

X-Ray-Screening von registriertem Gepäck: 2D- vs. 3D-Technologie und Evaluation eines On-Screen Alarm Resolution Protokolls

Nicole HÄTTENSCHWILER, Sarah MERKS, Myrta ISENSCHMID,
Adrian SCHWANINGER

*Institut Mensch in komplexen Systemen (MikS), Hochschule für
Angewandte Psychologie, Fachhochschule Nordwestschweiz (FHNW)
Riggenbachstrasse 16, 4600, Olten, Schweiz*

Kurzfassung: 2D Röntgensysteme mit multi-view Technologie werden an Flughäfen zur Erkennung von Sprengstoff bei registriertem Gepäck verwendet. Neuere Röntgensysteme basieren auf Computer Tomographie (CT), wobei Gepäckstücke als 3D-rotierbare Bilder dargestellt werden. Es steht zur Diskussion, ob Screener mit 2D Erfahrung spezifisches Training benötigen, bevor sie mit 3D arbeiten können. Zur Unterstützung der Alarmauflösung bei 3D gibt es zusätzlich On-screen Alarm Resolution Protokolle (OSARP), dessen Effektivität bisher noch nicht untersucht wurde. Um diesen Fragen nachzugehen, wurden in einem Experiment drei Bedingungen verglichen. In der ersten Bedingung bearbeiteten Screener eine Inspektionsaufgabe mit 2D multi-view Technologie (2D), in der zweiten Bedingung mit 3D CT Technologie (3D) und in der dritten Bedingung mit 3D und einem vorher durchgeführten Training zu OSARP. Die Ergebnisse zeigen, dass es keine Unterschiede bezüglich der Erkennungsleistung (d') zwischen 2D und 3D gibt. Unter der Befolgung eines OSARP erreichten die Screener eine höhere Erkennungsleistung als bei 2D und 3D. Die Reaktionszeiten waren bei den 3D Bedingungen länger als bei 2D. Die Resultate zeigen, dass Screener, welche momentan mit 2D multi-view Technologie arbeiten, kein spezifisches Training brauchen, um eine ähnliche Erkennungsleistung mit 3D Technologie zu erreichen. Die Nutzung eines OSARP könnte die Erkennungsleistung zusätzlich noch verbessern.

Schlüsselwörter: Mensch-Maschine Interaktion, visuelle Inspektion, Flughafensicherheit, Röntgentechnologie, 2D vs 3D Bildgebung, On-Screen Alarm Resolution

1. Einleitung

Sicherheitsbeauftragte an Flughäfen (Screener) analysieren Gepäckstücke mit Röntengeräten, um sicherzustellen, dass keine verbotenen Gegenstände an Bord eines Flugzeuges gelangen. Während bei der Inspektion von Handgepäckstücken mehrere verbotene Gegenstände von Relevanz sind (Bomben, Messer, Schusswaffen, u.a.), liegt der Fokus bei der Inspektion von registrierten Gepäckstücken (hold baggage screening: HBS) auf unkonventionellen Spreng- und Brandvorrichtungen (USBV; Bretz 2002). Röntgenmaschinen für HBS-Gepäckstücke sind deshalb mit automatischer Sprengstofferkennung (EDS) ausgestattet. Diese Systeme markieren sprengstoffverdächtige Gegenstände auf Röntgenbildern mit einem roten Rahmen

oder der Einfärbung des Gegenstandes (Wells & Bradley 2012). Die Aufgabe der Screener ist darauffolgend die visuelle Inspektion derjenigen Röntgenbilder, welche mit einem Alarm der Maschine markiert wurden, um zu entscheiden, ob es sich bei dem Alarm um eine USBV oder einen Fehlalarm des EDS handelt. Entsprechend kann die visuelle Inspektion von HBS-Gepäckstücken als Entscheidungsaufgabe eingestuft werden gemäss Signaldetektionstheorie (Koller et al. 2009). Um eine USBV in einem Röntgenbild zu erkennen, müssen Screener mit den folgenden Komponenten vertraut sein: Auslösemechanismus, Energiequelle, Drähte und explosives Material (Turner 1994; Wells & Bradley 2012). Diese erlernen Screener durch computer-basiertes Training, welches sich durch mehrere Studien als effektive und effiziente Methode erwiesen hat, um eine hohe Erkennungsleistung von USBV zu erzielen (Halbherr et al. 2013; Koller et al. 2008; Schuster et al. 2013; Schwaninger et al. 2007).

Maschinen mit 2D multi-view Röntgentechnologie sind momentan noch die am häufigsten verwendeten Systeme um HBS-Gepäckstücke zu röntgen. Diese Technologie ist jedoch nicht vollumfänglich in der Lage, jeden Winkel eines Gepäckstücks abzubilden, weshalb dicht bepackte Gepäckstücke und verdeckte Gegenstände schwer zu analysieren sind (Megherbi et al. 2010). Neuere Röntgengeräte basierend auf 3D CT Technologie. Sie generieren mehrere Schnitte eines Gepäckstücks, welche dann zu einem volumetrischen Bild zusammengerechnet werden (Singh & Singh 2003). Screener erhalten somit bei der Inspektion zusätzliche Tiefeninformationen und die Möglichkeit, das Bild eines Gepäckstücks um eine 360° Achse zu drehen, sowie die einzelnen Schnitte zu analysieren (Mouton & Breckon 2015; Sighn & Sighn 2003; Wells & Bradley 2012). Durch diese zusätzlichen Funktionalitäten könnte die visuelle Inspektion von 3D-rotierbaren Bildern die Inspektionsleistung von Screenern verbessern, insbesondere bei verdeckten Gegenständen oder dicht verpackten Gepäckstücken. Durch die regelmässige Interaktion mit 3D Objekten kann zudem angenommen werden, dass Screener detaillierte Objektrepräsentationen haben (Vuong & Tarr 2004), was die Erkennung von USBV in 2D und 3D Röntgenbildern erleichtert (Hättenschwiler, Mendes, & Schwaninger 2018). Gleichzeitig ist jedoch die Bildqualität von 3D CT Maschinen schlechter als bei 2D Röntgensystemen. Der Vergleich der Erkennungsleistung bei 2D mit neuerer 3D CT Technologie kann somit Aufschluss darüber geben, ob die Vorteile von 3D CT die negativen Effekte der schlechteren Bildqualität kompensieren können.

Der Vergleich von 2D und 3D Technologie weist weiter eine hohe praktische Relevanz auf, da sowohl Behörden, Flughäfen als auch Sicherheitsdienstleister momentan darüber diskutieren, ob Screener mit 2D Erfahrung spezifisches Training benötigen, bevor sie mit 3D arbeiten können. Eine erste Studie konnte zeigen, dass 2D Screener die gleiche Erkennungsleistung mit 2D und 3D Technologie erzielen ohne spezifisches Training (Hättenschwiler, Mendes, & Schwaninger 2018). Das erste Ziel unserer Studie ist die Replikation dieses Resultats. Desweiteren gibt es für 3D Technologie spezifische On-screen alarm resolution Protokolle (OSARP), welche Screener in ihrer Entscheidungsaufgabe bei der Alarmauflösung unterstützen sollen. Es wird davon ausgegangen, dass unter der Anwendung von OSARP die Erkennungsleistung verbessert werden kann, jedoch gibt es noch keine wissenschaftlichen Untersuchungen dazu. Das zweite Ziel dieser Studie war der Vergleich der Erkennungsleistung von Screenern bei 3D Bildern mit versus ohne die Befolgung eines spezifischen OSARP. Um diese Ziele zu verfolgen, wurde die Erkennungsleistung von 2D HBS Screenern in drei Experimentalbedingungen untersucht: Inspektion von 2D

multi-view Bildern versus 3D CT Bilder versus die Befolgung eines OSARP bei der Inspektion von 3D CT Bildern.

2. Methoden

2.1 Teilnehmende

62 HBS Screener eines internationalen Flughafens nahmen an der Studie teil. Diese verfügten über Erfahrung im HBS mit 2D multi-view Technologie (Tabelle 1).

Tabelle 1: Deskriptive Angaben der Screener aufgeteilt nach Experimentalbedingung.

Experimentalbedingung	n	Alter in Jahren	2D Arbeitserfahrung	% weiblich
2D	20	45.7 (12.8)	9.8 (6.6)	60
3D	22	42.5 (9.4)	9.4 (9.0)	52
OSARP	20	45.6 (11.6)	9.8 (5.8)	45

2.2 Design

Das Experiment wurde als Between-Subjects-Design umgesetzt. Alle Screener absolvierten zuerst einen Pretest, um sich mit den 2D und 3D Testsimulatoren und der Prozedur vertraut zu machen. Basierend auf diesen Testergebnissen, sowie den deskriptiven Variablen Alter und Arbeitserfahrung, wurden die Screener für den Haupttest in drei Bedingungen eingeteilt: Analyse von 2D multi-view Bildern (Abb. 1a/b), Analyse von 3D CT Bildern (Abb. 1c/d), oder Analyse von 3D CT Bildern (Abb. 1c/d) nach einem OSARP Training. Diese drei Experimentalbedingungen galten als unabhängige Variable und Inspektionsleistungsmasse (Sensitivität d' , Antwortendenz c , Reaktionszeiten RZ) wurden als abhängige Variablen berechnet.

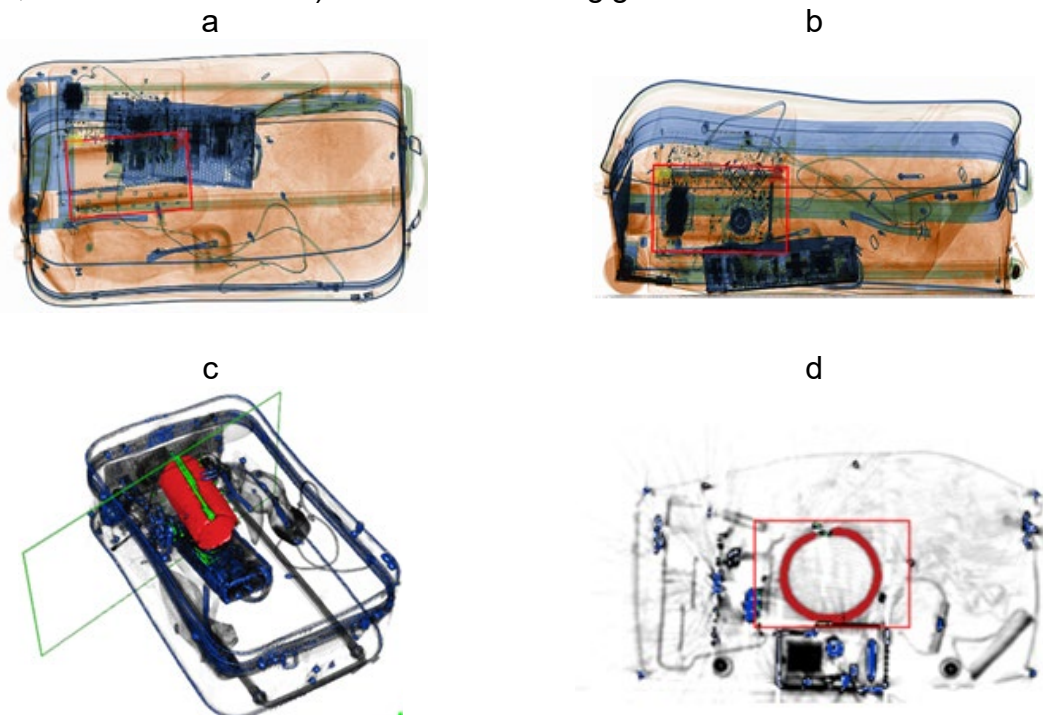


Abbildung 1: Abbildung eines HBS-Gepäckstücks aufgenommen mit einer 2D multi-view Maschine (a und b) und einer 3D CT Maschine (c und d). In der 2D multi-view Ansicht werden EDS Alarme durch rote Rahmen gekennzeichnet, wobei im 3D-rotierbaren und der Schnittansicht EDS Alarme rot eingefärbt werden.

2.3 Material

32 HBS Gepäckstücke wurden durch wiederholtes Umpacken zur Erstellung und Aufnahme von einzigartigen Gepäckstimuli verwendet. Als verbotene Gegenstände wurde in die Hälfte aller Gepäckstücke eine von 64 kreierten USBV eingebaut und in die andere Hälfte der Koffer ein Fehlalarmobjekt (z.B. Käse, Schokolade, Flüssigkeiten). Somit enthielten alle aufgenommenen Gepäckstücke einen Alarm.

Der Haupttest beinhaltete 256 Bilder von Gepäckstücken, wobei alle einmal mit einem 2D- (Abb. 1a/b) und einmal mit einem 3D CT Röntgengerät (Abb. 1c/d) aufgenommen wurden. Zur Sicherstellung der gleichen Systemreliabilität aller Testbedingungen (z.B. Rice & McCarley 2011) wurden die EDS Alarme der 3D CT Maschine als Referenz für die manuelle Setzung von roten Rahmen in 2D Röntgenbilder genommen. Die Prävalenz von USBVs lag bei 50%.

2.4 Vorgehen

Die Erhebung fand an zwei Zeitpunkten statt. Bei der ersten Testung führten die Screener den Pretest durch, bei der zweiten den Haupttest. Beide Tests bestanden aus einer simulierten HBS Screeningaufgabe, bei welcher die Screener instruiert wurden, bei jedem von den 256 Röntgenbildern von HBS-Gepäckstücken zu entscheiden, ob dieses eine USBV enthält (*NOT OK*) oder ein Fehlalarm der Maschine (*OK*) war. Nach der Bearbeitung von 10 Beispielbildern wurde der Haupttest gestartet, wobei ein Zeitlimit von 60s pro Bild gesetzt war.

Screener, welche in die Experimentalbedingung 3D mit OSARP Training eingeteilt waren, absolvierten vor dem Haupttest ein Training zur Erlernung des Protokolls. Das Originalprotokoll, welches im Realbetrieb in zwei Tagen trainiert wird, wurde zu Studienzwecken auf eine Dauer von 40 min gekürzt. Das OSARP enthält heuristische Schritte zur Diskriminierung zwischen einer USBV und einem Fehlalarm der 3D CT Maschine und ist für deren User-Interface entwickelt worden.

Der Haupttest wurde in zwei Blöcke unterteilt, dazwischen fand eine Pause von 10-15 min statt. Alle Screener absolvierten den Pretest in weniger als 40 min und den Haupttest in weniger als 1.5 h.

2.5 Analysen

Es wurden an Hand von SPSS Version 22 Varianzanalysen (ANOVA) berechnet mit den Experimentalbedingungen (2D, 3D, 3D mit OSARP) als unabhängige Variable und den Inspektionsmassen (d' , c , RZ) als abhängige Variablen ($\alpha = 0.05$). Die Inspektionsmasse d' und c wurden anhand der Formeln der Signaldetektionstheorie berechnet (z.B. Green & Swets 1966, MacMillan & Creelman 2005). Post-Hoc Tests wurden mit R (Version 3.22) berechnet, wobei Holm-Bonferroni Korrekturen angewendet wurden (Holm 1979).

3. Resultate und Diskussion

Abbildung 2a/b zeigt die Erkennungsleistung (d') sowie die Reaktionszeiten (RZ) abhängig von den drei Experimentalbedingungen 2D, 3D und OSARP.

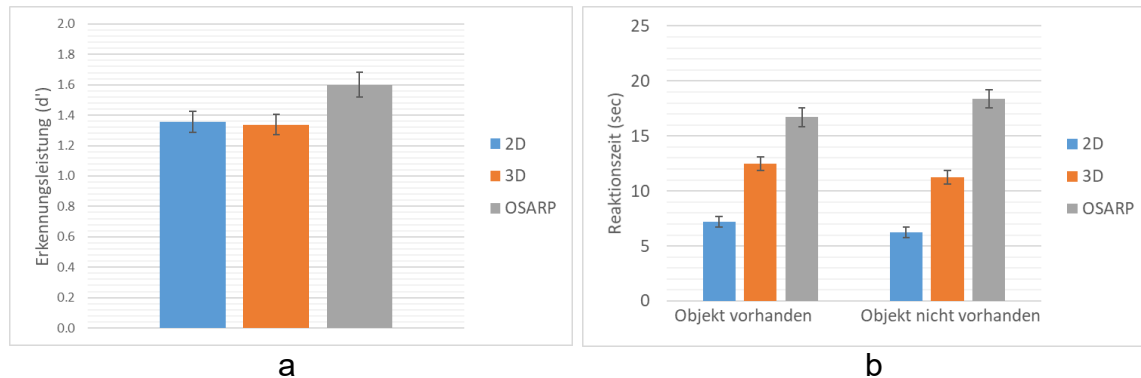


Abbildung 2: a) Detektionsleistung (d') und Standardfehler pro Bedingung. b) Reaktionszeiten (RZ) mit Standardfehler pro Bedingung, aufgeteilt in Bilder mit USBV vorhanden versus USBV nicht vorhanden (Bild mit Fehlalarm).

Eine univariate Varianzanalyse ergab einen signifikanten Haupteffekt für Bedingung, $F(2, 62) = 3.28$, $p = 0.045$, $\eta_p^2 = 0.100$. Der Post-Hoc-Test zwischen 2D und 3D wurde nicht signifikant, $t(40) = 0.145$, $p = 0.886$, $d = 0.05$, jedoch war die Erkennungsleistung mit OSARP besser als bei 3D, $t(40) = 3.05$, $p = 0.002$, $d = 1.05$. Die Analyse der Antworttendenz (c) ergab einen signifikanten Haupteffekt, $F(2, 62) = 15.69$, $p < 0.001$, $\eta_p^2 = 0.35$, wobei Post-Hoc-Tests zwischen den Bedingungen aufzeigten, dass OSARP zu einer tieferen und somit neutraleren Antworttendenz führte im Vergleich zu 2D ($p < 0.001$) und 3D ($p < 0.001$).

Somit konnten die Ergebnisse von Hättenschwiler et al. (2018) repliziert werden. 2D Screener ohne spezifisches Training erzielten die gleiche visuelle Inspektionsleistung mit 3D Technologie wie mit 2D Technologie. Die zusätzlichen Funktionalitäten von 3D CT scheinen somit die schlechtere Bildqualität zu kompensieren. Weiter wurde ersichtlich, dass die Erkennungsleistung mit 3D Technologie durch die Befolgung eines spezifischen OSARP bereits nach einem 40-minütigen Training verbessert werden konnte. Gleichzeitig veränderten Screener in der OSARP Bedingung ihre Antworttendenz indem sie mehr Gepäckstücke als gefährlich einstufen, was zu einer höheren Treffer- als auch Fehlalarmrate führte. Es ist aber davon auszugehen, dass Ergebnisse zur Antworttendenz sich im Realbetrieb auf Grund einer tieferen Objektprävalenz unterscheiden (Green & Swets 1966; MacMillan & Creelman 2005).

In einem nächsten Schritt wurden die Reaktionszeiten (RZ) analysiert (Abb. 2b), wobei die Bilder aufgeteilt wurden nach USBV vorhanden versus keine USBV vorhanden (Bild mit Fehlalarm). Eine mixed ANOVA ergab einen signifikanten Haupteffekt für Bedingung $F(2, 59) = 48.32$, $p < 0.001$, $\eta_p^2 = 0.62$, sowie eine signifikante Interaktion zwischen Bedingung und UBSV vorhanden vs nicht vorhanden, $F(2, 59) = 9.50$, $p < 0.001$, $\eta_p^2 = 0.24$. Die Post-Hoc-Tests ergaben signifikante Unterschiede zwischen allen Bedingungen (2D, 3D, OSARP), sowohl für USBV vorhanden ($p < 0.001$), als auch USBV nicht vorhanden ($p < 0.001$). Die Screener brauchten mehr Zeit um Bilder mit der 3D Technologie zu analysieren. Dies wurde so erwartet, weil die Verwendung der zusätzlichen Funktionalitäten von 3D CT mehr Zeit in Anspruch nimmt. Trotzdem ist davon auszugehen, dass 3D Systeme zu einer höheren Effektivität im Realbetrieb führen, da diese im Vergleich zu 2D Systemen eine tiefere Fehlalarmrate aufweisen (Flitton et al. 2013; Hättenschwiler et al. 2018; Mouton & Breckon 2015; Wells & Bradley 2012). Screener in der OSARP Bedingung brauchten

am längsten, da die genaue Befolgung des Protokolls längere RZ mit sich bringen. Mit Training sollte diese Zeit jedoch reduziert werden können. Der Vergleich zwischen USBV vorhanden versus nicht vorhanden zeigte bei USBV vorhanden längere RZ in der 2D ($p = 0.022$) und 3D Bedingung ($p < 0.001$), jedoch längere RZ für USBV nicht vorhanden in der OSARP Bedingung ($p = 0.041$). Dieses Ergebnis ist konsistent mit der Antworttendenz pro Bedingung.

4. Schlussfolgerung

Zusammenfassend konnte gezeigt werden, dass 2D Screener kein spezifisches Training benötigen, bevor sie erstmals mit neuer 3D CT Technologie arbeiten. Die Befolgung eines OSARP führte zu einer Steigerung der Erkennungsleistung und einer neutralen Antworttendenz bei 3D Bildern, wobei zukünftige Studien die Optimierung solcher Protokolle zur Alarmauflösung untersuchen sollten.

5. Literatur

- Bretz EA (2002) Slow takeoff. *IEEE Spectrum*, 39:37–39.
- Green DM, Swets JA (1966) *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Hättenschwiler N, Mendes M, Schwaninger A (2018) Detecting Bombs in X-Ray Images of Hold Baggage: 2D Versus 3D Imaging. *Human Factors*, doi: 10.1177/0018720818799215
- Halbherr T, Schwaninger A, Budgell G, Wales A (2013) Airport security screener competency: a cross-sectional and longitudinal analysis. *International Journal of Aviation Psychology*, 23:113-129.
- Holm S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70.
- Koller S, Drury C, Schwaninger, A (2009) Change of search time and non-search time in X-ray baggage screening due to training. *Ergonomics*, 52:644–656.
- Macmillan NA, Creelman CD (2005) *Detection theory: A users guide* (2nd ed.). New York: Cambridge University Press.
- Megherbi N, Flitton GT, Breckon TP (2010) A classifier based approach for the detection of potential threats in CT based baggage screening. *Proceedings of the International Conference on Image Processing ICIP, Hong Kong*, 1833–1836.
- Mouton A, Breckon TP (2015) A review of automated image understanding within 3D baggage computed tomography security screening. *Journal of X-ray Science and Technology*, 23:531–555.
- Rice S, McCarley J (2011) Effects of response bias and judgment framing on operator use of an automated aid in a target detection task. *Journal of Experimental Psychology: Applied*, 17:320–331.
- Schuster D, Rivera J, Sellers BM, Fiore SM, Jentsch F (2013) Perceptual training for visual search. *Ergonomics*, 56:1101–1115.
- Schwaninger A, Hofer F, Wetter O (2007) Adaptive computer-based training increases on the job performance of x-ray screeners. *Proceedings of the 41st Carnahan Conference on Security Technology, Ottawa*.
- Singh S, Singh, M (2003) Explosives detection systems (EDS) for aviation security. *Signal Processing*, 83:31–55.
- Turner S (1994) *Terrorist explosive sourcebook countering terrorist use of improvised explosive devices*. Boulder CO: Paladin Press.
- Vuong QC, Tarr JT (2004) Rotation direction affects object recognition. *Vision Research*, 44:1717–1730.
- Wells K, Bradley DA (2012) A review of X-ray explosives detection techniques for checked baggage. *Applied Radiation and Isotopes*, 70:1729–1746.



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