

Echtzeiterfassung psychischer Beanspruchungszustände

Yannick FUNK, Henrike HAASE, Barbara DEML

*Institut für Arbeitswissenschaft und Betriebsorganisation
Karlsruher Institut für Technologie
Engler-Bunte-Ring 4, D-76131 Karlsruhe*

Kurzfassung: In dieser Arbeit wird eine Probandenstudie (N=21) zur Generierung von Trainingsdaten und die Erarbeitung eines Algorithmus zur Echtzeitbewertung psychischer Beanspruchungszustände vorgestellt. Die Versuchspersonen nahmen zweimal im Abstand von mindestens drei Wochen an der Studie teil. Dabei wurden 18 okulare und zwei kardiovaskuläre Beanspruchungsindikatoren erhoben und die subjektive Beanspruchung mittels Rating Scale Mental Effort (RSME) abgefragt. Aus den Daten des ersten Termins wurde individuell für jede Versuchsperson ein Algorithmus trainiert, mit dessen Hilfe die subjektiv empfundene Beanspruchung, repräsentiert durch klassifizierte RSME-Werte (gering, mittel & hoch), vorhergesagt und mit den tatsächlichen Bewertungen verglichen wurde. Die Modelle erzielten eine mittlere Vorhersagegenauigkeit von 57.39% (Spannweite = 47.62% - 71.43%).

Schlüsselwörter: psychische Beanspruchung, Rating Scale Mental Effort, mental workload, maschinelles Lernen, adaptive Schnittstellen

1. Hintergrund

Im BMBF-geförderten Verbundvorhaben Fahrerkabine 4.0 wird eine adaptive Mensch-Maschine-Schnittstelle am Beispiel eines Mähdreschers entwickelt. Diese ist in der Lage, mithilfe physiologischer Daten das aktuelle Beanspruchungsniveau der Fahrerinnen und Fahrer zu erkennen. Basierend darauf werden situationsgerechte Handlungsempfehlungen abgeleitet. Bei geringer Beanspruchung, wie sie z. B. bei der automatisierten Ernte auftritt, können Bürotätigkeiten erledigt werden, die ansonsten noch nach dem Dreschen anfallen würden. Bei hoher Beanspruchung können unkritische Informationen oder Teilaufgaben zurückgehalten werden.

Beanspruchung wird dabei definiert als eine individuelle Reaktion auf externe Belastungen (Rohmert, 1984). Die psychische Beanspruchung ist nicht nur eine Funktion der Belastung, sondern hängt auch von spezifischen Fähigkeiten und Eigenschaften der arbeitenden Menschen ab (Schlick et al, 2018; DIN EN ISO 10075-1). Eine gleiche Belastung führt damit bei unterschiedlichen Personen zu verschieden hoher Beanspruchung. Aus dieser Erkenntnis ergibt sich die Forderung, Arbeitsplätze an ihre individuellen Benutzer und Benutzerinnen anzupassen, um deren Gesundheit und Leistungsmotivation zu erhalten und schließlich die Effizienz des Arbeitssystems zu erhöhen.

Diese Arbeit geht von De Waards (1996) Zusammenhang zwischen Leistung und psychischer Beanspruchung und Wickens (2002) Multipler Ressourcen Theorie aus. Hiernach sollte es möglich sein, ein Beanspruchungsspektrum von Unterforderung über optimale Belastung bis hin zur Überforderung zu induzieren und die individuellen physiologischen Reaktionen auf die Belastung zu messen. Ziel dieser Arbeit ist es, mit

einem zuvor validierten Versuchsaufbau (Funk et al., 2021) ein System zur Echtzeiterfassung psychischer Beanspruchungszustände zu realisieren.

2. Methode

Unsere Experimentalumgebung induziert psychische Belastung über eine bildschirmbasierte (1920x1080 Pixel, 22 Zoll) Überwachungstätigkeit in fünf randomisierten Geschwindigkeitsstufen, abschnittsweise ergänzt um eine visuelle und eine auditive Nebenaufgabe. Die subjektive Beanspruchung wird gemäß der Rating Scale Mental Effort bewertet (RSME; vgl. Zijlstra 1985). Für eine detaillierte Beschreibung der Experimentalumgebung siehe Funk et al. (2021). Alle Berechnungen wurden in Matlab R2021a durchgeführt.

2.1 Physiologische Indikatoren

In dieser Arbeit wurden zwei kardiovaskuläre und 18 okulare Parameter erhoben (vgl. Tabelle 1). Die Auswahl der Messgeräte erfolgte anhand ihrer Eignung für einen zukünftigen Einsatz in der Fahrerkabine 4.0. Ausschlaggebende Kriterien dabei waren die Genauigkeit der Geräte und eine geringe Beeinträchtigung der Arbeitstätigkeit.

Tabelle 1: *Physiologische Parameter (für die mit einem Fragezeichen markierten Indikatoren konnte keine Verbindung zu psychischer Beanspruchung, weder eine Korrelation noch eine Ablehnung als Indikator, in der Literatur gefunden werden.)*

Indikator	Korrelation mit psychischer Beanspruchung	Quelle
Herzfrequenz	+	Backs & Selios (1994)
Herzfrequenzvariabilität	-	Schlick et al. (2018)
Ø-Fixationsdauer	+	Maltz & Shinar (1999)
Fixationsrate	+	Maltz & Shinar (1999)
SD der Fixationsdauer	?	
Kumulierte Fixationsdauer	?	Pataki et al. (2005)
Ø-Sakkadendauer	-	Schneider (2019)
Sakkadenrate	-	Abich et al. (2013)
SD der Sakkadendauer	?	
Kumulierte Sakkadendauer	?	
Ø-Sakkadengeschwindigkeit	-	Di Stasi et al. (2010)
SD der Sakkadengeschwindigkeit	?	
Ø-Blinzeldauer	-	Schneider (2019)
Blinzelrate	-	Schneider (2019)
SD der Blinzeldauer	?	
PERCLOS	+	Dinges & Grace (1998)
Spektralverhältnis der Pupillenvariabilität	-	Peysakhovich et al. (2015)
Index of Cognitive Activity	+	Marshall (2000)
Index of Pupillary Activity	+	Duchowski et al. (2018)
Nearest Neighbor Index	+	Di Nocera et al. (2006)

Eye-Tracking-Parameter wurden mit einem zwei-Kamera-System, Smart Eye Pro dx (2021) und die Herzfrequenz mit einem kommerziellen Fitnesstracker, Fitbit Charge 3 (2021) erhoben. Die Genauigkeit des Fitnesstrackers wurde vorab in Gegenüberstellung mit einem Drei-Punkt-Elektrokardiogramm (Varioport, 2021) evaluiert.

2.2 Stichprobe

An dieser Studie nahmen 21 Personen (9 Frauen, 12 Männer) im Alter von 18 bis 44 Jahren ($M = 25.6$, $SD = 5.35$) teil. Die Stichprobe wurde aus Studierenden und Mitarbeitenden des KIT rekrutiert. Alle Versuchspersonen wurden vorab über den Untersuchungsablauf, mögliche Risiken, ihre Rechte und die Anonymität der Daten informiert und gaben ihr Einverständnis. Es wurde darauf geachtet, dass die Versuchspersonen keine Einschränkungen des Sehvermögens aufwiesen bzw. eine geeignete Sehhilfe für die Experimente benutzt wurde. Die Studie fand zwischen Juli und September 2021 statt und wurde durch die Ethikkommission des KIT genehmigt.

2.3 Versuchsablauf

Die Probanden und Probandinnen wurden gebeten, eine Versuchsbeschreibung, eine Datenschutzerklärung und eine Einverständniserklärung zu lesen und zu unterschreiben. Nach Erhebung der versuchsrelevanten demografischen Daten (Geschlecht, Alter, Sehhilfen) nahmen die Versuchspersonen vor einem Bildschirm Platz. Die Teilnehmenden wurden angewiesen, eine korrekte Antwort einer möglichst schnellen Antwort vorzuziehen. Die Versuchspersonen wurden randomisiert in zwei Gruppen eingeteilt. Ziel dieser Aufteilung war zu beantworten, ob die Belastungsdauer (60Sek, bzw. 120Sek) je Belastungsstufe einen Einfluss auf die Datenreliabilität und somit auf die Vorhersagegenauigkeit der Modelle hat.

Die Studie begann mit einer Übungsphase, in der alle drei Aufgabenteile (Hauptaufgabe, visuelle Nebenaufgabe und auditive Nebenaufgabe) separat bearbeitet wurden. Die Übungszeit für die Hauptaufgabe und für die visuelle Nebenaufgabe betrug jeweils zwei Minuten. Die Übung zur auditiven Nebenaufgabe bestand aus drei Fragen, die über Kopfhörer präsentiert wurden. Nach der Übungsphase wurde eine einminütige Baseline erhoben. In dieser Zeit wurde eine ruhige Mähdrescherfahrt, ohne Handlungsanforderung an die Versuchspersonen dargeboten, physiologische Maße erhoben und im Anschluss eine RSME-Bewertung abgefragt. Der Hauptteil des Experiments bestand aus einer Folge vordefinierter Kombinationen von Hauptaufgabe, visueller und auditiver Nebenaufgabe. Jede Kombination wurde je nach Gruppe 60 oder 120 Sekunden lang angezeigt.

Während der *Hauptaufgabe* wurde den Versuchspersonen ein Video einer Mähdrescherfahrt aus der Egoperspektive gezeigt. Während des Videos wurden randomisiert die Buchstaben „W“, „A“ oder „D“ in fünf Frequenzstufen auf dem Bildschirm eingeblendet (alle 15, 10, 5, 3 oder 2 Sekunden). Die Versuchspersonen mussten die angezeigten Buchstaben auf einer Computertastatur mit der linken Hand drücken.

Bei der *visuellen Nebenaufgabe* handelt es sich um eine Mentale Rotation (Eggemeier & Wilson, 1991). Dabei wurden im unteren rechten Bildschirmquadranten vier verschiedene Figurenpaare in einem Abstand von drei Sekunden randomisiert angezeigt. Die Figuren waren entweder nur zueinander gedreht (randomisierter Winkel) oder zusätzlich gespiegelt. Die Versuchspersonen mussten mit der rechten Hand anzeigen, ob es sich um eine Drehung (linke Pfeiltaste) oder um eine Drehung mit Spiegelung (rechte Pfeiltaste) handelte.

Die *auditive Nebenaufgabe* bestand aus zehn mal 13 Fragen, die den Versuchspersonen über Kopfhörer präsentiert werden. Die Antworten der Probanden und Probandinnen wurden dabei über ein externes Mikrofon aufgezeichnet. Die einzelnen Fragen wurden in vier Schwierigkeitsklassen eingeteilt (vgl. u.a. Batliner et

al. 2006). Beispiele sind: (1) Sehr gering: Wiederholung einfacher, kurzer Sätze. (2) Gering: Wiederholung zusammenhangsloser Silben. (3) Mittel: Einfache mathematische Operationen. (4) Hoch: Fragen zum Allgemeinwissen. Jeder Block enthielt drei Fragen sehr geringer, vier Fragen geringer, vier Fragen mittlerer und zwei Fragen hoher Schwierigkeit.

Nach jeder Aufgabenkombination wurden die Teilnehmenden gebeten, ihre subjektiv empfundene Beanspruchung auf einer Skala von 0 bis 150 abzugeben (RSME). Der Versuch wurde je Proband und Probandin zwei mal mit einem Abstand von mindestens drei Wochen durchgeführt. Je Versuchsperson ergaben sich somit 21 Trainingsdatenpaare (erster Termin) und 21 Validierungsdatenpaare (zweiter Termin), bestehend aus physiologischen Messdaten und korrespondierenden RSME-Bewertungen (20 einzigartige Aufgabenkombinationen und Baseline).

2.4 Maschinelles Lernen

In dieser Arbeit dienen klassifizierte RSME-Werte als die vorherzusagende Antwortvariable und die 20 physiologischen Indikatoren als Prädiktorvariablen. Ziel der RSME-Klassifizierung ist es, jeder der 21 Situationen einen Belastungsgrad zuzuordnen. Nach Zijlstra (1985) gilt eine Tätigkeit ab einer RSME-Bewertung von 37 als *etwas anstrengend* und ab 71 als *ziemlich anstrengend*. Anhand dieser Grenzwerte und der durchschnittlichen RSME-Bewertung wurde jede Situation einer der Belastungsklassen *gering* (unter 37), *mittel* (zwischen 37 & 71) und *hoch* (ab 71) zugeordnet. Diese Klassifikation wurde getrennt für die Bewertungen des ersten und des zweiten Termins durchgeführt.

Um eine Überanpassung („overfitting“) des Modells zu vermeiden, wurden die 20 Prädiktorvariablen individuell je Versuchsperson zunächst mittels Chi-Quadrat-Tests nach ihrer Relevanz sortiert. Alle Prädiktoren mit einem Relevanz-Wert kleiner eins wurden ausgeschlossen (vgl. Matlab, 2021). Die verbleibenden Indikatoren wurden auf Kollinearität geprüft und Prädiktoren mit starkem Zusammenhang ausgeschlossen (vgl. Belsley et al., 1980). Aus allen übrig gebliebenen Indikatoren wurde, zusammen mit den klassifizierten RSME-Bewertungen, je Versuchsperson ein Naive Bayes Algorithmus (Hastie et al., 2009) trainiert. Somit ergaben sich 21 Beanspruchungs-Modelle. Die Validität dieser Modelle wurde anschließend über den Validierungsdatensatz geprüft. Hierzu wurden anhand der Modelle und Prädiktorvariablen des zweiten Termins klassifizierte RSME-Bewertungen vorhergesagt und mit den tatsächlichen Bewertungen des zweiten Termins verglichen.

3. Ergebnisse

Die Vorhersagegenauigkeiten der entwickelten Modelle sind in Tabelle 2 dargestellt. Modellgenauigkeit bezeichnet dabei das Verhältnis der korrekt vorhergesagten zu den tatsächlich abgegeben (klassifizierten) RSME-Bewertungen des zweiten Termins. Die durchschnittliche Vorhersagegenauigkeit der individuellen Modelle beträgt 57.39% mit einer Spannweite von 47.62% bis 71.43%. Eine zufällige Zuordnung von Beanspruchungsbewertung zu Belastungssituation würde eine Vorhersagegenauigkeit von 33.34% ergeben. Es lässt sich eine Verbesserung der Beanspruchungsvorhersage gegenüber dem Zufall um durchschnittlich 24.05 Prozentpunkte verzeichnen. Ebenso erzielten alle individuellen Modelle eine besser als zufällige Vorhersagegenauigkeit (vgl. Tabelle 2).

Tabelle 2: Vorhersagegenauigkeit: Übereinstimmung der Modellvorhersage mit der tatsächlichen Bewertung.

Versuchsperson	Modellgenauigkeit	Versuchsperson	Modellgenauigkeit
VP 1	61.90%	VP 12	71.43%
VP 2	57.14%	VP 13	66.67%
VP 3	47.62%	VP 14	52.38%
VP 4	66.67%	VP 15	47.62%
VP 5	66.67%	VP 17*	52.38%
VP 6	52.38%	VP 18	66.67%
VP 7	47.62%	VP 19	42.86%
VP 8	66.67%	VP 20	52.38%
VP 9	71.43%	VP 21	57.14%
VP 11*	66.67%	Mittelwert	57.39%

*die Daten der VPs 10 & 16 sind beschädigt und konnten nicht ausgewertet werden.

4. Diskussion und Ausblick

Mit den vorgestellten Messsystemen und errechneten Modellen wurde ein Algorithmus entwickelt, der in nahezu Echtzeit (maximale Verzögerung = 120 Sek.) den Beanspruchungszustand einer vorab kalibrierten Person (Erstellung eines individuellen Modells) anzeigen kann.

Unsere zukünftige Forschung beschäftigt sich mit der Verbesserung der Vorhersagegenauigkeit und der Übertragung des Messsystems in eine Fahrsimulation. Ab September 2021 wurden weitere Versuchspersonen erhoben, mit dem Ziel, die Datengrundlage zu stärken und falls möglich, Cluster innerhalb der Stichprobe zu bilden. Über solche Cluster könnten „Gruppenmodelle“ mit einem erheblich größeren Trainingsdatensatz als die individuellen Modelle berechnet werden und die Vorhersagegenauigkeit verbessert werden. Ab März 2022 wird eine Kombination aus Fitnesstracker und Smart Eye System in einem Mähdrescher-Fahrsimulator verbaut, mit dem Ziel die Übertragbarkeit der Modelle und Vorhersagen auf tatsächliche Arbeitssituationen in Landmaschinen zu evaluieren.

Zusammenfassend ist festzustellen, dass unsere Beanspruchungsmodelle vielversprechende erste Ergebnisse liefern und, ergänzend zu Maschinendaten, Leistungsbewertung und Nutzerfeedback ein Standbein adaptiver, beanspruchungs-basierter Mensch-Maschine-Schnittstellen sein können.

5. Literatur

- Abich IV, J, Reinerman-Jones L, Taylor G S (2013, September) Investigating workload measures for adaptive training systems. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 57, No. 1, pp. 2091-2095). Sage CA: Los Angeles, CA: SAGE Publications.
- Backs R W, Seljos K A (1994) Metabolic and cardiorespiratory measures of mental effort: the effects of level of difficulty in a working memory task. International journal of psychophysiology, 16(1), 57-68.
- Batliner A, Steidl S, Schuller B, Seppi D, Laskowski K, Vogt T, Aharonson V (2006) Combining efforts for improving automatic classification of emotional user states.
- Belsley D A, Kuh E, Welsch R E (1980) Regression diagnostics: Identifying influential data and sources of collinearity. Wiley Series in Probability and Mathematical Statistics.
- De Waard D (1996) The measurement of drivers' mental workload. Traffic Research Center, Groningen, Niederlande.

- Dinges D F, Grace R (1998) PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. US Department of Transportation, Federal Highway Administration, Publication Number FHWA-MCRT-98-006
- DIN EN ISO 10075-1, 2018-01-00: Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung - Teil 1: Allgemeine Aspekte und Konzepte und Begriffe (ISO 10075- 1:2017); Deutsche Fassung EN ISO 10075-1:2017.
- Di Nocera F, Camilli M, Terenzi M (2006) Using the distribution of eye fixations to assess pilots' mental workload. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 50, No. 1, pp. 63-65). Sage CA: Los Angeles, CA: SAGE Publications.
- Di Stasi L L, Renner R, Staehr P, Helmert J R, Velichkovsky B M, Cañas J J, ... & Pannasch S (2010) Saccadic peak velocity sensitivity to variations in mental workload. *Aviation, space, and environmental medicine*, 81(4), 413-417.
- Duchowski A T, Krejtz K, Krejtz I, Biele C, Niedzielska A, Kiefer P, ... & Giannopoulos I (2018, April) The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In: Proceedings of the 2018 CHI conference on human factors in computing systems (pp. 1-13).
- Eggemeier T F, Wilson G F (1991) Performance-based and subjective assessment of workload in multi-task environments. In: D. L. Damos (Hrsg.), *Multiple-task performance* (S. 217–275). Taylor & Francis, London and Washington.
- Fitbit Charge 3 (2021). Abgerufen am 15. November, 2021. <https://www.fitbit.com/global/de/products/trackers/charge3>.
- Funk Y, Haase H, Deml B (2021) Entwicklung und Validierung einer Experimentalumgebung zur Messung mentaler Beanspruchungszustände. In: GfA, Dortmund (Hrsg.): Tagungsband 67. GfA-Frühjahrskongress Arbeit HUMAINE gestalten, Beitrag B.14.4, Bochum 03.03.-05.03.2021.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics. New York, NY: Springer. S.210 ff.
- Maltz M, Shinar, D (1999) Eye movements of younger and older drivers. *Human factors*, 41(1), 15-25.
- Marshall S P (2000) U.S. Patent No. 6,090,051. Washington, DC: U.S. Patent and Trademark Office.
- Matlab (2021). Abgerufen am 03. September, 2021. <https://de.mathworks.com/help/stats/fscchi2.html>.
- Pataki K, Schulze-Kissing D, Mahlke S, Thüning M (2005) Anwendung von Usability-Maßen zur Nutzeneinschätzung von Fahrerassistenzsystemen. *Beiträge zur Mensch-Maschine-Systemtechnik aus Forschung und Praxis*, 211-228.
- Peysakhovich V, Causse M, Scannella S, Dehais F (2015) Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort. *International Journal of Psychophysiology*, 97(1), 30-37.
- Rohmert W (1984) Das Belastungs-Beanspruchungs-Konzept. *Zeitschrift für Arbeitswissenschaft*, 38(4), 193-200.
- Schlick C, Bruder R, Luczak H (2018) *Arbeitswissenschaft*. Springer-Verlag.
- Schneider M (2019) Blickbasierte Beanspruchungsmessung: Entwicklung und Evaluation eines Kalibrierungssystems zur individuellen Bewertung der mentalen Beanspruchung in der Mensch-Technik-Interaktion (Vol. 3). KIT Scientific Publishing.
- Smart Eye Pro dx (2021). Abgerufen am 15. November, 2021. <https://smarteys.se/smart-eye-pro-new-features-and-demo/>.
- Varioport (2021). Abgerufen am 15. November, 2021. https://bisigma.de/pdf/varioport-b_english.pdf.
- Wickens C D (2002) Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2), 159-177.
- Zijlstra F R H, Van Doorn L (1985) The construction of a scale to measure perceived effort. Technical report, Delft University Press.

Danksagung: Dieser Beitrag entstand im Rahmen des vom Bundesministerium für Bildung und Forschung (BMBF) gefördert Verbundvorhabens „Fahrerkabine 4.0“. Das Forschungsvorhaben wird vom Projektträger Jülich (PTJ) betreut.



Gesellschaft für
Arbeitswissenschaft e.V.

Technologie und Bildung in hybriden Arbeitswelten

68. Kongress der
Gesellschaft für Arbeitswissenschaft

Otto-von-Guericke-Universität Magdeburg

Fraunhofer-Institut für Fabrikbetrieb und
Fabrikautomatisierung IFF, Magdeburg

02. – 04. März 2022

GfA-Press

Bericht zum 68. Arbeitswissenschaftlichen Kongress vom 02. – 04. März 2022

Otto-von Guericke-Universität Magdeburg;

Fraunhofer-Institut für Fabrikbetrieb und -automatisierung IFF, Magdeburg

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.

Sankt Augustin: GfA-Press, 2022

ISBN 978-3-936804-31-7

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle (s. u.) erhältlich.

Alle Rechte vorbehalten.

© **GfA-Press, Sankt Augustin**

Schriftleitung: Prof. Dr. Rolf Ellegast

im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Ohne ausdrückliche Genehmigung der Gesellschaft für Arbeitswissenschaft e.V. ist es nicht gestattet:

- den Kongressband oder Teile daraus in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) zu vervielfältigen,
- den Kongressband oder Teile daraus in Print- und/oder Nonprint-Medien (Webseiten, Blog, Social Media) zu verbreiten.

Die Verantwortung für die Inhalte der Beiträge tragen alleine die jeweiligen Verfasser; die GfA haftet nicht für die weitere Verwendung der darin enthaltenen Angaben.

Geschäftsstelle der GfA

Simone John, Tel.: +49 (0)30 1300-13003

Alte Heerstraße 111, D-53757 Sankt Augustin

info@gesellschaft-fuer-arbeitswissenschaft.de · www.gesellschaft-fuer-arbeitswissenschaft.de

Screen design und Umsetzung

© 2022 fröse multimedia, Frank Fröse

office@internetkundenservice.de · www.internetkundenservice.de