

# Potenziale multimodaler Benutzungsschnittstellen

Ansätze der Mensch-Maschine-Interaktion für die digitalisierte Produktion

Daniel-Leonhard Fox, André Ullrich und Norbert Gronau, Universität Potsdam

Die Digitalisierung verändert Fertigung und Produktion nachhaltig. Dabei ist die konkrete Ausgestaltung von Benutzungsschnittstellen im digitalisierten Produktionsumfeld von zentraler Bedeutung für eine erfolgreiche Transformation. In diesem Beitrag werden Einsatzpotenzial und Kombinationsmöglichkeiten von Ansätzen der Mensch-Maschine-Interaktion aufgezeigt sowie Anwendungsbeispiele vorgestellt. Dabei bieten multimodale Benutzungsschnittstellen einen hohen Grad an Immersion. Im Ergebnis werden Ansätze VR-, gesten- und sprachbasierter Interaktionsformen mithilfe der Grundsätze der Dialoggestaltung verglichen und deren Eignung im Einsatz als Arbeits- und Lernunterstützung für die Mitarbeiter dargestellt.

Grundlegender Bestandteil der digitalisierten Produktion sind Cyber-Physische Systeme (CPS), die informationstechnische und physikalische Welt verschmelzen lassen. Benutzungsschnittstellen auf Basis unterschiedlicher Ansätze der Mensch-Maschine-Interaktion (MMI) verbinden diese beiden Welten. Insbesondere technische Akteure und Systeme sind mit integrierten Verarbeitungs- und Kommunikationsfähigkeiten ausgestattet, die neue Ansätze für Benutzungsschnittstellen der MMI zugleich erfordern als auch befähigen. Benutzungsschnittstellen sind Komponenten von interaktiven Systemen, die Informationen und Steuerelemente bereitstellen, welche notwendig sind, damit die Benutzerin eine bestimmte Arbeitsaufgabe mit dem interaktiven System erledigen kann [1].

Interaktionen mittels traditioneller industrieller Benutzungsoberflächen beruhen in der Regel auf konventionellen unimodalen Benutzungsschnittstellen: ein System empfängt einen über Tastatur, Maus oder Touchscreen eingegebenen Befehl, woraufhin eine Systemantwort auf einem Bildschirm angezeigt wird. Insgesamt müssen Benutzungsschnittstellen zunehmend intuitiver und benutzungsfreundlicher werden, da vernetzte und intelligente Produktionsanlagen wesentlich

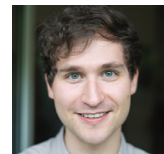
komplexer sind [2]. Ein vielversprechender Lösungsansatz zur Realisierung hoher Intuitivität und Benutzungsfreundlichkeit liegt im multimodalen Verbund der Interaktionen [3]. Technologische Fortschritte in den Bereichen Aktorik, Sensorik, Datenübertragung sowie Informationsverarbeitung bis hin zur künstlichen Intelligenz erhöhen die Usability neuartiger Interaktionskonzepte [4], wodurch auch der Arbeitsplatz selbst ergonomischer sowie Lernprozesse intuitiver gestaltet werden können. Die Weiterentwicklung von Benutzungsschnittstellen ist eine gesellschaftliche Gestaltungsaufgabe, da Lebensqualität und Produktivität von einem positiven Entwicklungsverlauf abhängen [4]. Beachtenswert sind im Paradigma der Industrie 4.0 auch die individuellen und unterschiedlichen mentalen Modelle der Arbeitskräfte, die das Verständnis und die Wahrnehmung der Akteure prägen. Um auf diese Unterschiede eingehen zu können, sind neue Ansätze notwendig, die Menschen in der Fabrik innerhalb ihrer jeweiligen mentalen Modelle unterstützen; insbesondere bei der immer umfassenderen und intelligenteren Vernetzung der Anlagen sowie der daraus resultierenden Komplexität als auch den großen Datenmengen.

## Potentials of Multimodal User Interfaces – A Comparison of Approaches of Human-Machine Interaction for the Digitalized Production Environment

Digitalization is changing manufacturing substantially. The design of user interfaces in the digitalized production environment is of utmost importance. In this article, potential combinations of approaches to human-machine interaction are shown and examples of applications are presented. Multimodal user interfaces offer a high degree of immersion. As a result, approaches of (among others) VR-, gesture- and speech-based types of interaction are compared with the dialogue principles and their suitability in practice as work and learning support for employees is presented.

### Keywords:

human-machine-interfaces (HMI), industry 4.0, multimodal user interfaces



Daniel-Leonhard Fox, M. Sc. ist wissenschaftlicher Mitarbeiter und Doktorand am Fachgebiet IT-Entrepreneurship des Hasso-Plattner-Instituts, Digital Engineering Fakultät, Universität Potsdam.



Dr. André Ullrich arbeitet im Rahmen der Nachwuchsforscherguppe PROMUT als Post-Doktorand am Lehrstuhl für Wirtschaftsinformatik, Prozesse und Systeme an der Universität Potsdam.



Univ.-Prof. Dr.-Ing. Norbert Gronau ist Inhaber des Lehrstuhls für Wirtschaftsinformatik, Prozesse und Systeme sowie Direktor des Forschungs- und Anwendungszentrum Industrie 4.0 an der Universität Potsdam.

aulrich@lswi.de.de  
www.lswi.de

MMI-Technologien für die Industrie 4.0

Zentrale MMI-Technologien stehen bislang noch am Anfang [4]. Der Entwicklungsstand befindet sich derzeit im Übergang von einer vorwiegend händisch stattfindenden Bedienung zu neuartigen Möglichkeiten, bei denen der gesamte Körper als Benutzungsschnittstelle fungiert und die Grenzen zwischen realer und virtueller Welt verschwimmen [5].

In Bild 1 werden Technologien aus dem letztveröffentlichten Hype-Cycle für MMI-Technologien hinsichtlich ihres Potenzials und Technologie-Reifegrad verortet. In der Abbildung sind die jeweiligen Schnittstellen einem Reifegrad zugeteilt, der den Technology Readiness Levels (TRL) entspricht [6]. So werden Potenzial, aktueller Reifegrad sowie prognostizierte Dauer bis zum Erreichen der Marktreife für die jeweilige Schnittstelle ersichtlich [6].

Der Reifeprozess beginnt mit Beobachtung und Beschreibung des Funktionsprinzips und dem darauffolgenden Nachweis der Funktionstüchtigkeit der Technologie. In diesen Phasen befinden sich momentan haut- und gestenbasierte Schnittstellen sowie Gehirn-Computer-Schnittstellen, die bis zur breiten Marktreife noch über 10 Jahre weiterentwickelt werden müssen [7]. Hautschnittstellen werden insbesondere durch sehr dünne hautähnliche Materialien ermöglicht, in denen Sensortechnik eingelassen ist. Die Sensoren erfassen Berührung, Dehnung, Feuchtigkeit oder Temperatur und werden insbesondere im multimodalen Verbund mit anderen Schnittstellen in der Produktion eingesetzt und weisen daher ein hohes Nutzungspotenzial auf. Gestenbasierte Schnittstellen beruhen auf der mathematischen Interpretation menschlicher Bewegungen durch einen Computer [8], die u. a. durch Kameras, Tiefen- und Beschleunigungssensoren erfasst werden. Ein produktiver Einsatz ist in fünf bis zehn Jahren vorstellbar, sofern eine zuverlässige und sichere Erkennung sicherheitsrelevanter Gesten möglich ist [8].

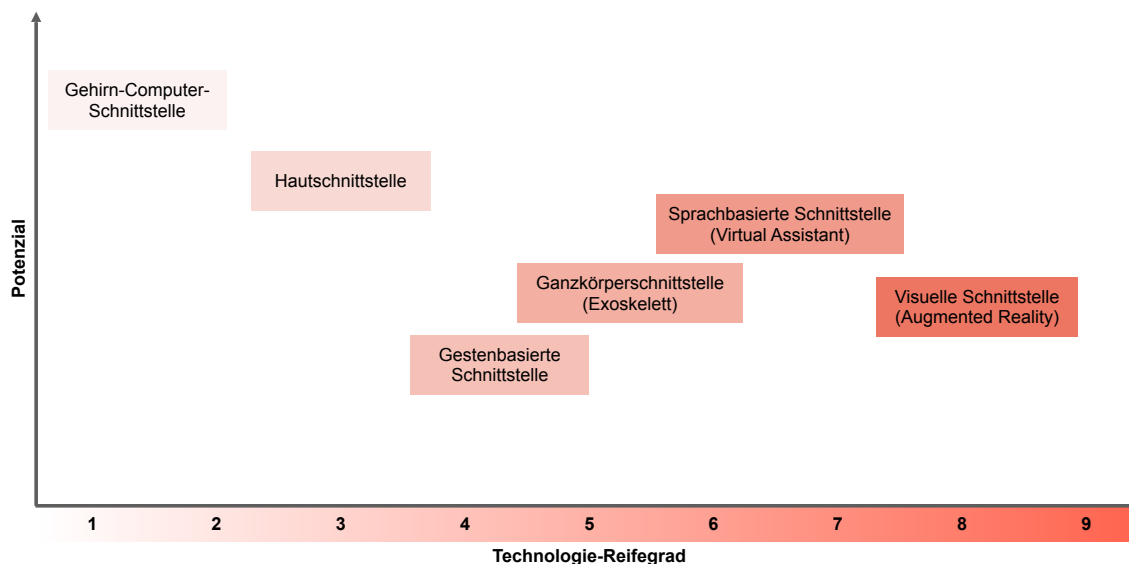
Auf dem Hochpunkt des erwarteten Potenzials, das über den bisherigen Möglichkeiten der Innovation liegt, befinden sich derzeit Gehirn-Computer-Schnittstellen (BCI). BCI messen die Aktivität des Zentralnervensystems und wandeln diese in künstlichen Output um [9]. Wenn gleich die technische Reife derartiger Schnittstellen noch gering ist und große Bedenken bezüglich des Datenschutzes bestehen [10], ist hier ein Einsatz in der Fabrik in deutlich über zehn Jahren vorstellbar. Der derzeitige Funktionsumfang beschränkt sich jedoch auf erste Erfolge bei der Steuerung von Roboterarmen [11].

Etwas ausgereifter sind Exoskelette, bei denen es sich um mechanische Strukturen handelt, die am Körper getragen werden und dessen natürliche Fähigkeiten verstärken. Testweise sind solche Skelette bereits in der Produktion im Einsatz, während das Potenzial dieser Schnittstelle derzeit gerade erst anwächst.

Sprachbasierte Schnittstellen, die auch ein hohes Nutzungspotenzial aufweisen, sind im privaten Bereich durch virtuelle Assistenten wie Alexa oder Siri bereits weit verbreitet, können jedoch auch in der Produktion zum Einsatz kommen. Bislang treten allerdings hohe Fehlerraten beim Erkennen natürlicher Sprache auf, weshalb diese erst prototypisch in Fabriken zum Einsatz kommen. Ein Produktiveinsatz ist hier erst in zwei bis fünf Jahren vorstellbar [12].

Zuletzt befinden sich visuelle Schnittstellen bereits in Fabriken im Einsatz. Im Produktionsbereich werden mit Augmented Reality (AR) relevante Informationen direkt ins Sichtfeld der Mitarbeiterinnen gerückt. Der gewünschte Immersionseffekt und damit das Nutzungspotenzial wird jedoch oftmals getrübt, da langes Tragen, Latenzen und eine dezentrale Bildpositionierung von AR-Brillen u. a.

Bild 1: Hype Cycle für Mensch-Maschine-Schnittstellen (eigene Abbildung in Anlehnung an [7]).



Zuletzt befinden sich visuelle Schnittstellen bereits in Fabriken im Einsatz. Im Produktionsbereich werden mit Augmented Reality (AR) relevante Informationen direkt ins Sichtfeld der Mitarbeiterinnen gerückt. Der gewünschte Immersionseffekt und damit das Nutzungspotenzial wird jedoch oftmals getrübt, da langes Tragen, Latenzen und eine dezentrale Bildpositionierung von AR-Brillen u. a.

Kopfschmerzen und Ermüdungserscheinungen auslösen können.

## Potenzielle multimodaler MMI

Für die nutzerfreundliche Interaktion von Mitarbeitern und CPS sind multimodale MMI-Konzepte geeignet [13], in denen zwei oder mehr Konzepte miteinander kombiniert werden [14]. Multimodale Systeme erlauben eine leistungsstarke sowie intuitive Bedienung und haben das Potenzial, die Interaktion zwischen Mensch und Maschine bspw. durch eine höhere Immersion zu verbessern [3]. So wird die Zuverlässigkeit durch die Kombination von Teilinformationsquellen erhöht und eine Personalisierung je nach Benutzer und Kontext sowie eine Nutzung durch mehrere Mitarbeiter ermöglicht [3]. Nur wenn CPS durch mehrere Schnittstellen mit dem Benutzer interagieren, können sie anhand intelligenter Systeme komplexe Aufgaben und Prozesse gut verstehen und durchführen [3]. Nachteile einzelner Konzepte in multimodalen Systemen werden zudem ausgeglichen, während die Bedienung unter Verwendung mehrerer menschlicher Sinne natürlicher wird [3]. Mögliche Kombinationen werden in Bild 2 nach potenzieller und aktuell vorstellbarer Kompatibilität zwischen den vorgestellten Bedienkonzepten in einem Ampelsystem aufgeführt. Da die MMI sowohl in Input- als auch in Outputschnittstellen klassifiziert werden können, ergeben sich unterschiedliche Substitutionspotenziale. Inputschnittstellen wie E-Skins oder BCI sind dadurch gekennzeichnet ausschließlich Signale zu empfangen, während Outputschnittstellen Signale senden. Schnittstellen wie virtuelle Assistenten können Signale sowohl empfangen als auch senden, wodurch sie in beide Kategorien fallen. Unauffindbare Kombinationen sind mit „Niedrig“ oder „Mittel“ gekennzeichnet, was einem niedrigen bzw. mittleren Substitutionspotenzial entspricht. Unauffindbar sind folgerichtig insbesondere Kombinationen mehrerer Outputschnittstellen. So sind insbesondere BCI, E-Skins und Gestensensorik untereinander substituierbar, da diese die Funktionen des jeweils anderen Konzepts beinhalten können.

## Vergleich multimodaler MMI

Allgemein ist AR das derzeit vielversprechendste Konzept für den multimodalen Einsatz, da es mit allen vorgestellten anderen Konzepten

		Inputschnittstellen			Outputschnittstellen		
		Augmented Reality	Virtual Assistant	Exoskelett	BCI	E-Skin	Gestensensorik
Inputschnittstellen	Augmented Reality		Hoch	Hoch	Hoch	Hoch	Hoch
	Virtual Assistant			Hoch	Mittel	Mittel	Hoch
	Exoskelett				Hoch	Mittel	Mittel
Outputschnittstellen	BCI					Niedrig	Niedrig
	E-Skin						Niedrig
	Gestensensorik						

ein hohes Kombinationspotenzial aufweist (Bild 1) und bereits in einigen Studien als Teil multimodaler Schnittstellen erforscht wurde. Auch in der Praxis findet das Konzept bereits Verwendung in der Microsoft HoloLens, das ein multimodales System mit AR, Gesten- und Sprachsteuerung ist. Schnittstellen wie AR müssen gesteuert werden, wozu mindestens eine weitere Schnittstelle zur Interaktion mit der Maschine benötigt wird. Auch für Exoskelette ergeben sich mittlere bis hohe Kombinationspotenziale mit den vorgestellten Ansätzen. So befinden sich in einem Prototypen eines multimodalen Exoskeletts mehrere gestenerfassende Sensoren im Handschuh, um bspw. einen Seilzug zu steuern, der eine Hebehilfe darstellt [15].

Zentral für die Interaktion von Menschen und Maschinen ist die Usability, also Nutzungstauglichkeit der Benutzungsschnittstelle. Kriterien für die Gestaltung derartiger „Dialoge“ sind in der ISO Norm 9241-110, auch bekannt als Grundsätze der Dialoggestaltung festgehalten [1, 16]. Die sieben Kriterien werden dem folgenden Vergleich der vorgestellten Konzepte zugrunde gelegt. Dabei wird eine Nutzung im industriellen Kontext angenommen. Anzumerken ist hierbei, dass einige der Kriterien mehr oder weniger auch von der verwendeten Software abhängen und damit für die Schnittstellen an sich unter Vorbehalt gegeben sind.

Das Kriterium der Aufgabenangemessenheit betrachtet, inwieweit die Benutzerin bei der effektiven und effizienten Aufgabenausführung unterstützt wird. Dabei sollen für eine vorliegende Aufgabe irrelevante Informationen auf ein Minimum reduziert werden sowie Eingabe- und Ausgabeformate auf aufgabenspezifische

**Bild 2: Kombinationspotenzial von Benutzungsschnittstellen.**

## Literatur

- [1] Handbuch Usability: ISO 9241. URL: [www.handbuch-usability.de/iso-9241.html](http://www.handbuch-usability.de/iso-9241.html), Abrufdatum 15.06.2020.
- [2] Vogel-Heuser, B.; Bauernhansl, T.; Ten Hompel, M. (Hrsg.): Handbuch Industrie 4.0. Bd. 4: Allgemeine Grundlagen, 2. Auflage. Berlin 2017.
- [3] Dumas, B.; Lalanne, D.; Oviatt, S.: Multimodal Interfaces: A Survey of Principles, Models and Frameworks. In: Lecture Notes in Computer Science (2009), S. 3-26.
- [4] Deutsche Akademie der Technikwissenschaften (Hrsg.): Innovationspotenziale der Mensch-Maschine-Interaktion. München 2016.
- [5] Reply: Zukunftskonzepte für Human Machine Interfaces - So kommen Menschen den Maschinen näher. URL: [www.reply.com/de/topics/artificial-intelligence-and-machine-learning/human-machine-interfaces-trend-report](http://www.reply.com/de/topics/artificial-intelligence-and-machine-learning/human-machine-interfaces-trend-report), Abrufdatum 15.06.2020.
- [6] Héder, M.: From NASA to EU: the evolution of the TRL scale in Public Sector Innovation. In: The Innovation Journal 22 (2017) 1.
- [7] Goertz, W.; Gartner: Hype Cycle for Human-Machine Interface. 2018. URL: [www.gartner.com/doc/3882879/hype-cycle-human-machine-interface](http://www.gartner.com/doc/3882879/hype-cycle-human-machine-interface), Abrufdatum 15.06.2020.

[8] Liu, H.; Wang, L.: Gesture recognition for human-robot collaboration: A review. In: International Journal of Industrial Ergonomics 68 (2018), S. 355-367.

[9] Vukelic, M.; Pollmann, K.; Peisner, M.; Spath, D.: Mensch-Technik-Interaktion mit Emotionen: Eine neurowissenschaftliche Untersuchung zu hirnpfysiologischen Prozessen emotional-affektiver Nutzerreaktionen die für technische Assistenz relevant sind. In: Freitag, M. (Hrsg): Mensch-Technik-Interaktion in der digitalisierten Arbeitswelt (2020), S. 39-58.

[10] Mittelstand 4.0 - Kompetenzzentrum Saarbrücken: Brain-Computer-Interfaces: Einsatz von Brain-Computer-Interfaces als Mittel der Mensch-Technik-Interaktion. URL: kompetenzzentrum-saarbruecken.digital/technologieradar-brain-computer-interfaces, Abrufdatum 15.06.2020.

[11] Hochberg, L. R. u. a.: Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. In: Nature 485 (2012) 7398, S. 372-375.

[12] Myers, C.; Furqan, A.; Nebolsky, J.; Caro, K.; Zhu, J.: Patterns for How Users Overcome Obstacles in Voice User Interfaces. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18 (2018), S. 1-7.

[13] Gorecky, D.; Schmitt, M.; Loskyll, M.; Zühlke D.: Human-machine-interaction in the industry 4.0 era. In: 12th IEEE International Conference on Industrial Informatics (IN-DIN) (2014), S. 289-294.

[14] Oviatt, S.: Advances in robust multimodal interface design. In: IEEE Computer Graphics and Applications 23 (2003) 5, S. 62-68.

[15] Bundesministerium für Bildung und Forschung: Logistik 4.0, mit dem Exoskelett SensHand gegen Gesundheitsschäden in der Lagerhaltung. URL: www.wissenschaftsjahr.de/2018/neues-aus-den-arbeitswelten/alle-aktuellen-meldungen/april-2018/exoskelett-senshand-bietet-sich-als-hebehilfe-in-der-lagerhaltung-an, Abrufdatum 15.06.2020.

[16] Böde, E.; Hartmann, E. A.; Lüdtke, A.; Oppenheimer, F.; Rötting, M.; Wegerich, A.: Mensch-Technik-Interaktion. In: Bundesministerium für Wirtschaft und Technologie (BMWi) Öffentlichkeitsarbeit: Autonomik - Autonome und simulationsbasierte Systeme für den Mittelstand 3. Berlin 2018.

[17] Teichmann, M.; Ullrich, A.; Wenz, J.; Gronau, N.: Herausforderungen und Handlungsempfehlungen betrieblicher Weiterbildungspraxis in Zeiten der Digitalisierung. HMD Praxis der Wirtschaftsinformatik. Faktor Mensch – Erwachsenenbildung. 2020. DOI: 10.1365/s40702-020-00614-x.

Informationen anpassbar sein [16]. Dieses Kriterium ist insbesondere bei AR und virtuellen Assistenten kritisch, da diese durch ihre informationsbereitstellende Funktion direkt betroffen sind. So ist eine Informationsüberflutung stets zu vermeiden. Mit Ausnahme des BCI kann dieses Kriterium für die Anderen als erfüllbar angesehen werden. BCI sind potenziell direkt an das Gehirn angebunden. Durch die fehlende Möglichkeit, die Hirnaktivität des Benutzers zu kontrollieren, wäre das ungewollte Senden aufgabenunspezifischer Informationen vorstellbar. Ohne eine weitere Modalität könnte demzufolge auch keine Unterstützung bei der Aufgabenausführung erfolgen.

Selbstbeschreibungsfähigkeit ist gegeben, wenn dem Benutzer unmittelbar jeder einzelne Bedienschnitt verständlich ist oder eine einfache, anforderbare Erläuterung zum Verständnis führt [16]. Dieses Kriterium ist insbesondere für virtuelle Assistenten von Bedeutung, da sie isoliert betrachtet ausschließlich über eine Sprachaus- und eingabe verfügen. Dies kann unter Umständen dazu führen, dass Bedienabläufe und mögliche Funktionen für den Benutzer unklar bleiben. Bei Schnittstellen, wie Gestensensoren oder E-Skins müssen ggf. zuerst spezielle Gesten erlernt werden, was ebenfalls nicht das Kriterium der unmittelbaren Verständlichkeit erfüllt. Exoskelette unterstützen intuitiv die Bewegungen des Nutzers, wodurch keine weiteren Verständlichkeitsbarrieren auftreten sollten. Im Falle von AR kann primär eine softwareseitige Ermöglichung der Fähigkeit erfolgen. BCI hingegen sind bislang auf ein (komplexes) Training der Gehirnaktivität zur Steuerung angewiesen und erfüllen das Kriterium somit derzeit nur sehr eingeschränkt.

Die folgenden vier Kriterien hängen von der verwendeten Software ab, können also für die vorgestellten MMI als gegeben angesehen werden. Steuerbarkeit umfasst die Kontrollierbarkeit des Ablaufs, dauerhafte Eingriffsmöglichkeiten sowie Korrekturmöglichkeiten [16]. Es ist insbesondere bei der Gestensteuerung kritisch, da auch sicherheitsrelevante Gesten absolut zuverlässig erkannt werden müssen.

Erwartungskonformität wird als konsistente Bedienung die den Erwartungen, Wissen und Erfahrungen der Bediener entspricht definiert [16]. Zuletzt sind die Kriterien Fehlertoleranz, die eine Fehlerkorrektur mit minimalem Aufwand ermöglichen und den Nutzer bei der Aufspürung und Beseitigung von Fehlern unterstützen soll [16], sowie Individualisierbarkeit, die die Anpassung an individuelle Aufgaben, Arbeitsumstände, Fähigkeiten und Vorlieben

ermöglichen soll, für alle Ansätze hardwareseitig gegeben.

Die Lernförderlichkeit sollte hoch sein, damit Abläufe und Inhalte dauerhaft im Gedächtnis verbleiben können [16]. Hierbei ist u. a. maßgeblich, wie natürlich die Schnittstelle wirkt und inwiefern bei Aus- und Weiterbildung der Umgang geschult wird [17]. Nach aktuellem Entwicklungsstand haben etwa Sprachassistenten Schwierigkeiten mit natürlichen Befehlen. Bei der Gestensteuerung kommt es darauf an, ob diese nahe an der natürlichen Verwendung von Gesten sind, um dieses Kriterium besser zu erfüllen. BCI erfüllen es erst, wenn sie ohne größeren Lernaufwand funktionieren und eine Gedankensteuerung ohne weiteres Training erlauben, was derzeit jedoch noch nicht absehbar ist. Den MMI ist gemeinsam, das sie inhärent natürliche Abläufe ermöglichen, wodurch ein leichter Verbleib im Gedächtnis der Schnittstellenbedienung angenommen werden kann.

## Zusammenfassung

In der digitalisierten Fertigung verleihen selbst organisierte Produktionsprozesse Mitarbeitern mit mobilen Arbeitsplätzen mehr Verantwortung. Diese Rolle eines kreativen Problemlösers kann durch multimodale Schnittstellen weitgehend unterstützt werden, während unimodale Schnittstellen angesichts zunehmender Datenmengen nicht mehr ausreichen. Multimodale Schnittstellen und Möglichkeiten der Immersion sind von besonderem Stellenwert, da sie Zugangsbarrieren für Fabrikmitarbeiter senken und somit die Bewältigung zunehmender Komplexität unterstützen. Dieser erleichterte Zugang durch neuartige Prozesse und Vorgehensweisen macht zudem eine Neuausrichtung von Arbeit und Lernprozessen denkbar, wodurch Arbeitskräfte auch im volatilen Umfeld beständig und variabel eingesetzt werden können.

*Dieser Beitrag entstand im Rahmen der Nachwuchsforschungsgruppe ProMUT „Nachhaltigkeitsmanagement 4.0 - Transformative Potentiale digital-verbundener Produktion für Mensch, Umwelt und Technik“ (Kennzeichen 01UU1705B), die vom Bundesministerium für Bildung und Forschung in Rahmen der Förderinitiative „Sozial-ökologische Forschung“ gefördert wird.*

Schlüsselwörter:

Mensch-Maschine-Interaktion (MMI), Industrie 4.0, Multimodale Benutzungsschnittstellen