

Evaluation einer multimodalen Mensch-Roboter-Schnittstelle

Marc SCHNEIDER, Barbara DEML

*Institut für Arbeitswissenschaft und Betriebsorganisation,
Karlsruher Institut für Technologie
Kaiserstraße 12, D-76131 Karlsruhe*

Kurzfassung: Dieser Beitrag berichtet über die Ergebnisse einer experimentellen Evaluation einer multimodalen Mensch-Roboter-Schnittstelle. Es wurden zwei Anwendungsszenarien aus der Roboterprogrammierung untersucht, die jeweils mit drei verschiedenen Eingabemodalitäten ausgeführt wurden. Die mentale Beanspruchung der Operateure wurde neben subjektiven Parametern auch über physiologische und objektive Indikatoren operationalisiert. Es zeigten sich signifikante Unterschiede hinsichtlich der mentalen Beanspruchung zwischen den Eingabemodalitäten in beiden Anwendungsszenarien. Die vorliegenden Ergebnisse verdeutlichen den Nutzen einer adaptiven Systemgestaltung in der Mensch-Roboter-Interaktion.

Schlüsselwörter: Mensch-Roboter-Interaktion, Beanspruchungsmessung, adaptive Systemgestaltung

1. Motivation

Aufgrund immer weiter steigender Dynamik im industriellen Alltag zeigt sich ein Trend weg von schweren Industrierobotern mit hohem Gefahrenpotential (Neudörfer 2014) hin zu Mensch-Roboter-Kooperationen unter Verwendung von Leichtbaurobotern (Spillner 2015). Eine besondere Herausforderung stellt dabei die Programmierung solcher Robotersysteme dar, die heute vielfach noch sehr komplex ist. Diese Tatsache verschärft sich vor dem Hintergrund des zunehmenden Fachkräftemangels und des demografischen Wandels. Aus dem Grund macht es Sinn, adaptive Schnittstellen vorzusehen, welche die individuellen Leistungsvoraussetzungen der Nutzer berücksichtigen (Fong et al. 2003). Die Evaluation verschiedener Eingabemodalitäten bezüglich der mentalen Beanspruchung im Rahmen dieser Arbeit stellt in diesem Zusammenhang eine wichtige Voraussetzung dar.

2. Methoden

Mentale Beanspruchung beschreibt die kognitive Reaktion des menschlichen Informationsverarbeitungssystems auf äußere Belastungen in Abhängigkeit von seinen persönlichen Voraussetzungen und den individuellen Bewältigungsstrategien (Packebusch 2003). Zur Messung des Konstrukts hat sich ein trimodaler Ansatz bewährt, bei dem gleichzeitig physiologische (siehe 2.1), subjektive (siehe 2.2) und objektive Leistungsparameter (siehe 2.3) erfasst werden.

2.1 Physiologische Parameter

Physiologische Parameter nutzen Reaktionen des menschlichen Körpers, um kontinuierlich und in hoher zeitlicher Auflösung auf die mentale Beanspruchung zu schließen (Manzey 1998). In dieser Studie werden sechs blickbasierte Parameter erfasst: Die *Blinzelrate* beschreibt die Anzahl des Blinzeln in einem definierten Zeitintervall von 30 Sekunden (Wilson 2009). Einige Studien zeigten erfolgreich den Zusammenhang zwischen der Blinzelrate und der mentalen Beanspruchung (Cardona & Quevedo 2014; Recarte et al. 2008). Die *Blinzeldauer* errechnet sich aus der gemittelten Dauer eines Blinzelvorgangs, wohingegen der Parameter *PERCLOS* die kumulierten Blinzeldauern in einem definierten Zeitintervall von 30 Sekunden beschreibt. Die Sensitivität der Blinzeldauer sowie des PERCLOS für die mentale Beanspruchung wurden bereits mehrfach gezeigt (Benedetto et al. 2011; Brookhuis & de Waard 2010). Ein weiterer Parameter ist die *Pupillendurchschnittsvariabilität*, die in vielen Studien als zuverlässiger Indikator für die mentale Beanspruchung bei unterschiedlichsten Anwendungen eingesetzt wurde (Demberg 2013; Schwalm 2009). Die Berechnung basiert auf dem Index of Cognitive Activity (ICA) nach Marshall (2000), welcher die psycho-sensorischen Reflexe der Pupille mittels einer Wavelet-Transformation erfolgreich von den Licht- und Adaptionsreflexen der Pupille trennt (Loewenfeld 1993). Darüber hinaus wird in dieser Studie der *Nearest Neighbor Index (NNI)* ausgewertet. Der NNI ist ein statistisches Maß für die Verteilung von Blickfixationen im Blickfeld, wobei die tatsächliche Verteilung ins Verhältnis zu einer rein zufälligen Verteilung gesetzt wird. Daraus ergeben sich drei mögliche Verteilungsformen – Gruppierungen, Regularität und zufällige Anordnung – welche mit unterschiedlich hohen mentalen Beanspruchungsniveaus verbunden werden (Clark & Evans 1954; Di Nocera et al. 2007). Letztlich wird auch die *Fixationsdauer* ausgewertet. Studien, unter anderem von Recarte und Nunes (2000), zeigten signifikante Zusammenhänge zwischen der mentalen Beanspruchung und der durchschnittlichen Fixationsdauer. Aktuelle Studien deuten darauf hin, dass die Zuverlässigkeit der sechs Parameter individuell verschieden ist. Gründe dafür liegen in den bewussten und unbewussten Anpassungen der Blickstrategien einzelner Personen (für eine ausführlichere Darstellung siehe Schneider & Deml in Druck).

2.2 Subjektives Empfinden

Zur Erfassung des subjektiven Beanspruchungsempfindens wurde in dieser Studie der NASA Task-Load-Index (NASA-TLX) nach Hart & Staveland (1988) verwendet. Der Fragebogen wurde hier mit fünf Items auf einer 20-stufigen Skala abgefragt: geistige Anforderungen, zeitliche Anforderungen, Leistung, Anstrengung und Frustration.

2.3 Leistungsdaten

Die Leistung in einer Nebenaufgabe ist ein guter Indikator für die mentale Beanspruchung (Wickens et al. 2013). In dieser Studie wurde dazu eine visuelle 1-back Nebenaufgabe eingesetzt, die gleichzeitig zur Bearbeitung der Hauptaufgabe (siehe Abschnitt 3) präsentiert wurde (Kane et al. 2007). Die Probanden bekamen alle vier Sekunden eine Buchstaben-Zahlen-Kombination präsentiert (z. B. A9 oder B3). Immer, wenn eine neue Kombination angezeigt wurde, sollten die Probanden dann verbal die vorangegangene Kombination nennen.

3. Experimentalaufbau

Für die Evaluation der multimodalen Mensch-Roboter-Schnittstelle wurden zwei typische Anwendungsszenarien aus der Roboterprogrammierung mit jeweils drei Eingabemodalitäten untersucht. Hier sind zum einen die Bahnplanung und zum anderen die Bahnkorrektur zu nennen. Vor dem Versuch wurden die Probanden eingehend mit Trainingsdurchgängen in die Steuerung des Roboters eingewiesen. Im ersten Anwendungsszenario sollten die Probanden einem Industrieroboter eine vordefinierte Schweißbahn einlernen. Die Bahn bestand dabei aus sieben Konturpunkten, die nacheinander eingelernt werden sollten. Das zweite Anwendungsszenario umfasste die Kontrolle und Korrektur der eingelernten Schweißbahn. Dazu sollten die Probanden die Bahn abfahren, den vierten der sieben definierten Punkte löschen und die Bahn erneut abfahren. Bei der verwendeten Schweißbahn handelte es sich um eine abgewandelte Form von Schweißprozessen auf Karosserieteilen in der Automobilfertigung (Abbildung 1).

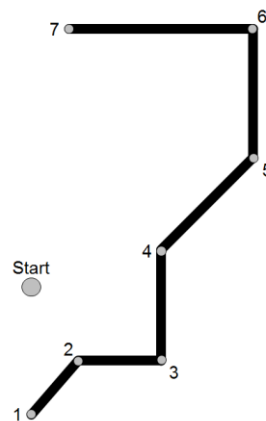


Abbildung 1: Verwendete Schweißbahn in Anwendungsszenario 1 und 2.

Beide Anwendungsszenarien wurden mit jeweils drei Steuerungsmodalitäten durchgeführt, wobei die Bearbeitungsreihenfolge randomisiert wurde. Bei der Sprach- und der Gestensteuerung nutzte der Proband definierte Sprach- beziehungsweise Gestenbefehle zum Speichern und Löschen von Punkten sowie zum Anfahren von bereits gespeicherten Punkten. Im Falle der Bahnplanung wurde der Roboter mittels eines Kraftmomentensensors (KMS) zu den neuen Punkten verfahren und diese wurden dann mit einer Sprach- beziehungsweise Gesteneingabe gespeichert. Die dritte Steuerungsmodalität war ein Tablet-PC, mit dem die Bahnplanung und -korrektur über eine grafische Oberfläche durchgeführt wurde. Während der Aufgabenbearbeitung trugen die Probanden ein kopfbasiertes Blickerfassungssystem (Ergoneers Dikablis Professional) und sie bearbeiteten zudem kontinuierlich die Nebenaufgabe. Nach jedem Versuchsdurchgang wurde der NASA-TLX gereicht, so dass der Fragebogen pro Proband insgesamt sechsmal (2 Aufgaben x 3 Eingabemodalitäten) ausgefüllt wurde.

4. Datenauswertung

Der Untersuchung liegt eine überwiegend studentische Stichprobe mit $N = 19$ Probanden zugrunde, die im Mittel 24,14 Jahre alt waren ($SD = 3,25$). Es wurden 16

männliche und drei weibliche Probanden untersucht, wobei insgesamt fünf Probanden eine Brille trugen. Die Daten wurden mit Hilfe von Varianzanalysen bei Messwiederholungen und Korrelationsanalysen ausgewertet.

4.1 Bahnplanung

Es zeigten sich für die Bahnplanung signifikante Haupteffekte zwischen den Steuerungsmodalitäten – sowohl für die NASA-TLX Gesamtwerte ($F(2, 36) = 7,08, p < 0,01$), als auch für die Fehlerquote in der Nebenaufgabe ($F(2, 36) = 3,99, p < 0,05$) (Abbildung 2). Die Tablet-Steuerung führte bei beiden Messungen zu signifikant geringeren Beanspruchungswerten als die Sprach- oder Gestensteuerung.

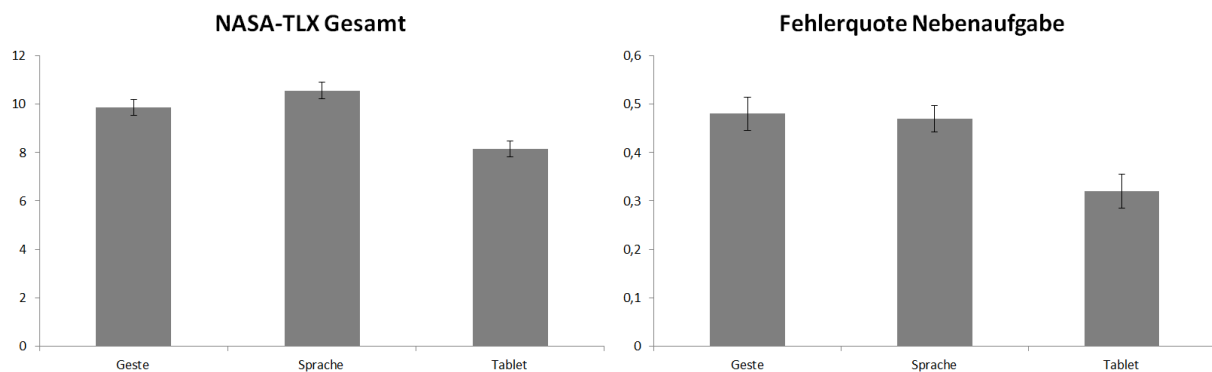


Abbildung 2: NASA-TLX Werte und Fehlerquote in der Nebenaufgabe für die Bahnplanung

Weitere interessante Befunde ergaben sich bei der Auswertung einzelner Items des NASA-TLX: Es lagen keine signifikanten Unterschiede im Item „Zeitliche Anforderung“ zwischen den Eingabemodalitäten vor, sodass davon ausgegangen werden kann, dass dieser Parameter keinen Einfluss auf die Gesamtbewertung hatte. Darüber hinaus war die Bewertung des Items „Frustration“ im Mittel sehr gering ($M = 7,95, SD = 1,70$), sodass anzunehmen ist, dass ein vereinzelt Nicht-Erkennen von Sprach- oder Gestenbefehlen die erlebte Beanspruchung nicht erhöhte.

4.2 Bahnkorrektur

Auch für die Bahnkorrektur zeigten sich signifikante Unterschiede zwischen den Steuerungsmodalitäten in den NASA-TLX Gesamtwerten ($F(2, 34) = 7,47, p < 0,01$) und der Fehlerquote in der Nebenaufgabe ($F(2, 34) = 3,89, p < 0,05$) (Abbildung 3). In diesem Anwendungsszenario wichen die subjektiven Bewertungen allerdings von den Leistungsdaten ab. Die subjektiven Bewertungen zeigten lediglich, dass die Gestensteuerung als signifikant höher beanspruchend erlebt wurde als die beiden anderen Eingabemodalitäten. Die Fehlerquote war hingegen bei der Tablet-Steuerung signifikant geringer als bei den beiden anderen Modalitäten. Im Falle der Tablet-Steuerung war, im Gegensatz zu den beiden anderen Steuerungsmodalitäten, keine Kontrolle des Erkennens einer Eingabe auf einem Feedbackbildschirm notwendig, was sich in höheren Fixationsdauern auf die Nebenaufgabe äußerte. Die Fehlerquote war daher bei der Tablet-Steuerung für die Bahnkorrektur nicht repräsentativ. Laut der NASA-TLX Werte war für dieses Anwendungsszenario also die Sprachsteuerung am besten geeignet.

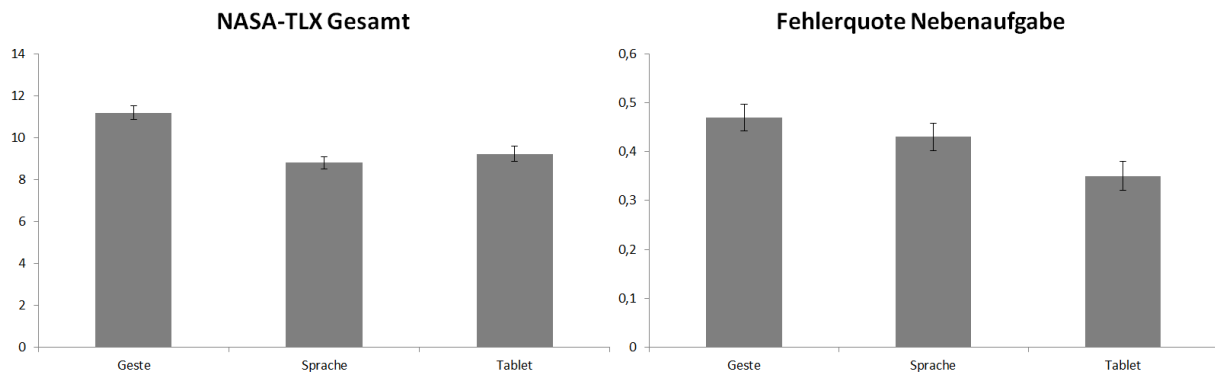


Abbildung 2: NASA-TLX Werte und die Fehlerquote für die Bahnkorrektur

4.3 Auswertung der physiologischen Parameter

Durch den Einsatz einer neuen Methode nach Schneider & Deml (in Druck) zur Gruppierung einer Stichprobe nach beanspruchungsbedingten Adaptionstrategien (bewusst und unbewusst), zeigten sich hoch signifikante Korrelationen zwischen allen physiologischen Parametern und dem NASA-TLX auf Probandenebene (Tabelle 1). Dabei wurden die Probanden anhand der Ausprägung ihrer individuellen Adaptionstrategien in Cluster eingeordnet. Durch bewusste Adaptionen über die Bearbeitungszeit der Aufgabe hinweg verändert sich beispielsweise die Blickstrategie, was sich direkt auf die NNI-Werte, aber auch auf die Fixationsdauern auswirkt. Ein Proband verändert jedoch nicht nur bewusst sein Verhalten, sondern ebenso unbewusst. Unbewusste Adaptionen zeigen sich unter anderem in der zeitlichen Veränderung der Pupillendurchschnittsvariabilität oder der Blinzelrate und Blinzeldauer (Schneider & Deml in Druck). Durch diese individuell verschiedenen Adaptionen kann jedes Cluster mit anderen physiologischen Parametern hoch zuverlässig beschrieben werden. Eine einheitliche Auswertung auf Ebene der gesamten Stichprobe war aufgrund des individuell verschiedenen Verhaltens, wie die Ergebnisse der Studie zeigen, nicht mit ausreichender Zuverlässigkeit möglich.

Tabelle 1: Korrelationen zwischen den physiologischen Parametern und dem NASA-TLX

	Blinzelrate	Blinzeldauer	PERCLOS	Pupillendurchschnittsvariabilität	NNI	Fixationsdauer
NASA-TLX	$r = -0,62^{**}$	$r = -0,58^{**}$	$r = -0,63^{**}$	$r = 0,66^{**}$	$r = 0,70^{**}$	$r = 0,60^{**}$

Korrelationskoeffizienten wurden Fishers' z-transformiert, gemittelt und rücktransformiert

5. Diskussion

Alle drei erhobenen Beanspruchungsindikatoren des trimodalen Ansatzes lieferten ein konsistentes Ergebnis bezüglich der Bewertung der Steuerungsmodalitäten. Für die Bahnplanung zeigte die Tablet-Steuerung die geringsten Beanspruchungswerte ($M_{TLX} = 8,14$, $SD = 3,06$), wohingegen im Falle der Bahnkorrektur die Sprachsteuerung die niedrigste Beanspruchung aufwies ($M_{TLX} = 8,80$, $SD = 2,62$). Die Ergebnisse der Studie zeigen deutlich, dass eine adaptive und multimodale Schnittstellengestaltung in der Mensch-Roboter-Interaktion zu einer Optimierung der Beanspruchungssi-

tuation führen kann. Es sind jedoch weitere Studien notwendig, um die Steuerungsmodalitäten in einem Anwendungsszenario zu kombinieren und zu überprüfen, ob ein dynamisches Wechseln der Modalitäten zu einer Optimierung der Beanspruchungssituation führen kann.

6. Literatur

- Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., & Re, A. (2011). Driver workload and eye blink duration. *Transportation Research Part F* 14, 199-208.
- Brookhuis, K., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis and Prevention*(42), 898-903.
- Cardona, G., & Quevedo, N. (2014). Blinking and Driving: the Influence of Saccades and Cognitive Workload. *Current Eye Research*, 239-244.
- Clark, P. J., & Evans, F. C. (1954). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology*(35), 445.
- Demberg, V. (2013). Pupillometry: The Index of Cognitive Activity in a dual-task study. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Hrsg.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Berlin: Austin TX.
- Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A Random Glance at the Flight Deck: Pilots' Scanning Strategies and the Real-Time Assessment of Mental Workload. Rom: University of Rome.
- Fong, T., Thorpe, C., & Baur, C. (2003). Collaboration, Dialogue, Human-Robot Interaction. *Robotics Research Volume 6*, 255-266.
- Hart, S., & Staveland, L. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload*, 139-183.
- Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working Memory, Attention Control, and the N-Back Task: A Question of Construct Validity. *J of Experimental Psychology* 33, 615-622.
- Loewenfeld, I. (1993). The pupil: Anatomy, physiology and clinical applications (Volume 1). Detroit: Iowa State University Press.
- Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung. In F. Rösler, *Enzyklopädie der Psychologie* (S. 799-864). Göttingen: Hofgrefe.
- Marshall, S. P. (2000). Patentnr. 6,090,051. USA.
- Neudörfer, A. (2014). Konstruieren sicherheitsgerechter Produkte. Methoden und systematische Lösungssammlungen zu EG-Maschinenrichtlinie (Bd. 6). Berlin: Springer Vieweg (VDI-Buch).
- Packebusch, L. (2003). *Psychische Belastung und Beanspruchung - Normung für die Praxis. Wirtschaftspsychologie aktuell* (4), 32-36.
- Recarte, M. Á., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied* 6, 31.
- Recarte, M. Á., Pérez, E., Conchillo, Á., & Nunes, L. M. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *The Spanish Journal of Psychology*, 374-385.
- Schneider, M & Deml, B. (in Druck). An integrated approach of mental workload assessment. In B. Deml, P. Stock, R. Bruder, & Ch. M. Schlick (Hrsg.). *Advances in Ergonomic Design of Systems, Products and Processes*. Berlin: Springer
- Schwalm, M. (2009). Pupillometrie als Methode zur Erfassung mentaler Beanspruchung im automotive Kontext. Saarbrücken: Universität Saarland.
- Spillner, R. (2015). Einsatz und Planung von Roboterassistenz zur Berücksichtigung von Leistungswandlungen in der Produktion. München: Utz.
- Wickens, C.D., Hollands, J., Banburry, S., & Parasuraman, R. (2013). *Engineering Psychology and Human Performance*. New York: Pearson.

Förderung: Die hier dargestellten Forschungsarbeiten wurden im Rahmen des Forschungsprojektes „Inklusion durch Innovation: Adaptive Mensch-Maschine-Schnittstellen für komplexe Automationssysteme (AMIKA)“ durchgeführt. Dieses Projekt wurde mit Mitteln des Bundesministeriums für Bildung und Forschung sowie des Europäischen Sozialfonds unter dem Förderkennzeichen 16SV6281K unterstützt.