

Interaktionsmodalitäten für die Mensch-Roboter-Interaktion – ein systematisches Review

Sarah SOMMER, Patricia H. ROSEN, Sascha WISCHNIEWSKI

*Fachgruppe „Human Factors, Ergonomie“,
Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)
Friedrich-Henkel-Weg 1-25, D-44149 Dortmund*

Kurzfassung: Durch Entwicklungen im Rahmen des Konzeptes Industrie 4.0 nehmen komplexe Arbeitssysteme und Arbeitsformen im Kontext des Einsatzes neuer Technologien zu. Robotische Systeme und Menschen arbeiten zunehmend in einem gemeinsamen Arbeitsbereich. Aus dieser direkten Mensch-Roboter-Interaktion (MRI) ergeben sich veränderte Anforderungen, zu deren Bewältigung multimodale Dialogsysteme einen Beitrag leisten können. Der vorliegende Beitrag liefert auf Basis einer systematischen Literaturrecherche eine Zusammenfassung multimodaler Interaktionsmöglichkeiten, ihren Vorteilen gegenüber einzelnen Möglichkeiten des Signalin- und -outputs und aktuelle Forschungsergebnisse im Bereich der multimodalen Dialoggestaltung innerhalb der direkten MRI. Die Analyse zeigt, dass die Kombination unterschiedlicher Ein- und Ausgabemodalitäten (vor allem sprachliche und gestenbasierte Modalitäten) Vorteile bezüglich der Flexibilität, der Effizienz und der Robustheit gegenüber unimodalen Interaktionen aufweist. Weitere Ansätze wie (vibro-)taktile Signale oder die Blickerkennung fördern die Natürlichkeit der MRI und unterstützen die Akzeptanz und die wahrgenommene Kontrollierbarkeit durch den Anwender.

Schlüsselwörter: Roboter, Interaktion, Multimodalität, Gestaltung

1. Einleitung

Die zunehmende Vielfalt von Produkten und der Bedarf einer immer flexibleren Fertigung setzen wandlungsfähige Fertigungssysteme voraus, die sich rechtzeitig und schnell an ständig ändernde Anforderungen anpassen lassen. Im Rahmen ebendieser komplexen Arbeitssysteme gewinnt der Einsatz industrieller Leichtbau-Roboter an Bedeutung. Bis 2020 werden nach Schätzungen des Weltbranchenverbandes International Federation of Robotics (IFR) insgesamt circa 3,05 Millionen Industrieroboter im Einsatz sein. Von 2016 bis 2017 konnte weltweit ein Anstieg von 30% an ausgelieferten Industrierobotern verzeichnet werden (IFR 2018). Deutschland zählt mit 322 Industrierobotern auf 10.000 Beschäftigte zu den führenden Nationen hinsichtlich der Roboterdichte (IFR 2018). Die zunehmende Nutzung industrieller Leichtbau-Roboter, die eine direkte Zusammenarbeit zwischen Mensch und Roboter ermöglichen, spiegelt sich in einem zunehmenden Forschungsinteresse an der (direkten) Mensch-Roboter-Interaktion (MRI) wider. Aus dieser direkten MRI ergeben sich neue Herausforderungen und veränderte Anforderungen, die einerseits die Kommunikation und andererseits weitere Faktoren wie die individuellen Voraussetzungen der Anwender und Umgebungsbedingungen umschließen. Zur Bewältigung

dieser Herausforderungen werden multimodale Dialogsysteme benötigt, die sowohl vorhandene Ein- und Ausgabekanäle (z. B. Schalter etc.) als auch neuere Möglichkeiten, wie z. B. Blickerkennung, bedienen können. Um dies zu erreichen, bieten multimodale Geräte verschiedene Wahrnehmungs- und Aktionsmodalitäten an, sodass der Anwender über unterschiedliche Kanäle bzw. Sinne Informationen erhält.

Die Integration verschiedener (sequenzieller oder paralleler) Ein- und Ausgaben ist neben der Auswahl geeigneter Ansätze die Hauptaufgabe der Gestaltung multimodaler MRI. Hierbei gilt es, die verschiedenen Eigenschaften der jeweiligen Modalitäten effizient und effektiv für die vorliegende Arbeitsaufgabe und in Bezug auf individuelle Bedürfnisse der Anwender und die spezifische Arbeitsumgebung zu kombinieren. Eine Anlehnung an Forschungserkenntnissen der Mensch-Computer-Interaktion sei nach Kiesler und Hinds (2004) zum Teil möglich, allerdings sei auf die Besonderheiten der robotischen Systeme gegenüber reinen Computern hinzuweisen: Zum einen sind die mentalen Modelle, die der Mensch über den Roboter hat, stärker anthropomorph und zum anderen haben Roboter gewisse Freiheitsgrade hinsichtlich ihrer Autonomie und müssen ihre Interaktion an die gegebenen Bedingungen schnell und flexibel anpassen können.

Ein Großteil der Forschungsansätze konzentrierte sich bislang auf den Anwendungsfall der Roboterprogrammierung. Ziel dieses Reviews war daher, anhand einer systematischen Literaturrecherche vorhandene multimodale Interaktionsmöglichkeiten zu analysieren, Einsatzmöglichkeiten verschiedener Modalitäten in Abhängigkeit von vorhandenen Faktoren und ihren Effekten auf unterschiedliche Outcome-Variablen (z. B. Leistung, Akzeptanz etc.) zu identifizieren und daraus Gestaltungsempfehlungen für die Mensch-Roboter-Zusammenarbeit abzuleiten.

2. Methode

Zur Erhebung der aktuellen Studienlage (multi-) modaler Interaktionen in der Mensch-Roboter-Zusammenarbeit wurde eine systematische Literaturrecherche durchgeführt. Die Literaturrecherche erfolgte über die drei Datenbanken EBSCOhost, SCOPUS und ScienceDirect sowie den dort inkludierten Datenbanken.

Der erste Schritt des später finalen Suchstrings umfasste die Begrifflichkeiten der direkten Zusammenarbeit zwischen Mensch und Roboter selbst (Mensch-Roboter-Interaktion/Modalität/Kooperation/Kollaboration). In einem weiteren Schritt der systematischen Eingrenzung der Suchergebnisse auf den Bereich der direkten MRI wurden Studien eingeschlossen, die a) einen direkten Bezug zur Industrie, Produktion, Fertigung oder einem ähnlichen Bereich aufwiesen, b) eine oder mehrere Ein- oder Ausgabemöglichkeiten direkt untersuchten bzw. miteinander verglichen und c) in deutscher oder englischer Sprache zur Verfügung standen.

Ausgeschlossen wurden Studien, die a) in einem medizinischen oder gesundheitsbezogenen Kontext stattfanden, b) den Bereich der Service-Robotik untersuchten, c) lediglich den Anwendungsfall der Roboterprogrammierung beinhalten und d) die keine humane, erwachsene Zielpopulation umfassten.

Mit den finalen Suchstrings (z. B. für die Datenbank EBSCOhost *“human robot interaction OR human robot collaboration OR human robot cooperation OR robot modality OR human robot modality NOT (medicine OR medical OR healthcare OR children) AND (modality OR surface OR industrial OR production OR automotive)”*) wurden die Datenbanken EBSCOHOST, SCOPUS und ScienceDirect analysiert. Die

so erhaltenen Treffer wurden auf Basis von Querverweisen und Literaturangaben per Handsuche abschließend ergänzt.

3. Ergebnisse

Das ungefilterte Ergebnis der Literatursuche zum Thema Mensch-Roboter-Interaktion / Kooperation / Kollaboration ergab aggregiert über alle Datenbanken $n = 218.252$ Suchtreffer. 3.672 Literaturstellen konnten unter Berücksichtigung der Ein- und Ausschlusskriterien herausgefiltert werden. 1.031 Suchtreffer wurden nach Ausschluss von Duplikaten und falschen Treffern auf Grund doppelter Wortbedeutungen weiter basierend auf ihrem Titel beurteilt. Auf dieser Ebene konnten 269 Studien in den weiteren Auswahlprozess eingeschlossen und auf Ebene ihrer Abstracts analysiert werden, sodass letztlich 103 Studien für ein Volltextscreening herausgefiltert wurden. Auf Volltextebene wurden Artikel ausgeschlossen, wenn eine andere Studiensprache als Deutsch oder Englisch vorlag oder die Interaktionsmodalität nicht im Vordergrund der Erhebung stand. Insgesamt gingen 89 Artikel in die weitere Betrachtung ein. Eine Übersicht über die einzelnen Reviewphasen zeigt Abbildung 1.

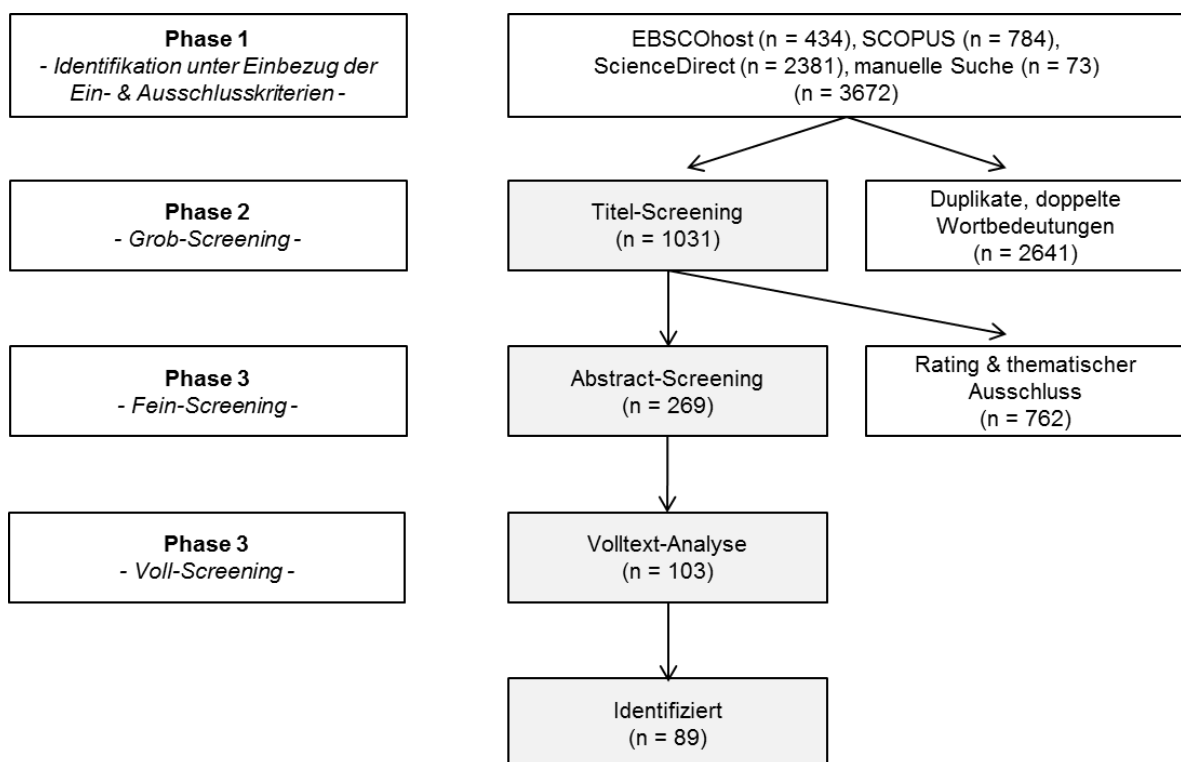


Abbildung 1: Prozess der systematischen Literaturrecherche

Bei den berücksichtigten Volltexten handelte es sich sowohl um Untersuchungen im Feld (6) als auch um Laborstudien mit der Fokussierung und Analyse einer (43) oder mehrerer Interaktionsmodalitäten (21). Weitere Artikel beschäftigten sich mit den allgemeinen Vorteilen multimodaler Interaktionsmöglichkeiten (7) und der Entwicklung ebendieser in den vergangenen Jahren auf der Basis theoretischer Überlegungen (12). Anhand der identifizierten Studien kann eine Aufteilung bzw. Kategorisierung nach Modalität (unimodal vs. multimodal) und nach betrachteten Outcomevariablen erfolgen. Die Outcomevariablen sind sehr heterogen und beziehen sich so-

wohl auf den Anwender direkt (Akzeptanz, Beanspruchung etc.), auf die Arbeitsleistung (Fehlerrate, Zeit etc.) oder auf allgemeine Sicherheitsaspekte, sodass letztlich eine Kategorisierung nach personen-, leistungs- oder sicherheitsbezogenen Faktoren möglich ist. Im Nachfolgenden werden in Anlehnung an die vorgeschlagene Kategorisierung zentrale Forschungsergebnisse und die allgemeinen Tendenzen der weiteren Forschungsarbeiten dargestellt.

Ruiz et al. (2010) geben einen Überblick über die Vorteile multimodaler Eingabeschnittstellen gegenüber unimodalen Möglichkeiten: Flexibilität, Fehlerreduktion und Ähnlichkeit zur Mensch-Mensch-Interaktion. Ohne fehlende Einarbeitung der Anwender tendieren diese jedoch zu weniger effizienten Ein- und Ausgabemodalitäten (Dániel et al. 2014). Die Bereitstellung multimodaler Interaktionsmöglichkeiten hilft dabei Ambiguitäten aufzuklären, die durch unimodale Gestaltungen auftreten können, sodass die Robustheit der Kommunikation erhöht wird (Salem et al. 2011).

Einerseits beeinflusst die Art und die Komplexität der Arbeitsaufgabe die Wahl der Modalität. Andererseits steigt die Wahrscheinlichkeit für Multimodalität mit zunehmender kognitiver Beanspruchung während der Arbeitsaufgabe, was wiederum direkt mit der Komplexität der Arbeitsaufgabe zusammenhängt (Oviatt et al. 2004). Weiterhin zeichnet sich eine Adaption der gewählten Modalität in Abhängigkeit der eingesetzten Modalität durch den Roboter ab. Agiert der Roboter mittels Sprachausgabe, nutzen Anwender häufiger die Spracheingabe trotz des Vorhandenseins weiterer Eingabemodalitäten (Bellik et al. 2009).

23 der gefilterten Untersuchungen bzw. Forschungsarbeiten konzentrierten sich auf die Anwendung sprachlicher Ein- und Ausgabemöglichkeiten (unimodal). Hierbei konnte gezeigt werden, dass die direkte verbale Interaktion mit dem Roboter von einem Großteil der befragten bzw. untersuchten Personen bevorzugt wird (unabhängig davon, ob dies die effizienteste Variante der direkten Kommunikation mit dem Roboter in der vorliegenden Arbeitssituation ist). Des Weiteren konnten verschiedene Studien zeigen, dass Faktoren wie Lärm, Geschlechtsunterschiede hinsichtlich der Stimmfarbe oder Stress die Nutzungsmöglichkeiten sprachlicher Signale verringern (Erickson et al. 2012; Jung et al. 2013).

14 der 23 oben erwähnten Studien und sieben weitere Studien beschäftigten sich mit dem Ansatz der gestenbasierten Interaktion bzw. mit der Kombination ebendieser mit sprachlichen Ein- und Ausgaben. Diese Studien deuten darauf hin, dass die Kombination von sprach- und gestengesteuerten Ein- und Ausgaben effizienter gegenüber unimodalen Kommunikationsmöglichkeiten ist und die gestengesteuerte Interaktion die natürliche Interaktion zwischen Mensch und Roboter fördert (z. B. Burger et al. 2012; Gleeson et al. 2013). Rossi et al. (2013) konnten eine höhere Genauigkeit in der MRI bei der kombinierten Nutzung von Gesten und Sprache gegenüber der separierten Nutzung der Interaktionsmöglichkeiten messen. Die Erkennung von Gesten kann zudem bei der Vermittlung von Informationen helfen, die Sprache alleine nicht erzeugen könnte wie beispielsweise bei referentiellen, räumlichen oder symbolischen Informationen (Salem et al. 2011). Ein zusätzlicher Vorteil hinsichtlich der Nutzung von Gesten durch den Roboter kann in einer erhöhten Kompetenzzuschreibung durch den Anwender und somit zu einer höheren Akzeptanz und einer positiveren MRI führen (Salem et al. 2011).

Wie bereits erkennbar, basiert neben den allgemeinen theoretischen Artikeln zur multimodalen Interaktion ein Großteil der identifizierten Forschungsarbeiten auf sprachlichen oder gestenbasierten Interaktionsmöglichkeiten. Neuere Forschungsansätze (und aufgrund dessen eine eher kleine Anzahl an Studien) konzentrieren

sich auf komplexere Ansätze wie beispielweise die Blickerkennung, taktile Ein- und Ausgabeschnittstellen sowie die Intentionserkennung.

Die Blickerkennung als Ergänzung zur sprachlichen oder taktilen Eingabe kann nach Bannat et al. (2009) vor allem in Situationen, in denen eine manuelle Eingabe aufgrund der Arbeitsaufgabe oder eine sprachliche Eingabe aufgrund von Lärm nicht möglich wären, eingesetzt werden.

Der Ansatz einer taktilen Erfassung und / oder Weitergabe von Informationen rückt zunehmend in den Fokus der Forschung zur MRI. Es konnte bereits gezeigt werden, dass der Einsatz visueller Informationen in Kombination mit der Weitergabe von Informationen durch taktile Sensoren zur Erfassung multipler Objekteigenschaften zu einer besseren Rekognition führen und dadurch keine zeitlichen bzw. die Ausführungsdauer betreffenden Nachteile entstehen (Liu et al. 2017). Des Weiteren können (vibro-)taktile Rückmeldungen genutzt werden, um die durch den Anwender subjektiv wahrgenommene Kontrollierbarkeit und Sicherheit der Roboter zu erhöhen (Sieber et al. 2015).

4. Diskussion

Die Industrie befindet sich in einem kontinuierlichen Wandel. Nicht nur der Trend hin zu einer schnelleren und flexibleren Produktion ist hier ein Resultat. Vielmehr werden immer häufiger wandlungsfähige Fertigungs- und Montagesysteme benötigt, deren Zentrum die direkte Zusammenarbeit von Mensch und Roboter bildet.

Ebendiese Mensch-Roboter-Zusammenarbeit bedarf einer multimodalen Interaktionsmöglichkeit. Multifaktorielle Bedingungen sind entscheidend für den Einsatz multimodaler Schnittstellen in der direkten MRI. Sowohl die Arbeitsumgebung (charakterisiert durch Lärm-, Licht- und Raumverhältnisse), als auch zeitliche Vorgaben, die Eigenschaften der Arbeitsaufgabe, das Geschlecht der Anwender oder der Grad des subjektiv erlebten Stresses können die richtige Wahl und die effiziente Kombination der ausgewählten Interaktionsmodalitäten beeinflussen.

Aus verschiedenen Untersuchungen lässt sich ein Vorteil in der Kombination von gesten- und sprachgesteuerten Ein- und Ausgabemöglichkeiten hinsichtlich Effizienz, Robustheit, Fehlerreduktion und Flexibilität ableiten. Des Weiteren können neuere Ansätze wie z. B. die Steuerung mittels Blickerkennung oder die Rückmeldung mit Hilfe taktiler Signale ein zusätzlicher Nutzen für die MRI sein.

In Anlehnung an die Mensch-Mensch-Interaktion und dem Wunsch der Übertragung ebendieser Prinzipien auf die direkte MRI für eine möglichst natürliche Dialoggestaltung ist ein multimodaler Ansatz grundsätzlich sinnvoll. Die Fokussierung auf die Kombination von gesten- und sprachgesteuerten Modalitäten ist ein erster Ansatz um dieses Ziel zu erreichen. Um weitere Herausforderungen wie z. B. Sicherheitsaspekte, die hinsichtlich einer direkten Zusammenarbeit zwischen Mensch und Roboter in das Zentrum der Betrachtung rücken, oder große Informationsmengen bewältigen zu können, ist die Nutzung weiterer Sinne (z. B. taktile Rückmeldungen) von besonderer Bedeutung.

Bislang ist die Anzahl der Studien, die genau diese kombinierten Ansätze untersuchen, sehr gering. Die Mehrheit der Forschungsansätze bezieht sich auf programmierbezogene Fragestellungen und den softwarebasierten Lösungsansätzen. Nur wenige Studien konzentrieren sich auf die Wirkung der multimodalen Dialoge in der direkten MRI im industriellen Setting unter Berücksichtigung der gegebenen Bedingungen und der Stichprobe.

Um fundierte Aussagen zu den Effekten von multimodalen Interaktionsmöglichkeiten in der direkten MRI auf das Wohlbefinden der Anwender, aber auch auf die Effizienz der Arbeitsleitung ableiten zu können, sind weitere Forschungsarbeiten angezeigt.

5. Literatur

- Bannat A, Gast J, Rehrl T, Rösel W, Rigoll G & Wallhoff F (2009) A Multimodal Human-Robot-Interaction Scenario: Working Together with an Industrial Robot. *Human-Computer Interaction 2*: 303-311.
- Bellik Y, Rebaï I & Machrouh E (2009) Multimodal Interaction within Ambient Environments: An Exploratory Study. In: T. Gross et al. (eds). *Human-Computer Interaction – INTERACT 2009*. INTERACT 2009. *Lecture Notes in Computer Science*: 5727. Berliner: Springer
- Burger B, Ferrané I, & Lerasle F (2010) Towards multimodal interface for interactive robots: challenges and robotic systems description. In: Abdellatif H (ed.) *Robotics 2010 - Current and future challenges*: 369-380.
- Erickson D, DeWees M, Lewis J & Matson ET (2013) Communication for Task Completion with Heterogeneous Robots. In: Kim JH et al. (eds) *Robot Intelligence Technology and Applications 2012*. *Advances in Intelligent Systems and Computing*: 873-882. Berliner: Springer
- Gleeson B, MacLean K, Haddadi A, Croft E & Alcazar J (2013). Gestures for industry Intuitive human-robot communication from human observation. *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*: 349–356.
- Jung MF, Lee JJ, DePalma N, Adalgeirsson SO, Hinds PJ & Breazeal C (2013) Engaging robots: easing complex human-robot teamwork using backchanneling. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*: 1555-1566.
- International Federation of Robotics (2018) *World Robotics - Industrial Robots*. VDMA
- Kiesler S & Hinds P (2004) Introduction to this Special Issue on Human-Robot Interaction. *Human Computer Interaction 19*: 1-8.
- Liu H, Yu Y, Sun F & Gu J (2017) Visual–Tactile Fusion for Object Recognition. *Proceedings of the IEEE Transactions on Automation Science and Engineering 14*: 996-1008.
- Oviatt S, Coulston R & Lunsford R (2004) When do we interact multimodally?: cognitive load and multimodal communication patterns. *Proceedings of International Conference on Multimodal Interfaces*: 129–136.
- Rossi S, Leone E, Fione M, Finzi A & Cutugno F (2013) An extensible architecture for robust multimodal human-robot communication. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*: 2208-2213.
- Ruiz N, Chen F & Oviatt S (2010) Multimodal Input. In Thiran JP et al. (eds.) *Multimodal signal processing*: 231–255. Elsevier.
- Salem M, Rohlfing K, Kopp S & Joubin F (2011) A Friendly Gesture: Investigating the Effect of Multimodal Robot Behavior in Human-Robot Interaction. *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*: 247-252.
- Sieber D, Selma M & Hirche S (2015) Multi-robot manipulation controlled by a human with haptic feedback. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Hinweis: Teile dieses Berichtes entstanden im Rahmen des Projektes Hybr-iT. Das Forschungs- und Entwicklungsprojekt Hybr-iT – Hybride Teams in wandlungsfähigen, cyber-physischen Produktionsumgebungen wird mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) gefördert und vom Projektträger im Deutschen Zentrum für Luft- und Raumfahrt e.V. Softwaresysteme und Wissenstechnologien (PT-SW) betreut (01IS16026H).



Gesellschaft für
Arbeitswissenschaft e.V.

Arbeit interdisziplinär analysieren – bewerten – gestalten

65. Kongress der
Gesellschaft für Arbeitswissenschaft

Professur Arbeitswissenschaft
Institut für Technische Logistik und Arbeitssysteme
Technische Universität Dresden

Institut für Arbeit und Gesundheit
Deutsche Gesetzliche Unfallversicherung

27. Februar – 1. März 2019

GfA-Press

Bericht zum 65. Arbeitswissenschaftlichen Kongress vom 27. Februar – 1. März 2019

**Professur Arbeitswissenschaft, Institut für Technische Logistik und Arbeitssysteme,
Technische Universität Dresden;
Institut für Arbeit und Gesundheit, Deutsche Gesetzliche Unfallversicherung, Dresden**

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.
Dortmund: GfA-Press, 2019
ISBN 978-3-936804-25-6

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle erhältlich.

Alle Rechte vorbehalten.

© **GfA-Press, Dortmund**

Schriftleitung: Matthias Jäger

im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

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

- den Konferenzband oder Teile daraus in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) zu vervielfältigen,
- den Konferenzband 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.

Screen design und Umsetzung

© 2019 fröse multimedia, Frank Fröse

office@internetkundenservice.de · www.internetkundenservice.de