindikator und mit hinreichender psychometrischer Güte zur Erfassung psychischer oder mentaler Beanspruchung beschrieben (z.B. Boucsein & Backs 2000; Izsó 2001; Luczak 1998; Mulder et al. 2000; Wierwille & Eggemeier 1993; Wiethoff 1997), obwohl einige Untersuchungen durchaus erhebliche und berechtigte Zweifel an der Validität dieses Maßes erkennen lassen (z.B. Apparies et al. 1998; Kamphuis & Frowein 1985; Jorna 1992; Paas et al. 1994; Veltman & Gaillard 1996; Wilson et al. 1995).

Während bei den in Labor- und Feldstudien zur Erfassung psychischer Beanspruchung herangezogenen HF- und HFV-Maßen auf eine Überprüfung der Reliabilität i.d.R. völlig verzichtet wird und diese wegen ihres ‚naturwissenschaftlichen‘ Messansatzes allenfalls als gegeben angenommen wird, beschränkt man sich in den meisten der bislang vorgelegten ‚Validitätsuntersuchungen‘ im wesentlichen darauf, für ein-



zelle oder für mehrere Maße eine Differenzierung zwischen extremen Belastungsunterschieden oder zwischen sogenannter Ruhesituation und (experimentell induzierter) Belastungssituation nachzuweisen, um dann je nach Ausprägung von Unterschieden „per Augenschein“ auf eine mehr oder weniger oder nicht vorhandene Validität zu schließen (Eilers 1999). Eine derartige Differenzierung zeigt streng genommen aber nur an, dass das kardiovaskuläre System auf die Zustandsänderung „Ruhe - Belastung“ reagiert hat, sagt jedoch nicht unbedingt etwas über die Art und Intensität der Folgen psychischer Belastung aus und wäre meist mit einfacheren Messverfahren reliabler abschätzbar (vgl. Schmidtke 1965, 2002). Auf Leistungskennwerte, Einschätzungen des subjektiven Beanspruchungsempfindens und deren Einsatz zur konkurrierenden Validierung bzw. Kontrolle des Zusammenhanges der Wechselwirkungen mit den HF- und HFV-Maßen bei gleichzeitiger Berücksichtigung individueller Anpassungsmöglichkeiten an die jeweilige Belastungssituation (Leistungs- vs. Beanspruchungskompensation; Haider & Rohmert 1978; Künstler 1980) wird in der Regel völlig verzichtet (Eilers 1999). Der Stand der Forschung erweist sich damit nicht nur unter inhaltlicher, sondern auch unter methodischer Perspektive als völlig unbefriedigend und erscheint schon gar nicht geeignet, unter praktischen Arbeitsbedingungen relevante Belastungsformen zu differenzieren, geschweige reliabel, valide, sensitiv und diagnostisch zu erfassen, um daraus „belastbare“ Gestaltungsempfehlungen ableiten zu können.

Aber nur durch valide und kontrolliert angelegte Reliabilitäts- und Validitätsprüfungen kann festgestellt werden, wie genau bzw. wie zuverlässig die Maße das messen, was gemessen werden soll - hier also psychische Beanspruchung und nicht die HF oder HFV an sich (das Problem ließe sich auf einer anderen (Integrations-)Ebene, der Definition des Anstiegs, Wendepunkts ebenfalls stellen, interessiert hier zur Überprüfung der psychometrischen Güte von *Beanspruchungsmaßen* aber nicht) - und ob und inwieweit es sich bei dem Gemessenen auch tatsächlich um eine Indikation psychischer oder mentaler Beanspruchung handelt. Dazu sind dann auch Ergebnisse zur Sensitivität und Diagnostizität der Maße vorzulegen, um auch Informationen darüber zu erhalten, welcher Intensitätsbereich wie fein aufgelöst differenziert werden kann und inwieweit solche Aussagen auf alle oder nur einzelne Dimensionen des Konstruktes (u.U. differenziell unterschiedlich) gültig sind.

Aus diesem Grunde haben wir uns in mehreren Studien systematisch und unter kontrollierten Experimentalbedingungen mit der Frage der Reliabilität, Validität, Sensitivität und Diagnostizität von HF- und HFV-Maßen beschäftigt, über deren Ergebnisse hier zusammenfassend berichtet wird. Eine detaillierte Darstellung der einzelnen Studien findet sich in den jeweils angegebenen Quellen.

3 EXPERIMENTELLE STUDIEN ZUR PSYCHOMETRISCHEN GÜTE VON HF- UND HFV-MASSEN ZUR ERFASSUNG PSYCHISCHER BEANSPRUCHUNG

3.1 Stabilität unter Ruhebedingungen

Für eine Überprüfung der Reliabilität und Validität von HF- und HFV-Maßen wurde auf Daten aus Daueraufmerksamkeitsversuchen zurückgegriffen, die im Rahmen eines von der Deutschen Forschungsgemeinschaft geförderten Projektes über „Auswirkungen von Tageszeit und Belastungsdauer auf Vigilanzleistungen bei unterschiedlicher Aufgabenschwierigkeit“ erhoben wurden. Der Versuchsplan war so angelegt, dass Mindestanforderungen an Indikatoren psychischer Beanspruchung für eine Reliabilitäts- und Validitätsbestimmung, d.h. die Differenzierung von Arten und Schwierigkeitsstufen der Informationsverarbeitung und deren Interaktion, zusammen mit Effekten der Versuchsdauer und der circadianen Lage, überprüft werden konnten. Eine ausführliche Darstellung dieser Untersuchungen findet sich bei Eilers (1999).

Bei der Zuverlässigkeit von HF- und HFV-Maßen interessiert die Frage nach der *intra*-individuellen Reliabilität mehr als die nach der *inter*-individuellen. Sinnvolle Aussagen über die intraindividuelle (Langzeit- und Kurzzeit-)Stabilität von HF- und HFV-Maßen lassen sich nur dann gewinnen, wenn für jede Versuchsperson (Vp) deutlich mehr als zwei Messungen unter sonst vergleichbaren Bedingungen vorgenommen werden. Von praktischer Bedeutung sind solche Messungen aber nur dann, wenn sie sich auch über einen längeren Zeitraum hinweg als zuverlässig erweisen (Hecker et al. 1980). Um mögliche circadiane Einflüsse auf die Stabilität von HF- und HFV-Maßen kontrollieren zu können, sollten nicht nur zu einer Tageszeit, sondern zu unterschiedlichen Tageszeiten jeweils mehrere Messungen im

zeitlichen Abstand von mehreren Tagen erfolgen. Mit dieser Versuchsanlage sind nicht nur die Voraussetzungen für eine Reliabilitätsbestimmung i.S.e. *intra*-individuellen *Langzeit-Stabilität* geschaffen, sondern auch die für eine *situations-* d.h. hier *tageszeitbezogene*. Bei einer *tageszeitbezogenen* Reliabilitätsbestimmung wird die Vp als Störfaktor angesehen und der Bedingung Tageszeit als Merkmalsträger begründbar Belastungswirksamkeit unterstellt, da die psychische Beanspruchung aufgrund einer geringeren physischen Aktivierung während der Nachtstunden höher sein dürfte als am Tag (vgl. Folkard 1996). Aussagen über die *Kurzzeit-Stabilität* von HF- und HFV-Maßen lassen sich aus der Erhebung mehrerer direkt aufeinanderfolgender Messungen zu ein und demselben Messzeitpunkt unter sonst gleichen Bedingungen ableiten.

An der Untersuchung nahmen 10 schichtarbeitserfahrene Versuchspersonen teil. Diese führten in einem Abstand von jeweils 20 Stunden zwischen den einzelnen Versuchssitzungen um 20, 16, 12, 08, 04 und 00 Uhr jeweils einen 2-stündigen Versuch mit einer visuellen Daueraufmerksamkeitsaufgabe vom Typ „perceptual speed“ (Speed- bzw. Sukzessivdiskriminations-Aufgabe) oder „flexibility of closure“ (Closure- bzw. Simultandiskriminations-Aufgabe) durch (vgl. hierzu auch Matthews et al. 1993), die sich in der Art ihrer Diskriminationsanforderungen (und damit der Art ihrer psychischen Belastung) unterschieden. Jeder Aufgabentyp war jeweils in drei unterschiedlichen Schwierigkeitsgraden zu bearbeiten, wobei die Vpn davon nicht in Kenntnis gesetzt wurden. Bei den Aufgaben handelt es sich um die von uns speziell für Vigilanzversuche entwickelten Diskriminationsaufgaben (vgl. Eilers & Nachreiner 1990; Hänecke 1995), die sich durch eine hohe Differenzierbarkeit der Diskriminationsanforderungen (und damit der psychischen Belastung) nach Art und Intensität auszeichnen.

Die sich aus dem 2 x 3 x 6 (Aufgabentyp x Schwierigkeit x Tageszeit) within-Design ergebenden 36 Einzelversuche wurden von den Vpn in randomisierter Reihenfolge von Aufgabentyp und Schwierigkeit im Abstand von jeweils 20 Stunden durchgeführt. Den einzelnen Versuchssitzungen war jeweils eine Ruhe- und Vorbereitungszeit von einer Stunde vorangestellt, die u.a. der Ermittlung einer 10-minütigen HFV-Ruhephase diente. Während der 10-minütigen Ruhephasen und der sich anschließenden Hauptversuche wurden die Herzschläge als Echtzeitergebnis-



Tabelle 1: Personen- und situationsbezogene Generalisierbarkeits-Koeffizienten (GK) ausgewählter HF- und HFV-Maße (nach Eilers 1999; Messwiederholung (M), Tageszeit (T), Messperiode (P), Versuchsperson (V), Erläuterung zu HF- und HFV-Maßen im Text)

HF- und HFV-Maße →		HF	HPD \bar{x}	HPD sd	HPD vk	RMSSD	CVS	N	S	ARQ	Hz Bd	0,03 Hz Bd	0,1 Hz Bd	0,3 Hz Bd
↓ Bezug des GK														
Person	Ep ² _v 6M 6T 2P	,97	,97	,98	,97	,98	,98	,97	,98	,97	,96	,92	,94	,96
Situation	Ep ² _t 10V 6M 2P	,86	,82	,78	,66	,75	,63	,00	,77	,00	,39	,32	,00	,57

reihe im Rechner gespeichert. Darüber hinaus hatten die Vpn nach Ende eines jeden Versuches die Anstrengungsskala (Eilers et al. 1986, 1989, 1990; Eilers 1999) und die BLV-Modifikation (Schütte 1988) zur Indikation des subjektiven Beanspruchungsempfindens zu beantworten.

Auf der Grundlage der den 36 Einzelversuchen vorangestellten 10-minütigen Ruhephasen erfolgten die Analysen zur Überprüfung der Reliabilität der dabei erhobenen HF- und HFV-Maße. Jede der Ruhephasen wurde dazu - ebenso wie die sich anschließenden 2-stündigen Einzelversuche - in 5-Minuten-Intervalle unterteilt. Für eine methodische Umsetzung bot sich die Generalisierbarkeitstheorie (vgl. Brennan 1992; Cronbach et al. 1972; Marcoulides 2000; Nußbaum 1987; Fisseni 1990; sowie auch Schütte & Nickel 2002) geradezu an, da sie die Möglichkeit bietet, sowohl eine intraindividuelle (Generalisierung über Kennwerte von Individuen als individuelles Merkmal) als auch eine situationsbezogene Reliabilitätsbestimmung (Generalisierung über Situationen - hier über Tageszeiten) unter kontrollierten Bedingungen vorzunehmen.

In der Tabelle 1 sind die personen- und die situations-, hier tageszeitbezogenen, Generalisierbarkeits-Koeffizienten (GK) für den hier untersuchten Versuchsplan mit 6 Messwiederholungen (2 Aufgabentypen x 3 Schwierigkeitsgrade = 6M), 6 Tageszeiten (6T), 2 Messperioden (2P) und 10 Versuchspersonen (10V) dargestellt. Eine hinreichende Generalisierbarkeit wird bei einem GK > 0,80 angenommen (Llabre et al. 1988). Wie die Ergebnisse der Analysen belegen, erhält man für sämtliche HF- und HFV-Maße personenbezogene GK, die weit über 0,90 liegen. Diese sind damit für alle HF- und HFV-Maße weitaus höher als die der tageszeitbezogenen GK, die lediglich für die beiden HF-Maße - HF (Herzschlagfrequenz) und HPD \bar{x} (mittlere Herzperiodendauer) -

einen GK > 0,80 erreichen, während alle HFV-Maße darunter liegen. Während für HPDsd (Standardabweichung der Herzperiodendauer), S (Summe der absoluten Differenzen aufeinanderfolgender Herzperiodendauern) und RMSSD (Quadratwurzel aufeinanderfolgender Herzperiodendifferenzen) die tageszeitbezogenen GK nur knapp unter 0,80 liegen, werden für die übrigen HFV-Maße Werte erzielt, die sich zwischen < 0,66 für HPDvk (Variationskoeffizient der Herzperiodendauer) und null für die 0,1 Hz Komponente der HFV, den ARQ (Arrhythmiequotienten) und N (Anzahl der Fluktuationen innerhalb eines Herzperiodenintervalls) bewegen (eine vollständige und ausführlichere Beschreibung der verwendeten Variabilitätsmaße findet man bei Eilers 1999). Da bei den drei letztgenannten HFV-Maßen die universale Varianz (hier die Tageszeitvarianz) bei null liegt, d.h. es lassen sich für diese Maße keine systematischen tagesperiodischen Unterschiede nachweisen, sind folglich auch die tageszeitbezogenen GK gleich null. Dies steht jedoch im Widerspruch zu den Temperaturverläufen der Vpn (hier nicht dargestellt), die auch trotz der Rückwärtsrotation über den gesamten Versuch hinweg einen stabilen circadianen Rhythmus zeigen.

Insgesamt betrachtet belegen die Ergebnisse der Reliabilitätsbestimmung eindeutig, dass es sich bei den ausgewählten HF- und HFV-Maßen um stabile Personenparameter handelt, die nur eine relativ geringe intraindividuelle Instabilität aufweisen. Dabei ist es ohne Bedeutung, ob die Messungen zu einer bestimmten Tageszeit vorgenommen werden oder ob zu unterschiedlichen Tageszeiten jeweils eine Messung erhoben wird. Das bedeutet, dass sowohl für die Belastungsbedingung Messwiederholungen (M) als auch für die Belastungsbedingung Tageszeit (T) eine hohe personenbezogene Langzeit-Stabilität vorliegt (vgl. hierzu Eilers 1999). Eine akzeptable bedingungsbezogene Stabilität lässt sich dagegen aus-

schließlich für die beiden HF-Maße sowie für HPDsd, RMSSD und S nachweisen. Die mit den Ergebnissen dieser Studie nachgewiesene hohe personenbezogene und z.T. auch tageszeitbezogene Generalisierbarkeit bzw. Langzeit-Stabilität der HF- und HFV-Maße gilt jedoch ausschließlich für Ruhebedingungen bzw. für mehr oder weniger homogene Situationen. Circadiane Effekte - von denen aufgrund der vorliegenden Erkenntnisse zur Circadianperiodik auszugehen war (Lavie 2001) - werden allerdings nur von einigen Maßen generalisierbar reflektiert, obwohl alle der hier verwendeten Variabilitätsmaße auf derselben HF-Datenbasis berechnet wurden. Das könnte für die beiden HF-Maße sowie für HPDsd, RMSSD und S als potentieller Hinweis auf eine positive Validitätsbeurteilung interpretiert werden, während dies für die anderen Maße schon nicht mehr zu erwarten ist.

3.2 Stabilität unter Belastungsbedingungen

Aus diesem Grunde war als nächstes der Frage nachzugehen, wie sich HF- und HFV-Maße in experimentell manipulierten, beanspruchungsbezogenen heterogenen Situationen verhalten, d.h., wenn die Vpn nicht Ruhesituationen zu unterschiedlichen Tageszeiten, sondern unterschiedlichen Belastungsbedingungen ausgesetzt sind, wie sie z.B. aus der Bearbeitung unterschiedlicher Aufgabentypen mit unterschiedlichen Schwierigkeitsgraden zu unterschiedlichen Tageszeiten resultieren. Für diese Validitätsprüfung wurden zur Berücksichtigung des zeitlichen Verlaufs von HF-, HFV-Maßen und der Leistungskennwerte (Trefferquote und Rate der falschen Alarme) die 2-stündigen Einzelversuche jeweils in 24 5-Minuten-Intervalle unterteilt. Für die weiteren Analysen stand somit ein 2 x 3 x 6 x 24 (Aufgabentyp x Schwierigkeit x Tageszeit x Belastungsdauer) within-Versuchsdesign zur Verfügung (vgl. Eilers 1999).

Für die methodische Umsetzung der Studie bot sich auch hier die Generalisierbarkeitstheorie an. Da es sich bei dieser nicht um ein reines Reliabilitätskonzept handelt, weil Aspekte der Reliabilität und Validität miteinander vereinigt werden, sind damit auch Aussagen zur Konstruktvalidität möglich (Nußbaum 1987). Dabei geben die aus der Generalisierbarkeitsstudie resultierenden Schätzwerte für die Varianzkomponenten Aufschluss darüber, wie homogen bzw. heterogen das zugrundeliegende Konstrukt „psychische Beanspruchung“ - hier bzgl. der



verwendeten Indikatoren HF- und HFV-Maße - hinsichtlich der verschiedenen Belastungsbedingungen ist. Voraussetzung für eine positive Validitätsbeurteilung von HF- und HFV-Maßen ist, dass die Varianzanteile, die auf die beanspruchungsrelevanten personenunabhängigen Varianzkomponenten Aufgabentyp (A), Schwierigkeit (S), Tageszeit (T) und Belastungsdauer (D) sowie deren Interaktionen zurückzuführen sind, im Vergleich zu den übrigen Varianzkomponenten relativ hoch ausfallen und somit als heterogen im Sinne der experimentellen Manipulation zu bezeichnen sind. Die auf dieser Basis ermittelten situations- bzw. belastungsbezogenen GK geben dann an, wie stabil die Messungen der unterschiedlichen HF- und HFV-Maße unter den jeweiligen Belastungsbedingungen Aufgabentyp, Schwierigkeit, Tageszeit und Belastungsdauer sowie deren Interaktionen sind.

Zur Überprüfung der Validität der HF- und HFV-Maße wurden nicht nur Absolut- oder Verlaufswerte (Vw), die die Höhe des jeweiligen physiologischen Leistungsaufwands als Folge einer bestimmten Belastungsbedingung zeigen sollen, sondern auch Differenz- bzw. Reaktionswerte (Rw), die den Vorgang der Zustandsänderung durch Differenzbildung von Verlaufswert und Ruhe- bzw. Ausgangswert (d.h. personenbezogene standardisierte Werte) berücksichtigen, analysiert.

Nach den Ergebnissen der Varianzkomponentenschätzungen sind die Anteile der *personenunabhängigen Bedingungsvarianz*, die auf die Belastungsbedingungen Aufgabentyp, Schwierigkeit, Tageszeit und Belastungsdauer sowie auf deren personenunabhängige Interaktionen entfallen sowohl für die Verlaufs- als auch für die Reaktionswerte insgesamt eher gering (vgl. Eilers 1999). Diese Varianzanteile variieren jedoch in Abhängigkeit vom Variabilitätsmaß sehr stark. Bei den *Verlaufswerten* variieren die

Summenwerte der personenunabhängigen Bedingungsvarianz zwischen 2,7 % für den ARQ bis zu 17,6 % für die HPDsd und bei den *Reaktionswerten* zwischen 2,2 % für das obere Frequenzband der HFV und 20,9 % für die HPDsd. Betrachtet man nun die Verteilung der Varianzen auf die unterschiedlichen Belastungsbedingungen und deren Interaktionen, so ist bei den *Verlaufswerten* festzustellen, dass fast die gesamte Bedingungsvarianz auf die Tageszeit und auf die Belastungsdauer entfallen. Dagegen entfallen nur sehr geringe Varianzanteile auf die Interaktionen TxD, AxSxT und AxSxTxD. Auf die unterschiedlichen Aufgabentypen, Schwierigkeitsgrade sowie deren Interaktion entfällt trotz ihrer eindeutigen Beeinflussung der Leistungsergebnisse (z.B. Rate der Treffer und Falschen Alarme) nahezu keine Varianz. Vergleichbare Ergebnisse zeigen die *Reaktionswerte*. Im Vergleich zu den Verlaufswerten ist hier bei nahezu allen HF- und HFV-Maßen eine Zunahme der Varianz, die auf die Belastungsdauer entfällt, zu verzeichnen, während die Tageszeit-Varianz sich stark verringert oder sogar gänzlich verschwindet. Letzteres ist darauf zurückzuführen, dass die Ruhewerte einiger Variabilitätsmaße einen mehr oder weniger stark ausgeprägten Tagesgang aufweisen (vgl. hierzu 3.1) und durch eine Subtraktion der jeweiligen Ruhewerte vom Verlaufswert die Tageszeit-Varianz bei acht von dreizehn Maßen ganz, bei den übrigen teilweise verschwindet (vgl. Eilers 1999).

Im Vergleich zur personenunabhängigen Bedingungsvarianz fällt allerdings die *interindividuelle Versuchspersonenvarianz* relativ hoch aus. Sie variiert in Abhängigkeit vom HFV-Maß bei den *Verlaufswerten* zwischen 16,9 % für HF und 66,8 % für N und bei den *Reaktionswerten* zwischen 4,2 % für HF und 24,8 % für HPD \bar{x} . Die geringere interindividuelle Versuchspersonenvarianz bei den Reaktionswerten ist darauf zurückzuführen, dass durch die Subtraktion des Ruhewertes vom Verlaufswert bereits ein

großer Teil der personenspezifischen Varianz eliminiert wurde.

In der Tabelle 2 werden die auf der Basis der Varianzkomponentenschätzungen ermittelten personen- und belastungsbezogenen GK für einige exemplarisch ausgewählte HF- und HFV-Maße und zwar für HPD \bar{x} , die HPDsd und für die 0,1 Hz Komponente der HFV (0,1 HFV) dargestellt. Betrachtet man zunächst die *personenbezogenen GK* (Tabelle 2: Spalte V), so findet man ausnahmslos - und zwar sowohl für die Verlaufs- (Zeilen Vw) als auch für die Reaktionswerte (Zeilen Rw) - sehr hohe GK. Dies bedeutet, dass die Vpn sich nicht nur - wie oben beschrieben - stark voneinander unterscheiden, sondern, dass die Messungen der HF- und der HFV-Maße innerhalb einer Versuchsperson *trotz unterschiedlicher Belastungsbedingungen als stabil* zu bezeichnen sind.

Anders sieht es bei den *belastungsbezogenen GK* aus. Wie die Ergebnisse sehr deutlich belegen, liegen fast alle belastungsbezogenen GK der ausgewählten Maße bei null bzw. $\ll 0,80$. Eine Ausnahme bilden lediglich die tageszeit- und die belastungsdauerbezogenen GK. Betrachtet man zunächst die tageszeitbezogenen GK (Tabelle 2: Spalte T), so finden wir für die Verlaufswerte von HPD \bar{x} und HPDsd Generalisierbarkeitskoeffizienten, die weit über 0,80 liegen, während für die Verlaufswerte der 0,1 Hz Komponente der HFV lediglich ein Wert von 0,74 erreicht wird. Geringere GK zeigen sich für die Reaktionswerte. Während für HPD \bar{x} und die 0,1 Hz Komponente der HFV Werte erzielt werden, die weit unter 0,80 bzw. bei null liegen, erhält man für HPDsd mit 0,78 einen tageszeitbezogenen GK, der nur knapp unter der Grenze von 0,80 liegt. Eindeutigere Ergebnisse zeigen sich jedoch für die belastungsdauerbezogenen GK (Tabelle 2: Spalte D). Hier ergeben sich für *alle* Variabilitätsmaße - sowohl für die Verlaufs- als auch für die Reaktionswerte - hohe GK, die weit über 0,80 liegen.

Tabelle 2:

Personen- (V) und belastungsbezogene (A, S, T, D, und Interaktionen) Generalisierbarkeitskoeffizienten (GK) der Verlaufswerte (Vw) und Ruhewerte (Rw) für einige exemplarisch ausgewählte HF- und HFV-Maße (s.a. Erläuterungen im Text)

		Verlaufswerte (V _w) und Reaktionswerte (R _w) [E ρ^2]															
		A	S	T	D	AS	AT	ST	AD	SD	TD	AST	ASD	ATD	STD	ASTD	V
HPD \bar{x}	V _w	,0	,0	,88	,98	,0	,0	,0	,0	,0	,0	,0	,0	,0	,0	,09	,97
	R _w	,25	,0	,0	,98	,0	,0	,0	,43	,0	,19	,0	,0	,0	,0	,04	,93
HPD _{sd}	V _w	,0	,0	,90	,97	,0	,0	,0	,0	,0	,43	,37	,0	,0	,0	,14	,97
	R _w	,0	,0	,78	,97	,0	,0	,0	,0	,0	,41	,29	,0	,0	,0	,13	,80
0,1 HFV	V _w	,0	,0	,74	,90	,0	,0	,0	,0	,0	,46	,43	,0	,0	,0	,0	,95
	R _w	,0	,0	,48	,90	,29	,48	,0	,0	,0	,45	,0	,0	,0	,0	,0	,88

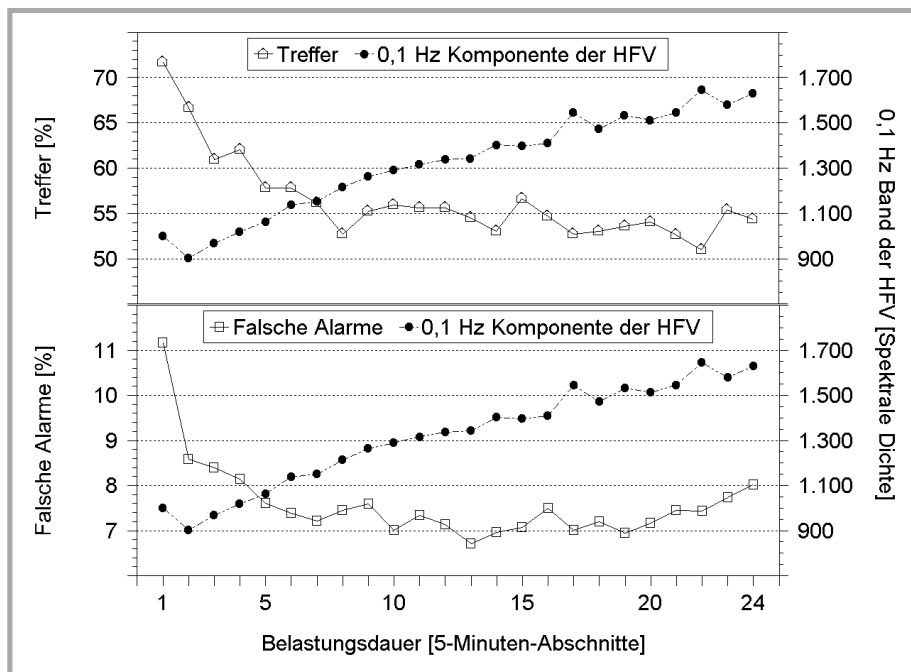


Bild 1: Prozentwerte der Treffer und der Falschen Alarme jeweils mit Ausprägungen der 0,1 Hz Komponente der HFV im Verlauf über die Belastungsdauer (vgl. Eilers 1999)

Differenzierte, nichtadditive Belastungswirkungen, wie sie sich etwa in Interaktionen zeigen sollten, sind nicht nachweisbar, allenfalls deuten sich Interaktionen mit der Belastungsdauer und/oder Tageszeit an, allerdings alle mit völlig unzureichenden GK.

Betrachtet man den Verlauf der 0,1 Hz Komponente der HFV über die Belastungsdauer hinweg (Bild 1), so wird deutlich, dass mit zunehmender Belastungsdauer die spektrale Dichte kontinuierlich zunimmt, wobei nach ca. 30 Minuten sogar die zuvor in der Ruhephase erreichten Werte überschritten werden. Theoretisch ist jedoch bei einem Ansteigen der Ausprägungen der 0,1 Hz Komponente der HFV über die Belastungsdauer hinweg davon auszugehen, dass der physiologische Leistungsaufwand und damit die psychische Beanspruchung geringer werden. Dies müsste einen entsprechenden Niederschlag in der Leistung finden. Für die Trefferrate und die Rate der Falschen Alarme (vgl. auch Bild 1) lässt sich jeweils ein signifikanter Haupteffekt für die Belastungsdauer nachweisen. Während die Trefferrate innerhalb der ersten 40 Minuten stark absinkt, im weiteren Verlauf jedoch konstant bleibt, findet man bei der Rate der Falschen Alarme ein starkes Absinken lediglich innerhalb der ersten 25 Minuten. Dann sinkt sie nur noch leicht ab, bevor sie während der letzten 30 Minuten wieder geringfügig ansteigt. Damit zeigt sich (nach ca. 30 Minuten) ein kontinuierlicher Anstieg der 0,1 Hz Komponente der HFV über die

Belastungsdauer bei nahezu gleichbleibender Leistung. Würde man nun annehmen, dass dieses HFV-Maß tatsächlich psychische Beanspruchung oder deren Folgen misst, dann wäre sein Anstieg über die Belastungsdauer hinweg als Abnahme psychischer Beanspruchung zu interpretieren, was bei gleichbleibender Aufgabenschwierigkeit zum Absinken der Leistung führen müsste. Da dies jedoch definitiv nicht der Fall ist, ist eher zu vermuten, dass die hier wegen der vergleichbaren Ergebnisse für die anderen HF- und HFV-Maße stellvertretend dargestellten Ausprägungen der 0,1 Hz Komponente der HFV eben nicht psychische Beanspruchung oder deren Folgen, sondern etwas ganz anderes erfassen: Dabei könnte es sich z.B. um Aktivierung handeln, die ja gelegentlich mit psychischer Beanspruchung einhergehen kann - aber nicht muss.

Diese Annahme wird unterstützt, wenn man die wiederum vergleichbaren Verläufe der HF- und der HFV-Maße über die Tageszeit hinweg betrachtet (vgl. Eilers 1999). Dabei fällt auf, dass die Variabilität während der ungünstigen Nachtbedingungen und morgens am höchsten und während der Nachmittagsstunden am geringsten ist. Danach wäre die psychische Beanspruchung während der Nacht und morgens am geringsten und nachmittags am höchsten. Auch hier zeigt sich wieder ein deutlicher Widerspruch zu den Ergebnissen der Leistung und des subjektiven Beanspruchungsempfindens (vgl. Eilers et al. 1989, 1990; Eilers 1999), dem darüber hinaus Erkenntnisse aus der

Circadian-Rhythmik entgegenstehen, so dass diese Verläufe der HF- und der HFV-Maße eher Ausdruck tagesperiodischer Schwankungen der allgemeinen Aktivierung als tageszeitabhängige, unterschiedliche Beanspruchung(sfolgen) sind, die sich jedoch unter der Bedingung Arbeit deutlicher zeigen als unter der Bedingung Ruhe, da die -aufgabenunspecifische - Aktivierung durch die Arbeit (vgl. dazu die Ergebnisse von Nickel (2001)) für eine gewisse Basisstabilisierung und Unterdrückung des Rauschens verantwortlich zu sein scheint.

Fasst man die Ergebnisse zusammen, so ist festzustellen, dass sich für die ausgewählten HF- und HFV-Maße durchweg eine hohe personenbezogene und für einige der Maße auch eine akzeptable situationsbezogene Reliabilität nachweisen lässt. Ihre Validität ist jedoch insgesamt negativ zu beurteilen: So lässt sich für alle HF- und HFV-Maße zwar eine hohe personenbezogene Generalisierbarkeit belegen, d.h. als Indikator der einzelnen V_p , die hier unter unterschiedlichen Belastungsbedingungen arbeitete, sind die HF- und HFV-Maße stabil und erlauben damit eine zuverlässige Differenzierung zwischen einzelnen Personen. Eine belastungsbezogene Generalisierbarkeit liegt entgegen aller Erwartungen jedoch nicht vor. Weder für die unterschiedlichen Aufgabentypen noch für die unterschiedlichen Schwierigkeitsgrade und/oder ihre Interaktionen lässt sich auf der Grundlage vorliegender Ergebnisse eine positive Validitätsbeurteilung ableiten. Lediglich für die Tageszeit und für die Belastungsdauer und damit aufgabenunspecifische Belastungsbedingungen zeigt sich eine signifikante belastungsbezogene Generalisierbarkeit. Dabei ist jedoch davon auszugehen, dass es sich bei diesen um Schwankungen der allgemeinen Aktivierung und nicht um solche der psychischen Beanspruchung handelt.

3.3 Automatisierte vs. gesteuerte Informationsverarbeitung

In Feldstudien im Öffentlichen Personennahverkehr (vgl. Nachreiner et al. 1999) war es mit Hilfe der 0,1 Hz Komponente der HFV nicht möglich, die aus - nach Expertenbeobachtungen - sich deutlich unterscheidenden Fahrsituationen resultierende unterschiedliche psychische Beanspruchung zu differenzieren. Dabei wurden analysetechnische und methodische Probleme des Indikators 0,1 Hz Komponente der HFV identifiziert, die sich sowohl in sehr fein aufgelösten Verläufen als auch in über



Belastungsabschnitte integrierten Werten zeigten (vgl. Nachreiner et al. 1998, 1999; Nickel et al. 1998). Es wurde daher als notwendig erachtet, sich mit diesen Problemen auch bezogen auf die beiden unterschiedlichen Auflösungsstufen des Indikators zunächst noch einmal experimentell auseinanderzusetzen. Dazu wurde in einer laborexperimentellen Studie von Lasner (1997) untersucht, inwieweit sich insbesondere mit der 0,1 Hz Komponente der HFV die nach Art und Intensität der Belastung variierten Bedingungen resultierende psychische Beanspruchung differenzieren lässt. Die Bearbeitung dieser Fragestellung erschien auch deshalb bedeutsam, weil die Diskriminationsfähigkeit der 0,1 Hz Komponente der HFV für automatisierte vs. gesteuerte Informationsverarbeitung in der Literatur immer wieder behauptet und darüber ihre Eignung als Indikator zur Erfassung kognitiver Beanspruchung postuliert wird (z.B. Wierwille & Eggemeier 1993). In der letzten Zeit wird dieses Maß daher auch vermehrt zur Evaluation von praktischen Arbeitssituationen eingesetzt (z.B. Izsó 2001; Wiethoff 1997), ohne dass jedoch die psychometrischen Kriterien für ein solches Vorgehen belegt wären oder würden.

Die Analysen der sehr fein aufgelösten Verläufe der Ausprägungen der 0,1 Hz Komponente der HFV (vgl. Lasner 1997) führten allerdings wiederum zu einer - wie bereits von Nickel et al. (1998) auf der Basis von Feldstudien festgestellten - negativen Beurteilung der psychometrischen Güte. Aus den Analyseergebnissen für die über Belastungsabschnitte integrierten Indikatoraus-

prägungen lassen sich allerdings Überlegungsansätze ableiten, die auch für die Verwendung von HF- und HFV-Maßen insgesamt von großer Bedeutung sind und über die daher hier zusammenfassend berichtet wird.

In der Studie von Lasner (1997) wurde zur Differenzierung unterschiedlicher Arten der Belastung auf die bereits beschriebenen Speed- und Closure-Aufgaben zurückgegriffen. Mit einem höheren Anteil kritischer Signale (20%) als unter Vigilanzbedingungen üblich, konnte mit den hier leicht modifizierten Aufgaben eine deutliche Differenzierung zwischen automatisierter und gesteuerter Informationsverarbeitung erreicht werden. Neben dieser unterschiedlichen Art der Belastung (durch speed-Aufgaben mit konsistenter vs. closure-Aufgaben mit variiertem Zuordnung von Reiz und erforderlicher Reaktion) wurde auch jeweils die Intensität der Belastung (leichte vs. deutlich schwierigere Diskriminationsanforderungen) variiert. Zur Differenzierung unterschiedlicher Belastungsdauer wurden eine nachträglich in vier 10-Minuten-Abschnitte unterteilte Bearbeitungsdauer von 40 Minuten realisiert. Somit war es über die variierten Belastungsbedingungen möglich, erste Aussagen zur Sensitivität als auch zur Diagnostizität der 0,1 Hz Komponente der HFV unter einem Signalentdeckungsparadigma abzuleiten.

In die Analysen der Untersuchung konnten Daten von 12 Vpn einbezogen werden. Die Vpn wurden per Zufall zwei Experimentalgruppen (leichte vs. schwierige Aufgaben-

arten) zugeteilt. Im Anschluss an drei Trainingsabschnitte wurde jeweils die gleiche Referenzaufgabe (eine einfache Reaktionszeitaufgabe) gefolgt von einmal Speed- und dann Closure-Aufgaben (bzw. umgekehrt; balanciert für Gruppen) bearbeitet. Über die gesamte Versuchsdauer wurden die Herzperiodendauer und die Leistungsdaten kontinuierlich registriert und anschließend für die weiteren Analysen als Kennwerte (0,1 Hz Komponente der HFV, Antwortzeiten, Treffer, Falsche Alarmer, Diskriminationsvermögen, Entscheidungskriterium) aufbereitet (vgl. Lasner 1997).

Die Ergebnisse in Bild 2 deuten zwar sowohl für das Diskriminationsvermögen (d') als auch für die 0,1 Hz Komponente der HFV zunächst auf eine Differenzierung der Belastungsbedingungen (z.B. helle Balken mit/ohne Muster für ‚leicht‘ vs. dunkle Balken mit/ohne Muster für ‚schwer‘) hin, allerdings treten bei genauerer Betrachtung deutliche Widersprüche in den HFV- und zu den Leistungsergebnissen zu Tage. Während das Diskriminationsvermögen (d' -Werte; rechte y-Achse) für die leichten Aufgaben größer als für die schweren Aufgaben (gefüllte vs. leere Symbole) und für die Speedgrößer als für die Closure-Aufgabe (Kreis vs. Dreieck) ist, liegen die d' -Werte für die leichte Variante der Speed-Aufgabe deutlich über denen der übrigen Varianten. Dieses Ergebnis spiegelt sich auch in den varianzanalytischen Auswertungen mit einem hochsignifikantem F-Wert für den Interaktionseffekt für ‚Level x Aufgabe‘ wider. Somit stützen die Ergebnisse die Operationalisierung der zwei unterschiedlichen Schwierigkeitsgrade durch die Variation der Aufgabenparameter, d.h. leichter und schwerer Aufgabenlevel desselben Aufgabentyps unterscheiden sich in erwarteter Richtung. Allerdings ist dabei davon auszugehen, dass es sich bei leicht vs. schwer um differenzielle Schwierigkeitsgrade bei unterschiedlichen Aufgabentypen handelt, d.h. dass z.B. die schwere Speed und die schwere Closure *nicht* den gleichen Schwierigkeitsgrad darstellen. Das lässt sich relativ einfach am unterschiedlichen d' ablesen, aber selbst ein identisches d' wäre keine Garantie für eine gleiche Schwierigkeit oder Anforderungshöhe, da eben unterschiedliche Funktionen gefordert werden (Simultan vs. Sukzessivdiskrimination) und die Schwierigkeitsunterschiede damit auf unterschiedlichen Dimensionen liegen könnten, die sich nur im Leistungsmaß, nicht aber in den Aufwandsmaßen integriert abbilden. Insofern sind signifikante Interaktionswerte hier durchaus akzeptabel. Die Ergebnisse für

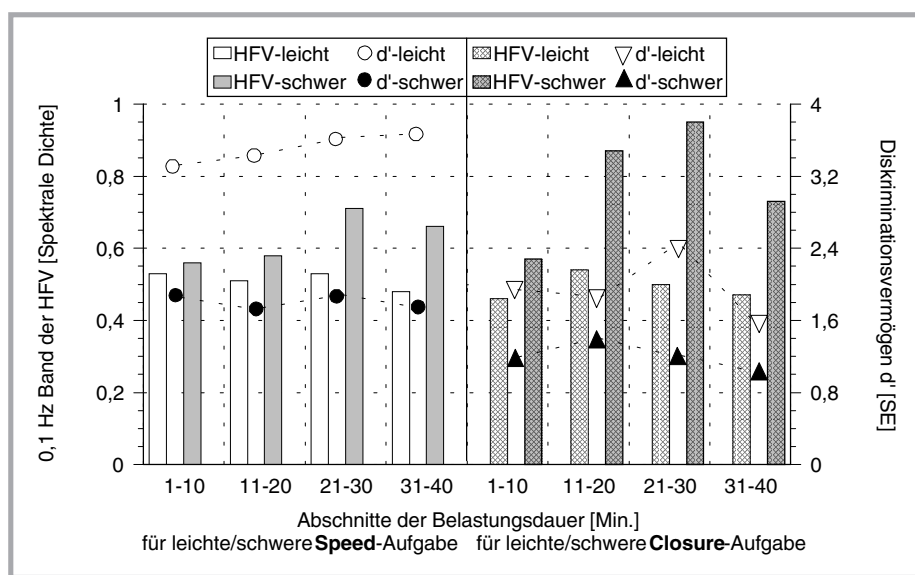


Bild 2: Ausprägungen der 0,1 Hz Komponente der HFV und des Diskriminationsvermögens (d') für die beiden Arten der Informationsverarbeitung (speed: linke Seite; closure: rechte Seite) jeweils in unterschiedlichen Schwierigkeitsgraden (leicht; schwer) über die vier Abschnitte der Belastungsdauer



die übrigen Leistungsdaten, Treffer- und Falsche-Alarme-Raten sowie Antwortzeiten, korrespondieren mit diesen Ergebnissen. Da sich für das Entscheidungskriterium (β) für keinen der Faktoren signifikante Effekte ergeben, ist auch nicht von einer Verschiebung des Kriteriums bzw. einer Veränderung zu risikoreicheren oder vorsichtigeren Entscheidungen auszugehen.

Die Ergebnisse für die Ausprägungen der 0,1 Hz Komponente der HFV (Bild 2: linke y-Achse) stehen allerdings nicht nur im *völligen Widerspruch* zu den Parametern der Leistungsdaten, sie zeigen sich darüber hinaus auch im Widerspruch zu den theoretischen Annahmen. Da davon ausgegangen wird, dass mit einer steigenden Unterdrückung der 0,1 Hz Komponente der HFV ein Anstieg psychischer oder mentaler Beanspruchung einhergeht, weisen die hohen Ausprägungen auf eine relativ niedrige Beanspruchung hin und umgekehrt. Danach würde die Bearbeitung der leichten Speed-Aufgabe auf einem tendenziell höheren Beanspruchungsniveau stattfinden als die der schweren Aufgaben gleichen Typs. Da sich entsprechendes auch für den Aufgabentyp Closure-Aufgabe zeigt, wäre davon auszugehen, dass die Bearbeitung der leichten Aufgaben beanspruchender ist als die der schweren. Nach den statistischen Analysen zu den Ausprägungen der 0,1 Hz Komponente der HFV zeigen sich aber weder bedeutsame Haupt- oder Interaktionseffekte, womit es nach der hier vorgestellten Untersuchung mit diesem Maß weder möglich ist, die aus unterschiedlichen *Arten* der Belastung noch die aus unterschiedlichen *Intensitäten* der Belastung resultierende Beanspruchung zu diskriminieren.

Würde man nun, obwohl die unterschiedlichen Aufgaben zu deutlichen Unterschieden in den Leistungsparametern führen, davon ausgehen, dass die in dieser Studie vorgenommene Abstufung der Belastungsintensität nicht ausreichend unterschiedlich war, würden sich damit einerseits Ergebnisse anderer Untersuchungen bestätigen, nach denen sich der Indikator nur als sensitiv für extreme Belastungsunterschiede zeigt (z.B. Mulder et al. 2000; Paas et al. 1994), lässt dann allerdings auch die Brauchbarkeit dieses Indikators zur Differenzierung von Belastungssituationen bei praktischen Arbeitstätigkeiten mehr als fraglich erscheinen. Andererseits konnte über diese Studie die wiederholt berichtete und auch zur Bestätigung theoretischer Konzepte als sehr bedeutsam erachtete Differenzierung zwischen automatisierter und gesteuerter

Informationsverarbeitung (Mulder & Mulder 1981; Aasman et al. 1987) zumindest über die hier vorgenommene Operationalisierung durch unterschiedliche Signal-Entdeckungsaufgaben *nicht* bestätigt werden.

Eventuell ist aber die mangelnde Diskriminationsfähigkeit des Indikators auf die - die belastungsbedingungsbezogenen Effekte u.U. maskierende - hohe interindividuelle Varianz in den HFV-Daten rückführbar (vgl. Eilers 1999; Nickel et al. 1998, 2001). Zur Prüfung dieser Hypothese wurden Kovarianzanalysen der HFV-Werte mit den als Bezugsbasis verwendeten HFV-Daten aus den einfachen Reaktionstests zur Untersuchung von a-priori-Unterschieden zwischen den Vpn herangezogen. Das Herausparsieren des signifikanten Einflusses der Kontrollvariablen (Bezugsbasis) hatte aber nicht den Effekt, dass durch die Ausprägungen der 0,1 Hz Komponente der HFV nun Unterschiede zwischen den Belastungssituationen erkennbar wurden. Dies führte im Gegenteil dazu, dass die gesamte Varianz des within-Faktors Level (leicht vs. schwer) auf null zurückging (!). Kontrolliert man also die personenspezifische Varianz, verschwindet die in den Daten enthaltene Varianz - diese kann also nur auf *interindividuelle* Unterschiede der Vpn, nicht aber auf unterschiedliche Belastungsbedingungen zurückgeführt werden. Aber auch die mangelnde Differenzierungsfähigkeit zwischen den *Arten* der Belastung (hier Anforderungen der automatisierten vs. gesteuerten Informationsverarbeitung der Aufgabenarten), die sich sehr deutlich in allen Leistungsdaten widerspiegelt, bleibt bemerkenswert.

Die hier vorgestellten Ergebnisse sprechen damit ebenso wie die von Eilers (1999) dafür, dass die 0,1 Hz-Komponente der HFV eher *interindividuelle* und damit personenspezifische als *intersituative*, d.h. auf unterschiedliche Grade mentaler Belastung zurückführbare Beanspruchungsreaktionen differenziert. Das würde bedeuten, dass das Verfahren als Maß psychischer oder mentaler Beanspruchung zur Beurteilung der *Arbeitsbedingungen* nicht zu empfehlen ist.

Offensichtlich muss es sich um sehr gravierende Unterschiede der Belastung handeln, um eine Differenzierung der Beanspruchung anhand der HFV sichtbar werden zu lassen. Als durchaus plausibel erscheint daher nach den bisher vorliegenden Befunden die Überlegung, dass dem Ansprechen dieses Indikators eine bestimmte Form der Beanspruchung, z.B. durch Zeit- oder Verant-

wortungsdruck oder die Gefahr des Scheiterns, zugrunde liegt. Auch dahingehend ließen sich die Ergebnisse der Kovarianzanalysen interpretieren, sofern den Aufgaben enge Zeitbindung als gemeinsame Komponente zugrunde liegt, die dann zur Varianzreduktion führt, z.B. der Reduzierung emotional bedingter Varianzkomponenten. Damit würden sich Hinweise aus anderen (z.B. Kamphuis & Frowein 1985; Pruyn 1986) und der hier berichteten Studien verdichten, wonach der Indikator ein Maß unspezifischer Aktivierung (z.B. durch Bedrohung des Scheiterns einer erfolgreichen Aufgabenbearbeitung, Stress) ist und wonach die 0,1 Hz Komponente der HFV wahrscheinlich lediglich emotionale, nicht aber mentale oder informatorische Beanspruchungsreaktionen abbildet.

3.4 Belastung durch unterschiedliche kognitive Anforderungen

Nun könnte nach den vorgestellten Untersuchungen eingewandt werden, dass die 0,1 Hz Komponente der HFV lediglich nicht sensitiv und diagnostisch für die psychische Beanspruchung durch die in den Studien von Eilers (1999) und Lasner (1997) eingesetzten Daueraufmerksamkeitsbelastung ist. Das könnte darauf hinweisen, dass dieser Indikator die Inanspruchnahme psychischer oder mentaler Leistungsfunktionen zwar nicht bei den oben gewählten (und für die Praxis durchaus relevanten), hingegen aber die bei anderen Aufgaben - und dann diagnostisch und dabei u.U. sogar auch sensitiv - differenzieren könnte.

In einer laborexperimentellen Studie wurde daher von Nickel (2001) der Frage nachgegangen, ob mit der 0,1 Hz Komponente der HFV die (qualitativ und quantitativ unterschiedliche) psychische Beanspruchung aus der Bearbeitung sehr unterschiedlicher *Arten* von Aufgaben mit unterschiedlicher Beanspruchung mentaler Funktionen in jeweils unterschiedlichen *Graden* ihrer Schwierigkeit diskriminiert werden kann. Mit den Standardized Tests for Research with Environmental Stressors (STRES, AGARD 1989) wurde auf eine explizit für die internationale Mental Workload Forschung entwickelte Batterie von 14 Tests, die sich z.T. deutlich nach den zur Bearbeitung erforderlichen Funktionen unterscheiden, zurückgegriffen. Aus den Untersuchungen (s.a. Schütte & Nickel 2002) konnten Daten von insgesamt 14 Vpn in die Analysen einbezogen werden. Jede Vp bearbeitete alle 14 Einzelaufgaben mit jeweils direkt anschlie-

ßenden Ruheabschnitten über jeweils fünf Minuten, mit einem Wiederholungsversuch an einem der nachfolgenden Tage (vgl. Nickel 2001).

Erwartungskonform deuten die Ergebnisse für die 0,1 Hz Komponente der HFV (Bild 3: dunkle Balken für Arbeit, helle Balken für Ruhe; hier nur für ausgewählte Aufgaben dargestellt) auf einen auch statistisch gesicherten Unterschied zwischen den Abschnitten der Aufgabenbearbeitung und den Abschnitten der jeweils direkt anschließenden Ruhephase hin. Damit weist der Indikator die Arbeits- bzw. Aufgabenabschnitte auf einem deutlich *höheren* Beanspruchungsniveau aus als die Ruheabschnitte - sofern man zunächst von den Ergebnissen für GR absieht, die überraschend und unerwartet auf ein den Ruheabschnitten vergleichbares und niedriges Beanspruchungsniveau hinweisen. Nach den Analysen der subjektiv empfundenen Schwierigkeit (ZEIS, Pitrella 1998; im folgenden Schwierigkeitsindex; Bild 3: schwarz gestufte Linie) wird allerdings die Bearbeitung der Aufgabe zum logischen Denken (GR) als am schwierigsten und die Bearbeitung der Reaktionszeitaufgabe „basis“ (RT) als am leichtesten skaliert. Alle anderen Aufgaben werden dazwischen breit gestreut skaliert. Während nach den Ergebnissen des Schwierigkeitsindex unterschiedliche Aufgabenarten deutlich differenziert werden können, ist das mit der 0,1 Hz Komponente der HFV nicht möglich. Es könnte nun davon ausgegangen werden, dass die Ergebnisse der 0,1 Hz Komponente der HFV und des Schwierigkeitsindex nur deshalb inkonsistent sind, weil bei letzterem

von den Vpn das Beanspruchungsniveau nur subjektiv (und möglicherweise nicht korrekt) eingeschätzt wurde. Bezieht man nun die Ergebnisse für den Leistungsindex (Bild 3: grau gestufte Linie) mit ein, so zeigt sich konsistent bei denjenigen Aufgaben, deren Bearbeitung subjektiv als schwieriger empfunden wurde auch ein deutlich niedrigeres Leistungsergebnis, was immer dabei Ursache und was Effekt sein mag. Für die 0,1 Hz Komponente der HFV ergibt sich somit ein sowohl zum Leistungsindex als auch zum Schwierigkeitsindex widersprüchliches und nicht erwartungskonformes Bild.

Geht man nun dennoch davon aus, dass die 0,1 Hz Komponente der HFV tatsächlich psychische Beanspruchung indiziert, so deuten die varianzanalytischen Auswertungen darauf hin, dass dieses Messverfahren vergleichbar den Ergebnissen bei Eilers (1999; s.a. Mulder et al. 2000) sensitiv zwischen den zwei unterschiedlichen Belastungsintensitäten, nämlich Arbeit vs. Ruhe, differenzieren kann; eine Differenzierung, die einfacher zu haben ist als mit einer derartig aufwendigen Technik. Hinweise auf eine Sensitivität für graduelle Belastungsabstufungen, wie sie durch die Aufgaben gleicher Art aber unterschiedlicher Belastungsintensität (z.B. DU (=Doppelaufgaben aus Tracking + Sternberg) in Bild 3) realisiert wurden, ergeben sich hingegen nicht. Da insbesondere für die Bearbeitung der GR-Aufgaben (grammatical reasoning) - stärker als bei allen anderen Aufgaben - gesteuerte Informationsverarbeitung notwendig ist und der Indikator für diese Aufgabe auf ein dem Ruheabschnitt vergleich-

bares Beanspruchungsniveau hinweist, scheint es ebenfalls *nicht* gerechtfertigt, der 0,1 Hz Komponente der HFV Diagnostizität zur Differenzierung von Belastungsarten zuzuschreiben.

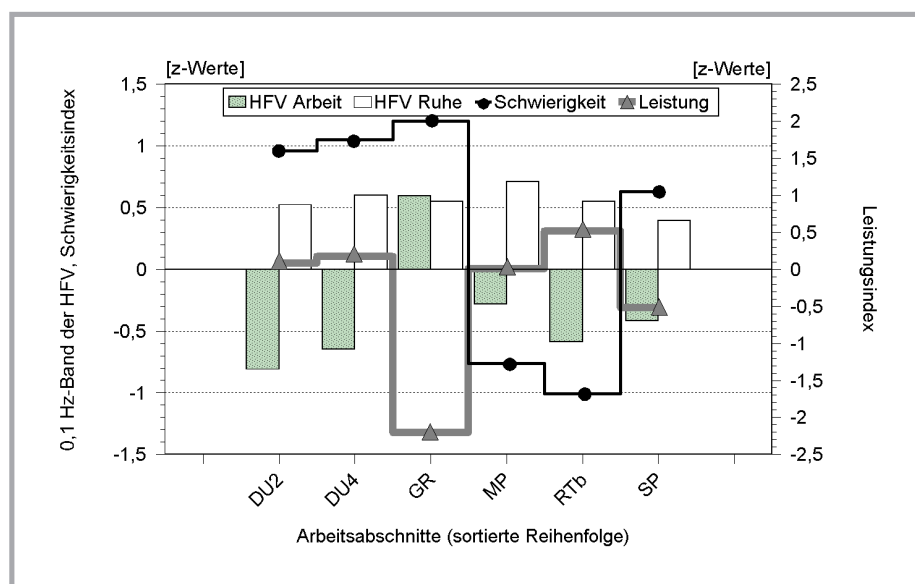
Weitergehende Überlegungen und Erwägungen verschiedener plausibler Erklärungen für die bisher gefundenen Ergebnisse zur 0,1 Hz Komponente der HFV ließen vermuten, dass ähnlich des plausiblen Erklärungsansatzes für die Ergebnisse der Studie von Lasner (1997; s.a. Nickel et al. 2001), auch bei dieser Studie ein über nahezu alle Aufgaben gemeinsamer Faktor (z.B. Zeitdruck durch Pacing-Effekte bzw. Effekte der Aufgabentaktung) die Ausprägungen des Indikators beeinflusst oder sogar maßgeblich bestimmt (vgl. Nickel 2001).

Auf der Grundlage eines bei Nickel (2001) vorgestellten Pacing-Konzeptes (Pacing i.S.v. Arbeitstakt) würden die Ergebnisse dann auf einen graduell abgestuften Effekt zwischen selbst- und fremdgetaktet („self-paced“ vs. „machine-paced“) dargebotenen Aufgaben, und damit eine unterschiedliche Bedrohung durch das Scheitern einer ziel-führenden Aufgabenbearbeitung und eine dadurch ggf. ausgelöste unterschiedliche Angst- oder bedrohungsinduzierte und damit wieder aufgabenunspezifische Aktivierung hindeuten.

3.5 Pacing-Effekte

Zur Überprüfung des Einflusses eines solchen Pacing-Effektes wurde eine eigenständige Untersuchung konzipiert, mit der untersucht wurde, inwieweit sich mit Hilfe der 0,1 Hz Komponente der HFV entweder tatsächlich Unterschiede der mentalen Beanspruchung oder doch eher ein graduell abgestufter Effekt zwischen self-paced und machine-paced dargebotenen Aufgaben und dadurch ausgelöste Aktivierungsunterschiede differenzieren lassen (vgl. Nickel 2001). Zur Operationalisierung unterschiedlicher Arten und Intensitäten der Belastung wurde hier auf die Reaktionszeitaufgabe „basis“ (RT) und die Aufgabe zum logischen Denken (GR) aus den STRES (AGARD 1989) zurückgegriffen, die nach einem in Nickel (2001) beschriebenen Pacing-Konzept als machine-paced bzw. als self-paced klassifiziert werden konnten. Darüber hinaus wurde die GR so modifiziert, dass neben der o.g. self-paced-Variante auch eine machine-paced-Variante (auf Basis der jeweils individuellen Arbeitsleistung) vorlag. Diese beiden Varianten der GR unter-

Bild 3: Ausprägungen der 0,1 Hz Komponente der HFV, des Leistungsindex und des Schwierigkeitsindex für ausgewählte Aufgaben (DU, GR, MP, RT, SP) aus den AGARD-STRES aggregiert über Versuchspersonen und -durchgänge (vgl. Nickel 2001)



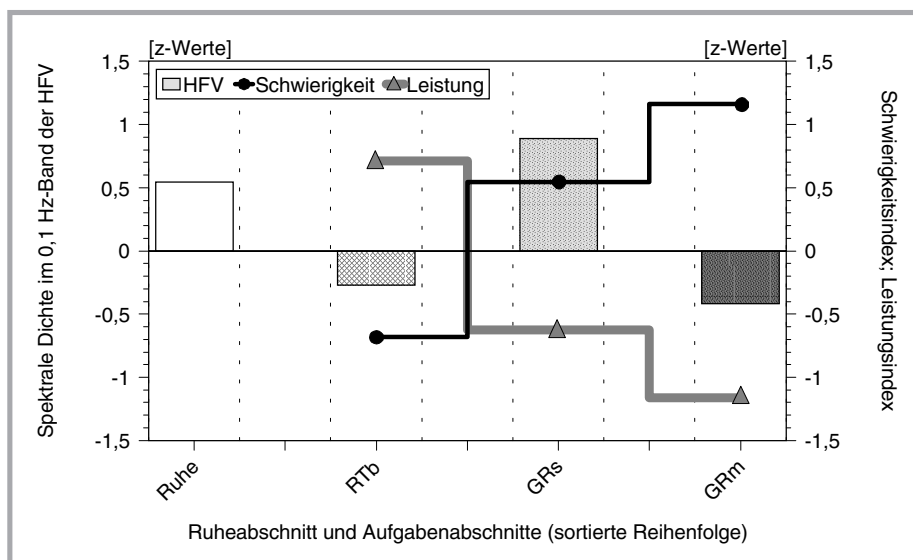


Bild 4: Ausprägungen der 0,1 Hz Komponente der HFV, des Leistungsindex und des Schwierigkeitsindex aggregiert über Versuchspersonen und -bedingungen und wiederholte Aufgabenbearbeitungen (vgl. Nickel 2001)

schieden sich allerdings weder in ihrer Art noch in der Schwierigkeit der Anforderungen an mentale/kognitive Funktionen zur Aufgabenbearbeitung. Es wurden lediglich Veränderungen der Anforderungen an das Zeitmanagement zur Aufgabenbearbeitung (Inter-Stimulus-Intervall und Response-Stimulus-Intervall) vorgenommen. Durch die engeren zeitlichen Rahmenbedingungen bei der machine-paced-Variante wurden die Versuchsperson zeitlich stärker ‚gedrückt‘, wodurch es zu einer Bedrohung durch Scheitern oder sogar Nichterreichen von Aufgabenzielen (z.B. Aufgaben möglichst schnell in der vorgegebenen Zeit zu bearbeiten und dabei Fehler vermeiden), zum Erleben von potentieller Überforderung und Zeitdruck und damit zu Frustrationen, zu Angst(ähnlichen)-Zuständen kommen kann (vgl. Rutenfranz 1960; Nachreiner & Wucherpennig 1975). Als beanspruchungswirksame Folge wird ein emotionaler Erregungszustand erwartet, der sich dann als erhöhte Aktivierung in der 0,1 Hz Komponente der HFV widerspiegeln sollte.

Die drei unterschiedlichen Aufgabenvarianten (RT, GRself-paced, GRmachine-paced) wurden im mehrmaligen Wechsel insgesamt 10 Vpn über jeweils ca. drei Stunden präsentiert (Nickel 2001).

Die Ergebnisse der Analysen sind in Bild 4 hoch aggregiert zusammengefasst, da sich keine bedeutsamen Effekte für die wiederholten Bearbeitungen der jeweiligen Aufgaben ergaben. Die Bearbeitung der einfachen, fertigkeitbasierten Reaktionszeitaufgaben „basis“ (RT) wurde anhand des Schwierigkeitsindex (Bild 4: schwarz gestufte Linie) als sehr leicht eingeschätzt,

wohingegen die Bearbeitung beider GR-Varianten als deutlich schwieriger empfunden wurde. Darüber hinaus wurde die Bearbeitung der machine-paced-Variante relativ zur self-paced-Variante der GR als schwieriger empfunden. Diese Ergebnisse sind konsistent mit denen für den Leistungsindex (Bild 4: grau gestufte Linie) und ließen sich ebenfalls durch die statistischen Analysen absichern. Ein hoher Schwierigkeitsindex fällt damit mit einem niedrigen Leistungsindex zusammen und umgekehrt. Da eine steigende Unterdrückung der 0,1 Hz Komponente der HFV (Bild 4: Balken) mit einem Anstieg psychischer oder mentaler Beanspruchung einhergehen soll, weisen niedrige Ausprägungen des Indikators auf eine hohe Beanspruchung hin und umgekehrt. Damit weist der Indikator lediglich für fremdgetaktete Aufgaben eine hohe Beanspruchung aus, und zwar für die einfache, fertigkeitbasierte Reaktionszeitaufgabe ‚basic‘ (RT) sowie für die sehr viel komplexere, regelbasierte machine-paced dargebotene Variante des GR. Für die von der Aufgabenbearbeitung ebenso komplexe, aber self-paced dargebotene Variante des GR weist die 0,1 Hz Komponente der HFV dagegen ein Beanspruchungsniveau aus, das deutlich unter dem der anderen beiden Aufgaben liegt und wiederum dem Niveau des Ruheabschnittes entspricht. Auch diese Ergebnisse sind statistisch abgesichert (vgl. Nickel 2001).

Die auf der Grundlage des Pacing-Konzepts vorgenommene Klassifizierung der Aufgaben wird sowohl durch die Ergebnisse für den Leistungsindex als auch für den Schwierigkeitsindex abgebildet. Zu diesen Ergebnissen stehen die Ausprägungen der

0,1 Hz Komponente der HFV allerdings im *völligen Widerspruch*. Geht man nun davon aus, dass die Differenzierung des Beanspruchungsindikators zwischen self-paced und machine-paced dargebotenen Aufgaben eine unterschiedliche Belastungsintensität abbildet, dann wäre die 0,1 Hz Komponente der HFV ausreichend sensitiv, um die aus den Belastungsbedingungen ‚self-paced‘ vs. ‚machine-paced‘ resultierende Beanspruchung zu diskriminieren. Allerdings wurde dabei keine Variation der Aufgabenart (i.S.d. kognitiven Anforderungsstruktur zur Aufgabenbearbeitung) vorgenommen, sondern bei der Variation handelt es sich lediglich um Veränderungen der Zeitsteuerung der Aufgabenbearbeitung (hier Zeitdruck durch die Bedrohung des Scheiterns). Damit liegen den unterschiedlichen Ausprägungen der 0,1 Hz Komponente der HFV Veränderungen unspezifischer, allgemeiner Aktivierung (hervorgerufen durch z.B. Arousal, Angst, Frustration, Stressreaktionen) und damit wahrscheinlich emotionale Beanspruchungskomponenten - im Gegensatz zu mentaler oder informatorischer Beanspruchung zugrunde.

4 DISKUSSION

Aufgrund der Ergebnisse der hier dargestellten Studien zeigt sich, dass die ausgewählten HF- und HFV-Maße - und dies gilt auch für die 0,1 Hz Komponente der HFV - zur Erfassung der aus variierten Belastungsbedingungen resultierenden psychischen Beanspruchung nicht geeignet sind.

Bezogen auf die *Reliabilität* weisen die Analysen von Eilers (1999) insgesamt keine *intraindividuelle Kurzzeit-Stabilität* für die HF- und HFV-Maße unter Ruhebedingungen nach. Ursachen dafür können zumindest für die 0,1 Hz Komponente der HFV in den bei feiner auflösenden Verlaufsanalysen erheblichen auch kurzzeitigen Instabilitäten (Spontanfluktuationen) gesehen werden, wie sie auch in Ruheabschnitten aus Feldstudien (Nachreiner et al. 1999; Nickel et al. 1998) beobachtet werden konnten. Eine *intraindividuelle Langzeit-Reliabilität* ließ sich unter Ruhebedingungen für alle HF- und HFV-Maße feststellen; für einige der Maße - allerdings nicht für die 0,1 Hz Komponente der HFV - sogar eine situations- bzw. tageszeitbezogene Langzeit-Reliabilität.

Zur Ableitung von Aussagen über die *Validität* und dann auch *Sensitivität* und *Diagnostizität* ist eine differenzierte Ergebnis-



betrachtung notwendig. In allen Validierungsstudien zeigt sich, dass *interindividuelle* Unterschiede recht zuverlässig fassbar und reproduzierbar sind. Damit wird ein relativ zur Gesamtvarianz hoher personenspezifischer Anteil über die HF- und HFV-Maße gebunden, der eine für die Beurteilung der Belastungssituationen - über die resultierende psychische Beanspruchung - nicht interessierende Differenzierung von Individuen ermöglicht. Das spiegelt sich in der hohen interindividuellen Personenvarianz fast aller HF- und HFV-Maße wider (Eilers 1999), wird in Lasner (1997) mit der vollständigen Elimination der Varianz des Faktors ‚Level‘ nach Herauspartialisieren von a-priori-Unterschieden deutlich und konnte in Nickel (2001) nur durch eine über die Belastungsbedingungen jeder Personen vorgenommene z-Transformation der Daten nicht mehr so offensichtlich werden. Die Ergebnisse belegen ferner, dass man auch durch eine Reduktion der interindividuellen Varianz durch geeignete Kontrolltechniken nicht zu günstigeren Ergebnissen kommt. Obwohl die Personen unter deutlich unterschiedlichen Belastungsbedingungen arbeiteten, zeigten sich die HF- und HFV-Maße personenbezogen stabil (Eilers 1999; Lasner 1997).

Aus der für die Beurteilung der HF- und HFV-Maße zur Indikation psychischer Beanspruchung interessierenden *intersituativen* Differenzierung ergeben sich über alle Studien zunächst scheinbar unterschiedliche Ergebnisse. Wenn davon ausgegangen wird, dass es sich bei den in den Studien von Eilers (1999) und Nickel (2001) verwendeten Abschnitten der Ruhe und der Aufgabebearbeitung um eine Variation der Belastungsintensität handelt, so zeigen sich die jeweils verwendeten Maße hinreichend *sensitiv*, um die Belastungsbedingungen Ruhe vs. Arbeit über die resultierenden Beanspruchungskennwerte zu differenzieren. Allerdings ist dies nach aller praktischen Lebenserfahrung auch über einfachere Verfahren möglich. Andererseits war es über die hier analysierten Kennwerte in keiner der fünf vorgestellten Studien möglich, feiner abgestufte Belastungsgrade zu differenzieren, obwohl sowohl bei den verwendeten Leistungsmaßen als auch bei den Einschätzungen des subjektiven Beanspruchungsempfindens klare Effekte sichtbar wurden. Das betrifft die Schwierigkeitsabstufungen für die Vigilanz-Aufgaben bei Eilers (1999) und die Signal-Entdeckungsaufgaben bei Lasner (1997) als auch die von ihrer Art grundsätzlich anderen Aufgaben bei Nickel (2001).

Bezieht man die Dynamik der Belastung mit ein, so kann mit zunehmender Dauer und mit steigender Belastungsintensität, die schließlich sogar zu einer Überschreitung der ‚Dauerleistungsgrenze‘ führen kann, erwartet werden, dass sich dies auch in den Ausprägungen der Beanspruchungsmaße widerspiegelt. Es lässt sich allerdings weder aus den Studien von Eilers (1999) noch von Lasner (1997) mit Hilfe der jeweils verwendeten Maße ein Effekt der Belastungsdauer i.S. einer Zunahme der Beanspruchung ableiten. Vielmehr weisen die Ergebnisse für die 0,1 Hz Komponente der HFV in den zeitlich längeren Abschnitten bei Eilers (1999) - bezieht man sowohl Leistungsergebnisse als auch die Einschätzungen des subjektiven Beanspruchungsempfindens ein - eher auf eine Deaktivierung über den Versuchsdurchgang hin. Nach Ergebnissen für verschiedene HF- und HFV-Maße (Eilers 1999) und die 0,1 Hz Komponente der HFV aus allen hier vorgestellten Studien ergeben sich *keine Hinweise auf eine hinreichende Sensitivität*, die über die offensichtliche Differenzierung zwischen Arbeit vs. Ruhe hinausgeht, weder für eine Differenzierung zwischen feineren Graden psychischer - oder genauer mentaler - Beanspruchung (Intensität und Dynamik), und zwar weder aufgabenunspezifisch (z.B. Belastungsdauer) noch aufgabenspezifisch (z.B. Abstufungen kognitiver Anforderungsstrukturen bei Aufgaben gleicher Art). Vielmehr weisen auch diese Ergebnisse die 0,1 Hz Komponente der HFV wieder als einen aufgabenunspezifischen (z.B. auf Ruhe vs. Arbeit, auf die Enge der Zeitbindung bezogenen) oder personenspezifischen Indikator der Aktivierung aus.

Als möglicher Grund für die relative Insensitivität der 0,1 Hz Komponente der HFV wird von Aasman et al. (1987) auf eine hohe personenspezifische Varianz verwiesen, was durch die hier durchgeführten Analysen bestätigt wird. Darüber hinaus weisen Mulder et al. (2000) darauf hin, dass nur relativ große Unterschiede zwischen Aufgaben durch die 0,1 Hz Komponente indiziert werden. Sie verweisen dabei auf ältere Untersuchungen (Mulder & Mulder 1981), nach denen mit dem Indikator eine Differenzierung zwischen leichten und schweren kognitiven Anforderungen oder automatisierter und gesteuerter Informationsverarbeitung möglich war, wenngleich solche Ergebnisse lediglich von Aasman et al. (1987) und auch nur in Ansätzen repliziert werden konnten. Damit wird aber bereits auf die *Diagnostizität* von HF- und HFV-Maßen verwiesen. Nimmt man an, dass sich unterschiedliche Tageszeiten i.S.v. Belastungs-

bedingungen unterscheiden, dann zeigen sich nach Eilers (1999) einige der HF- und HFV-Maße zur Differenzierung der unterschiedlichen Belastung (siehe auch Tabelle 2) geeignet. Mit der Diskriminierung unterschiedlicher Tageszeiten unterliegt die Ausprägung der 0,1 Hz Komponente der HFV - unabhängig von der zu bearbeitenden Aufgabenstellung und der damit einhergehenden und nicht indizierten psychischen Inanspruchnahme von Leistungsvoraussetzungen - circadianen Schwankungen der allgemeinen Aktivierung. Diagnostizität i.S. einer Differenzierung zwischen unterschiedlichen Arten der erforderlichen Informationsverarbeitung (automatisiert vs. gesteuert, Sukzessiv- vs. Simultandiskrimination), wie sie über die Speed- und die Closure-Aufgaben in den Studien von Eilers (1999) und Lasner (1997) verwendet wurden und sich nach Ergebnissen der verwendeten Maße der Leistung und des subjektiven Beanspruchungsempfindens abzeichnen, kann allerdings für die jeweils verwendeten HF- und HFV-Maße - entgegen der von Mulder et al. (2000) berichteten Ergebnisse - nicht abgeleitet werden. Auch nach den Ergebnissen von Nickel (2001) ließ sich für die nach ihrer Art sehr verschiedenen Aufgaben keine positive Beurteilung der Diagnostizität ableiten. Dabei wäre zunächst abzusehen von der - nach dem subjektiven Empfinden und den Leistungskennwerten - als am schwierigsten einzustufenden GR-Aufgabe, die nach den Ausprägungen der 0,1 Hz Komponente der HFV als am geringsten beanspruchend - vergleichbar einem Ruheabschnitt - einzustufen wäre, sofern man an einer Diagnostizität i.S.v. Diskriminierung von GR-bezogenen Anforderungen vs. nicht-GR-bezogenen Anforderungen interessiert ist. Dabei ist der entscheidende Unterschied in den Anforderungen dann die Zeitbindung, die man auch mit einfacheren Methoden erkennen kann.

Anhand der Pacing-Studie (Nickel 2001) konnte allerdings eine Diskriminierung von Aufgabebearbeitung mit vs. ohne Zeitdruck und damit ggf. Diagnostizität zur Erfassung *emotionaler* vs. *mentaler* Beanspruchung nahegelegt werden. Darüber hinaus konnte gezeigt werden, dass sich dieser Erklärungsansatz auch für die Ergebnisse anderer Untersuchungen zur 0,1 Hz Komponente der HFV (z.B. Aasman et al. 1987; Kamphuis & Frowein 1985; Wiethoff 1997) und anderer HF- und HFV-Maße (z.B. Coury & Drury 1986; Rutenfranz 1960; Sharit & Salvendy 1982) als gültig erweisen könnte. Ebenso wird darüber nachvollziehbar, warum sich auch in den Studien von Eilers (1999) mit

verschiedenen HF- und HFV-Maßen und Lasner (1997) mit der 0,1 Hz Komponente der HFV keine Sensitivität zur Differenzierung der experimentellen Belastungsvariationen zeigen konnte.

Auf der Grundlage der vorgestellten Studien erscheinen HF- und HFV-Maße zwar als (personenbezogen) hinreichend reliabel, nicht aber als hinreichend valide, sensitiv oder diagnostisch zur Erfassung psychischer oder mentaler Beanspruchung durch unterschiedliche Arten und/oder Intensitäten psychischer Belastung, es sei denn man ist lediglich an der Erfassung einer (Teil-)Komponente psychischer Beanspruchung, nämlich einer unter - z.B. Bedingungen circadianer Rhythmik oder Zeitdruckbedingungen auftretenden - aufgabenunspezifischen Erregung oder Aktivierung bzw. emotionalen Beanspruchung interessiert. Inwieweit der Indikator allerdings für verschiedene Arten emotionaler Beanspruchung diagnostisch ist und auch Abstufungen nach ihrer Intensität sensitiv differenziert, kann auf der Grundlage der hier vorgestellten Studien nicht beurteilt werden.

5 THEORETISCHE UND PRAKTI-SCH E KONSEQUENZEN

Die vorgestellten Ergebnisse belegen, dass HF und HFV-Maße zur Erfassung der psychischen Beanspruchung nicht geeignet sind. Ein valider Rückschluss von der psychischen Beanspruchung auf die psychische Belastung über diese Indikatoren kann daher nicht gelingen.

Werden dennoch derartige Messverfahren, die weder psychische noch mentale sondern lediglich - wenn überhaupt - emotionale Beanspruchung indizieren, zur Beurteilung der psychischen Belastung eingesetzt, so können Arbeitssysteme nach einer vermeintlich geringen psychischen Beanspruchung fälschlicherweise (und zwar nicht auf der Grundlage von Fakten, sondern von Artefakten) als gering belastend beurteilt werden, mit der Konsequenz, dass die Notwendigkeit präventiver Interventionen, z.B. zur Vermeidung psychischer Ermüdung, nicht identifizierbar ist, mit ggf. weitreichenden Konsequenzen für Sicherheit und Gesundheit nicht nur des Operators. Vom Einsatz solcher - und ähnlicher, nicht hinreichend auf ihre psychometrischen Gütekriterien untersuchter - Messverfahren als Bewertungsinstrumente zur Beurteilung der psychischen Beanspruchung und darüber der psychischen Belastung ist daher abzuraten.

Was diese Verfahren erfassen, scheint nach den vorliegenden Ergebnissen lediglich der, wodurch auch immer bedingte, Aktivierungszustand des Menschen. Diesen mit der psychischen Beanspruchung gleichzusetzen erscheint durch nichts zu rechtfertigen.

Um den Forderungen auch der betrieblichen Praxis - wie sie z.B. aus der Bildschirmarbeitsverordnung resultieren - nach geeigneten Verfahren zur Erfassung und Beurteilung psychischer Belastung entgegenzukommen, dürfte die Entwicklung spezifischer Mess- und Beurteilungskonzepte, ggf. für spezifische Arbeitstätigkeiten und Fragestellungen, angemessener sein. Dazu könnten dann Verfahren zur Erfassung von spezifizierten Komponenten eines mehrdimensionalen (Komposit-)Konstruktes „psychische Beanspruchung“ geeignet sein, die spezifisch, konzeptadäquat und für bestimmte, begrenzte Gültigkeits- und Anwendungsbereiche konstruiert und nach ihrer systematischen Evaluation auch mit nachvollziehbar konkreten Angaben zur Erfüllung psychometrischer Gütekriterien dokumentiert werden können (vgl. Eilers et al. 1990; Eilers 1999; Nickel 2001). Auf valide universelle, d.h. für beliebige Arbeitstätigkeiten und beliebige Fragestellungen einsetzbare Verfahren wird man auch in absehbarer Zukunft - wenn nicht sogar generell - verzichten müssen, auch und insbesondere im Bereich psychophysiologischer Indikatoren.

LITERATUR

Aasman, J.; Mulder, G. & Mulder, L.J.M.: Operator effort and the measurement of heart-rate variability. *Human Factors*, 29, 161-170, 1987

AGARD (Advisory Group for Aerospace Research and Development): *Human performance assessment methods* (AGARDograph No. 308). Neuilly sur Seine: NATO, 1989

Apparies, R.J.; Riniolo, T.C. & Porges, S.W.: A psychophysiological investigation of the effects of driving longer-combination vehicles, *Ergonomics*, 41, 581-592, 1998

BildscharbV: *Verordnung über Sicherheit und Gesundheitsschutz bei der Arbeit an Bildschirmgeräten* (Bildschirmarbeitsverordnung). In: Bundesgesetzblatt, Teil I, 1841, 1996

Boucsein, W. & Backs, R.W.: Engineering psychophysiology as a discipline: Historical and theoretical aspects. In: R.W. Backs & W. Boucsein (eds.), *Engineering psychophysiology. Issues and*

applications (3-30). Mahwah: LEA, 2000

Brennan, R.L.: *Elements of generalizability theory*. Iowa City: ACT Publications, 1992

Caldwell, J.A.; Wilson, G.F.; Cetinguc, M.; Gailard, A.W.K.; Gundel, A.; Lagarde, D.; Makeig, S.; Myhre, G. & Wright, N.A.: *Psychophysiological assessment methods* (AGARD-Advisory Report 324). Neuilly sur Seine: NATO, 1994

Coury, B.G. & Drury, C.G.: The effects of pacing on complex decision-making inspection performance. *Ergonomics*, 29, 489-508, 1986

Cronbach, L.J.; Gleser, G.C.; Nanda, H. & Rajaratnam, N.: *The dependability of behavioural measurements: Theory of generalizability for scores and profiles*. New York: Wiley, 1972

Eilers, K.: *Herzfrequenz und Herzfrequenzvariabilität als Indikatoren psychischer Beanspruchung. Zur Reliabilität und Validität von Herzfrequenz- und Herzfrequenzvariabilitätsmaßen als Indikatoren psychischer Beanspruchung bei Daueraufmerksamkeitsbelastungen*. Frankfurt/Main: Lang, 1999

Eilers, K. & Nachreiner, F.: Time of day effects in vigilance performance at simultaneous and successive discrimination tasks. In: G. Costa, G. Cesna, K. Kogi & A. Wederburn (eds.), *Shift-Work: Health, sleep, and performance* (467-472). Frankfurt/Main: Lang, 1990

Eilers, K.; Nachreiner, F. & Hänecke, K.: Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. *Zeitschrift für Arbeitswissenschaft*, 40, 215-224, 1986

Eilers, K.; Nachreiner, F. & Böning, E.: Zur subjektiven Skalierung psychischer Beanspruchung. Teil 1: Experimentelle Überprüfung der Validität verankerter Relativurteile. *Zeitschrift für Arbeitswissenschaft*, 43, 217-223, 1989

Eilers, K.; Nachreiner, F. & Böning, E.: Zur subjektiven Skalierung psychischer Beanspruchung. Teil 2: Überprüfung der Validität verankerter Relativurteile in einer Felduntersuchung. *Zeitschrift für Arbeitswissenschaft*, 44, 24-29, 1990

Fisseni, H.-J.: *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe, 1990

Folkard, S.: Effects on performance efficiency. In: W.P. Colquhoun; G. Costa; S. Folkard & P. Knauth (eds.), *Shiftwork. Problems and solutions* (65-87). Frankfurt/Main: Lang, 1996

Gopher, D. & Donchin, E.: Workload - An examination of the concept. In: K.B. Boff; L. Kaufman & J.P. Thomas (eds.), *Handbook of perception and*



human performance, Volume II: Cognitive processes and performance (41/1-41/49). New York: Wiley, 1986

Hacker, W. & Richter, P.: *Psychische Fehlbeanspruchung. Psychische Ermüdung, Monotonie, Sättigung und Streß.* Berlin: Springer, 1984

Haider, E. & Rohmert, W.: Leistungs- und Beanspruchungskompensation bei informatorischer Arbeit. *Zeitschrift für Arbeitswissenschaft*, 32, 1-5, 1978

Hänecke, K.: *Antwortverhalten als Ergebnis einer aktiven Auseinandersetzung mit der zeitlichen Verteilung von Ereignissen. Analysen des Antwortverhaltens bei Vigilanzaufgaben.* Frankfurt/Main: Lang, 1995

Hecker, R.; Schmidtke, H. & Wegener, H.: *Reliabilität und Validität spektraler EEG-Parameter als Indikatoren der psychischen Beanspruchung.* Meisenheim: Anton Hain, 1980

Izsó, L.: *Developing evaluation methodologies for human-computer-interaction.* Delft: University Press, 2001

Jenkins, J.G.; Mitchell, R.H. & McClure, B.G.: Heart rate variability in the newborn infants. *Automedica*, 4, 263-270, 1983

Jorna, P.G.A.M.: Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology*, 34, 237-257, 1992

Kamphuis, A. & Frowein, H.W.: Assessment of mental effort by means of heart rate spectral analysis. In: J.F. Orlebeke, G. Mulder & L.J.P. van Doornen (eds.), *Psychophysiology of cardiovascular control* (841-853). New York: Plenum Press, 1985

Klimmer, F.: Verfahren zur Abschätzung der mentalen und emotionalen Belastung und Beanspruchung des arbeitenden Menschen. In: F. Nachreiner (Hrsg.), *Aktuelle Probleme der Belastungs- und Beanspruchungsforschung. Festschrift zum 60. Geburtstag von Joseph Rutenfranz* (212-258). Frankfurt/Main: Lang, 1988

Künstler, B.: Psychische Belastung durch die Arbeitstätigkeit - Theoretisches Rahmenkonzept der Entwicklung eines Fragebogens zum Belastungserleben. *Probleme und Ergebnisse der Psychologie*, 74, 45-67, 1980

Lasner, L.: *Herzfrequenzvariabilität als Beanspruchungsindikator bei Signalentdeckungsaufgaben mit automatisierter vs. kontrollierter Informationsverarbeitung. Eine experimentelle Untersuchung zur Diagnostizität der 0,10 Hz-Komponente*

der Herzfrequenzvariabilität (Diplomarbeit). Oldenburg: Carl von Ossietzky Universität, 1997

Lavie, P.: Sleep-wake as a biological rhythm. *Annual Review of Psychology*, 52, 277-303, 2001

Llabre, M.M.; Ironson, H.G.; Spitzer, S.B.; Gellman, M.D.; Weidler, D.J. & Schneiderman, N.: How many blood pressure measurements are enough? An application of generalizability theory to the study of blood pressure reliability. *Psychophysiology*, 25, 97-106, 1988

Luczak, H.: Psychophysiologische Methoden zur Erfassung psychophysiologischer Beanspruchungszustände. In: U. Kleinbeck & J. Rutenfranz (Hrsg.), *Arbeitspsychologie* (Enzyklopädie der Psychologie, D/III/1) (185-259). Göttingen: Verlag für Psychologie, Hogrefe, 1987

Luczak, H.: *Arbeitswissenschaft.* Berlin: Springer, 1998

Manzey, D.: Psychophysiologie mentaler Beanspruchung. In: F. Rösler (Hrsg.), *Ergebnisse und Anwendungen der Psychophysiology* (Enzyklopädie der Psychologie, C/1/7) (799-864). Göttingen: Hogrefe, 1998

Marcoulides, G.A.: Generalizability theory. In: H.E.A. Tinsley & S.D. Brown (eds.), *Handbook of applied multivariate statistics and mathematical modelling* (527-551). San Diego: Academic Press, 2000

Matthews, G.; Davies, D.R. & Holley, P.J.: Cognitive predictors of vigilance. *Human Factors*, 35, 3-24, 1993

Meshkati, N.; Hancock, P.A.; Rahimi, M. & Dawes, S.M.: Techniques in mental workload assessment. In: J.R. Wilson & E.N. Corlett (eds.): *Evaluation of human work. A practical ergonomics methodology* (749-782). London: Taylor & Francis, 1999

Moray, N.: Mental workload since 1979. *International Review of Ergonomics*, 2, 123-150, 1988

Mosso, A.: *Die Ermüdung* (aus dem Italienischen übersetzt von J. Glinzer). Leipzig: Hirzel, 1892

Mulder, G.: Sinusarrhythmia and mental workload. In: N. Moray (ed.), *Mental workload - Its theory and measurement* (327-343). New York: Plenum Press, 1979

Mulder, G. & Mulder, L.J.M.: Information processing and cardiovascular control. *Psychophysiology*, 18, 392-402, 1981

Mulder, G.; Mulder, L.J.M.; Meijman, T.F.; Veldman, J.B.P. & Van Roon, A.M.: A psychophysiological approach to working conditions. In:

R.W. Backs & W. Boucsein (eds.), *Engineering psychophysiology. Issues and applications* (139-159). Mahwah: LEA, 2000

Nachreiner, F.: Psychische Belastung und Beanspruchung: Normen und rechtliche Bestimmungen. *Zeitschrift für Arbeitswissenschaft*, 56, 2002, in diesem Heft.

Nachreiner, F. & Wucherpfennig, D.: Arbeits- und sozialpsychologische Aspekte der Arbeit unter Zeitdruck. *Betriebsärztliches*, 22-36, (12)1975

Nachreiner, F.; Nickel, P. & Yanagibori, R.: Analyse der Herzfrequenzvariabilität eines Busfahrers vor, während und nach einem kritischen Ereignis im Fahrdienst des ÖPNV. In: W. Hacker (Hrsg.), *Computer-gestützter Abstractband zum 41. Kongress der Deutschen Gesellschaft für Psychologie* (04.05-2). Dresden: Technische Universität, 1998

Nachreiner, F.; Grzech-Šukalo, H. & Möhlmann, D.: *Beteiligungsorientierte und sozialverträgliche Arbeitszeit- und Pausengestaltung im Stadt- und Regionalverkehr (SuRV)* (unter Mitarbeit von P. Nickel und K. Trauernicht). Stuttgart: ÖTV, 1999

Nickel, P.: *Sensitivität und Diagnostizität der 0,1 Hz-Komponente der Herzfrequenzvariabilität zur Erfassung psychischer Beanspruchung*, Dissertation. Oldenburg: Carl von Ossietzky Universität, 2001. (im Druck)

Nickel, P.; Nachreiner, F.; Zdobyh, A. & Yanagibori, R.: Zur Beurteilung der psychischen Beanspruchung mit Hilfe der 0,1 Hz-Komponente der Herzfrequenzvariabilität - einige methodische und analysetechnische Probleme. *Zeitschrift für Arbeits- und Organisationspsychologie*, 42, 205-212, 1998

Nickel, P.; Seehase, L. & Nachreiner, F.: Die 0,1 Hz-Komponente der HFV zur Indikation psychischer Beanspruchung unterschiedlicher Belastungssituationen oder unterschiedlicher Personen? In: K. Moser & R. Soucek (Hrsg.), *2. Tagung der Fachgruppe Arbeits- & Organisationspsychologie der Deutschen Gesellschaft für Psychologie (DGPs)*, 19.-21.09.2001, Nürnberg, Abstractband (158). Erlangen-Nürnberg: Friedrich-Alexander Universität, 2001

Nußbaum, A.: Das Modell der Generalisierbarkeitstheorie. In: K.J. Klauer (Hrsg.), *Kriteriumsorientierte Tests* (114-136). Göttingen: Hogrefe, 1987

Obrist, P.A.: *Cardiovascular psychophysiology. A perspective.* New York: Plenum Press, 1981

Opmeer, C.H.J.M.: The information content of successive RR-interval times in the ECG. Preliminary results using factor analysis and frequency analysis. *Ergonomics*, 16, 105-112, 1973

Paas, F.G.W.C.; van Merriënboer, J.J.G. & Adam, J.J.: Measurement of cognitive load in instructional

research. *Perceptual and Motor Skills*, 79, 419-430, 1994

Pitrella, F.D.: *Sequential judgement scale: Cognitive model, scale development, use and validation (Be-richt Nr. 491)*. Wachtberg-Werthhoven: FGAN-FAT, 1998

Pruyn, A.T.H.: *Performance and activation under social control*. Delft: Eburon, 1986

Rutenfranz, J.: Über das Verhalten der Pulsfrequenz bei Arbeit unter Zeitdruck. *Internationale Zeitschrift für angewandte Physiologie einschließlich Arbeitsphysiologie*, 18, 264-279, 1960

Schmidt, R.F. & Thews, G.: *Physiologie des Menschen*. Berlin: Springer, 2000

Schmidtke, H.: *Die Ermüdung. Symptome, Theorien, Messversuche*. Bern: Huber, 1965

Schmidtke, H.: Mentale Beanspruchung durch in-formationelle Belastung. In: H. Schmidtke (Hrsg.), *Ergonomie*. (143-160). München: Hanser, 1993

Schmidtke, H.: Vom Sinn und Unsinn der Messung psychischer Belastung und Beanspruchung. *Zeit-*

schrift für Arbeitswissenschaft, 56, 2002, in diesem Heft

Schütte, M.: Zur Modifikation des Belastungs-verlaufstests (BLV). *Zeitschrift für Arbeits- und Organisationspsychologie*, 32, 188-192, 1988

Schütte, M. & Nickel, P.: Multivariate Analyse der bedingungsbezogenen Messgenauigkeit von Beanspruchungsindikatoren. *Zeitschrift für Arbeitswissenschaft*, 56, 2002, in diesem Heft.

Sharit, J. & Salvendy, G.: External and internal attentional environments. II. Reconsideration of the relationship between sinus arrhythmia and information load. *Ergonomics*, 25, 121-132, 1982

Van Dellen, H.J.; Aasman, J.; Mulder, L.J.M. & Mulder, G.: Time domain versus frequency domain measures of heart-rate variability. In: J.F. Orlebeke, G. Mulder & L.J.P. van Doornen (eds.), *Psychophysiology of cardiovascular control. Models, methods, and data* (353-374). New York: Plenum Press, 1985

Veltman, J.A. & Gaillard, A.W.K.: Physiological indices of workload in a simulated flight task. *Biological Psychology*, 42, 323-342, 1996

Wierwille, W.W. & Eggemeier, F.T.: Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors*, 35, 263-281, 1993

Wiethoff, M.: *Task analysis is heart work. The investigation of heart rate variability: A tool for task analysis in cognitive work*. Delft: Delft University Press, 1997

Wilson, G.F.; Swain, C.R. & Brookings, J.B.: *Workload related changes in eye, cardiac, respiratory and brain activity during simulated air traffic control (U)* (AL/CF-TR-1995-0156, ADA306676). Springfield: National Technical Information Service, 1995

ANSCHRIFT DER VERFASSER

Dipl.-Psych. Peter Nickel

Dr. Karin Eilers

Dipl.-Psych. Liane Seehase

Prof. Dr. Friedhelm Nachreiner

Carl von Ossietzky Universität Oldenburg

AE Arbeits- & Organisationspsychologie

26111 Oldenburg

E-Mail:peter.nickel@uni-oldenburg.de