

## Technische Beschreibung einer dynamischen und einer statischen Variante der Zeigegeste

Tobias NOWACK, Sabine WENZEL, Peter KURTZ

*Fachgebiet Arbeitswissenschaft, TU Ilmenau  
Max-Planck-Ring 12, D-98693 Ilmenau*

**Kurzfassung:** Die Mensch-Roboter-Kooperation (MRK) als ein Schwerpunkt in der digitalisierten Fabrik („Industrie 4.0“) verlangt alternative Bedien- und Steuerungsstrategien. Die gestenbasierte Steuerung des Roboters kommt der menschlichen Interaktionsform sehr nahe, ist aber gleichzeitig auch problembehaftet. Eine solche Geste besteht aus fünf Phasen. Die eigentliche Gestenausführung ist dabei ein dynamischer Prozess, während gerade das Zeigen durch eine abschließende statische Körperhaltung die Information überträgt. Theoretisch sollte das Zeigen in der dynamischen sowie der statischen Phase einer Geste beschrieben werden können. Aktuelle Daten zeigen aber, dass bei der Erkennung der intuitiv genutzten Zeigegeste eine dynamische Beschreibung die Geste nur qualitativ erfasst.

**Schlüsselwörter:** Gestenerkennung, Zeigegeste, Mensch-Roboter-Kooperation (MRK)

### 1. Hintergrund

Ein Vorzug von „Industrie 4.0“ bzw. deren Auswirkung ist die konsequente Individualisierbarkeit von Produkten (Stückzahl 1) zu Kosten, die bisher nur im Kontext der Massenfertigung erreicht wurden. Um dieser Anforderung sowie den Anforderungen aus der aktuellen Altersstruktur in der manuellen Montage gerecht zu werden, ist auch die Mensch-Roboter-Kooperation (MRK) weiter zu entwickeln.

Im Rahmen des BMBF-geförderten Projektes „KLARA - kleine autonome Transport- und Handhabungshilfe“ wird für die Zielgruppe 60+ ein persönliches, mobiles Unterstützungssystem zum Heben und Tragen von Kleinladungsträgern entwickelt. In der Altersgruppe 50+ sind laut Gesundheitsreport der Techniker Krankenkasse von 2014 (Grobe 2014) über 150 Arbeitsunfähigkeitstage (AU-Tage) je 100 Versicherten aufgrund von Rückenschmerzen gemeldet. Die Verkehrs- und Lagerberufe belegen dabei mit 283 AU-Tagen je 100 Versicherte Platz 4 nach den Bau-, Bauneben-, Holz- und Metallberufen.

Zur Auswahl der Kleinladungsträger (KLT; Schäferkisten mit einer Größe von max. 60 cm x 40 cm und einem Gewicht von bis zu 50 kg) ist eine ergonomisch günstige Körperhaltung zu wählen. Orientiert an der Haltungswichtung, beschrieben unter anderem in der Leitmerkmalmethode Heben und Tragen (BAuA 2015), ist eine Bedienung (für KLARA die Kommandos: Heben, Tragen, Absetzen) in stehender Körperhaltung vorzuziehen.

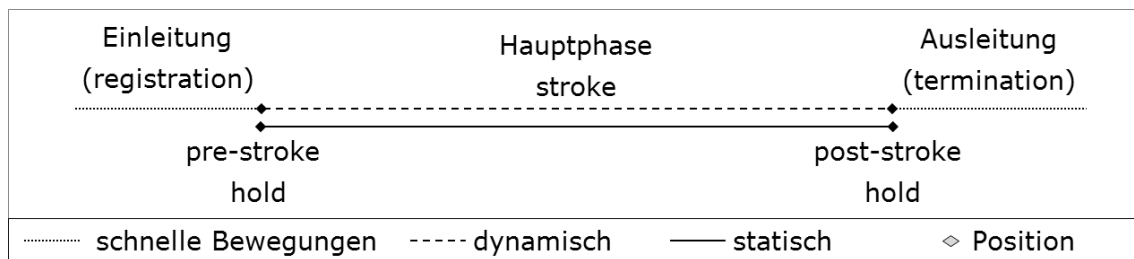
In einem zweistufigen Auswahlprozess soll für die Grobauswahl des gewünschten KLT die Zeigegeste verwendet werden, die Feinauswahl bzw. Bestätigung erfolgt über einen Touchscreen. Dieser Ansatz entspricht in Teilen der Forderung nach berührungsloser Interaktion (Strassmann 2010) bei der interaktiven Zusammenarbeit

von Mensch und Roboter im gemeinsamen Arbeitsraum (Koch 2002). Die zusätzliche Bestätigung erlaubt dem Menschen, Erkennungsfehler zu beheben oder unsichere Aktionen zu unterbinden.

Die Nutzung einer Spracherkennung ist für das Projekt nach den zurzeit in der Industrie gültigen Lärmvorschriften ausgeschlossen. Für die Werkhalle sind bis zu 80 dB(A), teilweise noch bis zu 85 dB(A) erlaubt (LärmVibrationsArbSchV 2007). Die für eine erfolgreiche Zusammenarbeit notwendige Kommunikation mit Spracherkennung ist somit nur eingeschränkt oder mit hohem technischem Aufwand (z.B. Funkmikrofon am Werker) möglich.

## 2. Grundlagen der Gestensteuerung

Das Fünf-Phasen-Modell (s. Abb. 1) der sprachbegleitenden Gesten ist die Grundlage für die Entwicklung einer industriell nutzbaren Gestensteuerung. Das von der Spracherkennung bekannte „live mic problem“ (Wigdor & Wixon 2011) ist auch für eine Gestensteuerung in der „Registrierungsphase“ zu lösen. Bei der sprachbegleitenden Geste wie sie von McNeill (1992) untersucht wurde, ist das notwendige Startkommando aus den Kontextinformationen der Rede zu entnehmen. Zur Erkennung beliebiger Freiraumgesten (wie z.B. Zeigen auf den KLT) müssen zusätzlich andere Kontextinformationen genutzt werden. Dey (2001) bezeichnet als Kontextinformationen jedes Wissen, das zur Charakterisierung der Situation genutzt werden kann. Für die „regISTRATION“ heißt das, es muss „DER BEDIENER“ erkannt werden.



**Abbildung 1:** Fünf-Phasen-Modell der Gestenausführung nach McNeill (1992), Kendon (2004), Walter et al. (2013) und Pavlovic et al. (1997)

Was macht den Nutzer zum Bediener? Johannsen (1993) definiert den Benutzer als passiv in der Nutzung, den Bediener jedoch als interaktiven Nutzer des Mensch-Technik-Systems. Somit ist der Wunsch zur Nutzung des Assistenzsystems entscheidend. Dazu muss das System, den Menschen als Nutzer erkennen und anschließend **den** Bediener, der das Kommando übernimmt, identifizieren. Dazu sind verschiedene Ansätze in der MRK denkbar, z. B. der Einsatz von RFID-Tags zur Bediener- oder Objekterkennung<sup>1</sup>, aber auch eine einfache, durch den Bediener im Rahmen der „regISTRATION“ und dem anschließenden „pre-stroke hold“ (s. Abb. 1) durchgeführte Anmeldegeste.

Auch für die Zielerkennung und damit die Verbesserung der Gestenerkennung sollten bekannte Kontextinformationen eingesetzt werden. So wird in der Mensch-Mensch-Interaktion die Zeigegeste überwiegend nur eingesetzt, um dem Interakti-

<sup>1</sup> Vorgestellt durch Firma SICK auf der SPS IPC Drives 2015, Nürnberg vom 24.-26.11.2015

onspartner eine grobe Richtungsinformation mitzuteilen. Die Information, nach welchem Zielobjekt der Kommunikationspartner suchen muss, wird meist verbal übermittelt.

Dank dieser Kontextinformationen ist für die Kommunikationspartner der folgende nonverbale Informationsaustausch eindeutig und problemlos durchführbar, unabhängig davon, wie in der Hauptphase („stroke“, „post-stroke hold“) die Geste ausgeführt wird. Eine Geste ist in der Hauptphase eine dynamische Bewegung (Kendon 2004; McNeill 1992) und kann als reine dynamische Geste („stroke“) oder auch als erreichter statischer Zustand („post-stroke hold“) beschrieben werden.

### 3. Gestenbeschreibung

Zur Analyse der Zeigegeste wird auf die zwei beschriebenen, grundsätzlichen Ansätze (statisch/dynamisch) zurückgegriffen. Dabei gilt das besondere Augenmerk der dynamischen Beschreibung, da eine statische Variante bereits für die MRK implementiert wurde und zurzeit ausgiebig getestet wird (Nowack et al. 2015a). Diese statische Beschreibung wird zunächst kurz vorgestellt.

#### 3.1 Umsetzung (statische Beschreibung)

Im Rahmen des Projektes „KLARA“ wird die Zeigegeste mit der statischen Beschreibung während des „post-stroke hold“ interpretiert. In der „registration“ meldet sich der Bediener am System an und definiert mit der Anmeldung den Arm, mit dem die Geste ausgeführt wird. Systemseitig wird dazu in zwei Stufen erst der Nutzer mit Hilfe der Gesichtserkennung (OpenCV-Bibliothek) und anschließend mit der Skeletterkennung (Nite-Bibliothek) identifiziert. Damit der Nutzer als Bediener interpretiert wird, muss er die Anmeldepose (Hand auf/über Kopfhöhe) einnehmen.

Wenn der Bediener erkannt wurde, signalisiert das System, dass es bereit ist für die Erkennung der Zeigegeste. Dazu werden momentan farbige LED am System entsprechend des angepassten Ampelprinzips eingesetzt:

- Rot: System sucht den Bediener
- Gelb: System erwartet die Zeigegeste
- Gelb/Grün: Pose (statische Körperhaltung) der Zeigegeste erkannt
- Grün: Zeigevektor erkannt.

Die Zeigepose ist beschrieben durch einen Elevationswinkel am Schultergelenk von größer  $20^\circ$  und einem Extensionswinkel des Ellbogengelenkes von größer  $160^\circ$ . Da die Zeigegeste von den Bedienern als statische Beschreibung einer Körperhaltung gesehen wird, gilt die Pose als bestätigt, wenn der Bediener diese über ca. 1 s unverändert hält.

Für die Zielerkennung wird der durch die Geste beschriebene Zeigevektor (Skelettpunkt-Kopf Richtung Skelettpunkt-Hand) weiter geführt, bis dieser eine erkennbare Tiefenebene durchstößt. Da durch die Anforderungen im Projekt bekannt ist, dass sich an dem ermittelten Zielpunkt ein KLT oder eine für einen KLT ausreichende Freifläche befinden muss, können diese Kontextinformationen eingesetzt werden, um die Zielobjekterkennung robuster zu gestalten.

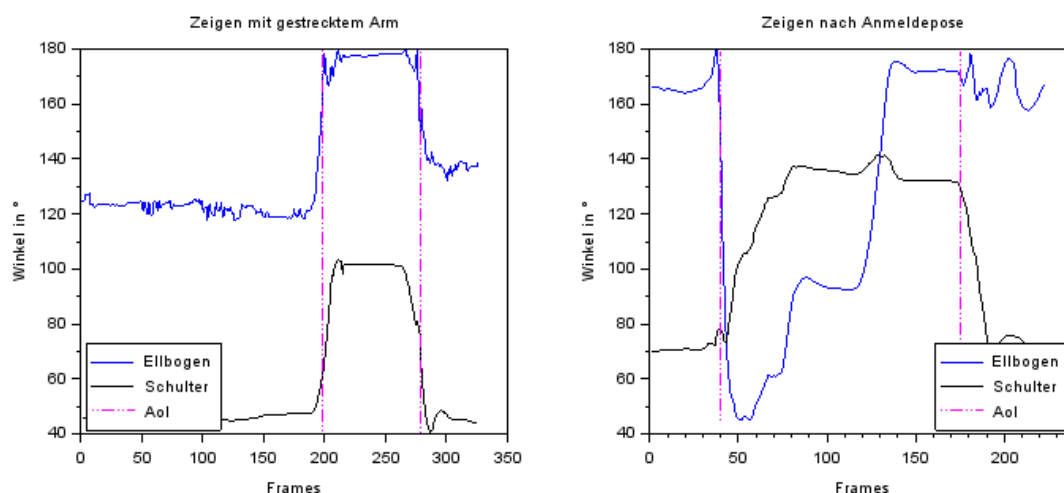
### 3.2 Alternativer Ansatz (dynamische Beschreibung)

Alternativ gilt es, eine dynamische Beschreibung des Zeigevorganges zu evaluieren. Unter einer dynamischen Beschreibung wird die Definition der Geste ausschließlich in der „stroke“ Phase ohne anschließenden „post-stroke hold“ verstanden.

Für die dynamische Interpretation der Geste müssen die Skelettpositionen aufeinanderfolgender Bilder die Bewegung sinnvoll beschreiben. Dazu sollen ca. 30 Frames (1 s) interpretiert werden. Die dynamische Geste kann somit als Abfolge diskreter Zustände, aber auch als echter kontinuierlicher Prozess betrachtet werden.

Aus den zurzeit vorliegenden Daten lassen sich unterschiedliche Bewegungsabläufe zur Erreichung der Zielpose (statische Zeigegeste, „post-stroke hold“) erkennen. Diese können sich grundlegend unterscheiden. Zwei dieser Möglichkeiten werden im Folgenden vorgestellt:

In der ersten Variante führt der Bediener den bereits gestreckten Arm (Extension des Ellbogenwinkels  $> 160^\circ$ ) aus der Ruheposition (Arm hängt entspannt seitlich des Körpers) direkt in die Zielrichtung und verharrt bei Erreichen kurz (Abb. 2 links). Alternativ wird die Hand des Bedieners aus der Anmeldepose durch Strecken in Zielrichtung bewegt. Der für die Anmeldepose  $90^\circ$  betragende Ellbogenwinkel steigt monoton bis ca.  $160^\circ$  (Abb. 2 rechts, Frame 125 bis 140). Diese Bewegung der Hand kann teilweise auch mittels linearer Regression in den drei Raumebenen abgebildet werden, wird aber durch die anatomischen Einschränkungen der beteiligten Gelenke immer von einer Kreisbewegung überlagert.



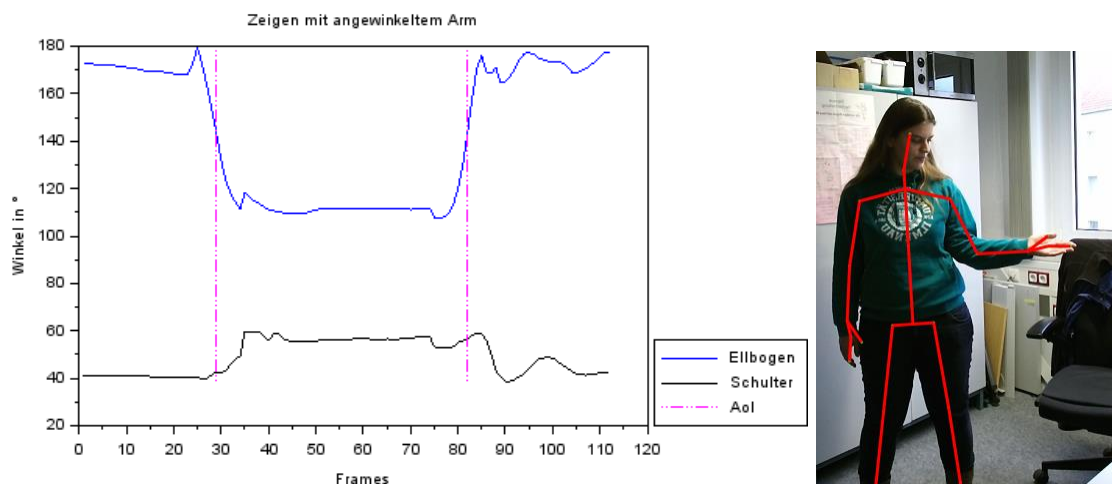
**Abbildung 2:** Charakteristische Winkeländerung an Ellbogen und Schulter während des Zeigens (links) mit gestrecktem Arm, (rechts) nach der Anmeldepose (ab Frame 125)

Neben den charakteristischen Winkeländerungen bei der Bewegung ist beim Zeigen immer auch eine statische Beschreibung möglich und für die Identifikation des Zeigevektors notwendig. Im Rahmen einer Fragebogenuntersuchung mit 90 Probanden wurde die Zeigegeste bereits im Vorfeld als statische Geste bestätigt (Nowack et al. 2015b). Diese Einschätzung wird durch die vorgestellten Daten belegt. Das bedeutet, dass die Auswertung der Pose (siehe Abschnitt 3.1) in der „post-stroke“ Phase sinnvoller ist, und die dynamische Definition kann bei der Erkennung der Zeigegeste nur unterstützend wirken.

#### 4. Diskussion und Ausblick

Gesten enthalten unabhängig vom Nutzerverständnis mindestens eine dynamische „stroke“ Phase (siehe Abschnitt 2). In diesem Artikel sollte gezeigt werden, dass diese dynamische Phase auch zur Beschreibung der Zeigegeste verwendet werden kann. Theoretisch ist ein dynamischer Ablauf des Zeigens entlang einer virtuellen Linie (vergleichbar der Handbewegung beim Werfen eines Dartpfeils) möglich. Der Bediener führt diese Geste jedoch meist nicht entsprechend exakt und ohne zusätzlichen, sich anschließenden Halt („post-stroke hold“) aus. Damit ist es für die technische Erkennung einfacher, die Zeigegeste in dieser Haltephase zu interpretieren und den Zeigevektor abzuleiten.

Die Bewegungsabläufe beim Zeigen können wie an den zwei Beispielen gezeigt sehr unterschiedlich ausfallen. Neben den Zeigeposen mit gestrecktem Arm, sind auch Zeigeposen mit angewinkelterm Arm (Abb. 3) gebräuchlich. Bei dieser Ausführungsvariante der Zeigegeste ist ein vergleichbarer charakteristischer Bewegungsablauf erkennbar, auch wenn die Winkel an Schulter und Ellenbogen invertiert und anders ausgeprägt sind.



**Abbildung 3:** links) Charakteristische Winkeländerung an Ellenbogen und Schulter während des Zeigens mit angewinkelterm Arm rechts) Körperhaltung des Probanden

Soll eine beliebige, intuitiv-ausgeführte Zeigegeste durch ein technisches System automatisch erkannt werden, sind noch weitere Varianten des Zeigens zu evaluieren und über einen Entscheidungsbaum richtig zuzuordnen. Ergänzt werden sollte dieser durch eine geeignete Erkennung zusätzlicher Kontextinformationen. Bediener aktueller MRK-Systeme ohne Kontexterkenkung und mit eingeschränktem Entscheidungsbaum werden spezielle Gesten zur Bedienung und Kommunikation erlernen müssen.

## 5. Literatur

- Bundesanstalt für Arbeitsschutz und Arbeitsmedizin - BAuA (2015). Gefährdungsbeurteilung mithilfe der Leitmerkmalmethode.  
<http://www.baua.de/de/Themen-von-A-Z/Physische-Belastung/Gefaehrdungsbeurteilung.html>, Zuletzt geprüft 16.Dezember 2015.
- Dey, A. K. (2001): Understanding and Using Context, Personal Ubiquitous Comput. 5:4-7
- Grobe, T. (2014): Gesundheitsreport 2014 – Veröffentlichungen zum Betrieblichen Gesundheitsmanagement der TK, Band 29. Hsg. Techniker Krankenkasse. Schmid Druck+Medien, Kaisheim.
- Johannsen, G. (1993): Mensch-Maschine-Systeme. Berlin: Springer, Springer-Lehrbuch
- Kendon, A. (2004): Gesture. Visible action as utterance. Cambridge University Press, Cambridge.
- Koch, B. (2002): Mensch und Roboter - ein tolles Team. Hg. v. Fraunhofer Magazin.  
<http://www.archiv.fraunhofer.de/archiv/magazin/pflege.zv.fhg.de/german/publications/df/df2002/mag2-2002-48.pdf>, Zuletzt geprüft 16.Dezember 2015.
- Lärm- und Vibrations-Arbeitsschutzverordnung vom 6. März 2007 (BGBl. I S. 261), die zuletzt durch Artikel 3 der Verordnung vom 19. Juli 2010 (BGBl. I S. 960) geändert worden ist.  
[http://www.gesetze-im-internet.de/bundesrecht/l\\_rmvibrationsarbschv/gesamt.pdf](http://www.gesetze-im-internet.de/bundesrecht/l_rmvibrationsarbschv/gesamt.pdf), Zuletzt geprüft 16.Dezember 2015.
- McNeill, D. (1992): Hand and mind. What gestures reveal about thought. University of Chicago Press, Chicago.
- Nowack, T., Jehring, S., Kurtz, P. (2015): Zeigen – Objektauswahl für ein industrielles Assistenzsystem zum Heben und Tragen von Kleinladungsträgern; in Wienrich, C., Zander, T. O., & Gramann, K.. Trends in Neuroergonomics. 11th Berlin Workshop of Human-Machine Systems, Proceedings, 7th - 9th October, 2015. Berlin: Universitätsverlag der TU Berlin. ISBN-Nr.: 978-3-7983-2803-7 (online).
- Nowack, T.; Suzaly, N.; Lutherdt, S.; Schürger, K.; Jehring, S.; Witte, H.; Kurtz, P. (2015): Phases of technical gesture recognition. In: Kurosu (Hg.)– Human-computer interaction, S. 130–139.
- Pavlovic, V. I., Sharma, R., Huang, T. S. (1997): Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. In: IEEE Trans. Pattern Anal. Mach. Intell. 19 (7), pp. 677–695. Washington.
- Strassmann, B. (2010): Gehorsam auf Fingerzeig. Die Zeit Nr. 10/2010.  
<http://www.zeit.de/2010/10/Touchpads>, Zuletzt geprüft 16.Dezember 2015.
- Walter, R., Bailly, G., Müller, J. (2013): StrikeAPose: Revealing Mid-air Gestures on Public Dis-plays. In: Wendy E. Mackay und A. Special Interest Group on Computer-Human Interaction (Hg.): Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. [S. I.]: ACM, pp. 841–850.
- Wigdor, D.; Wixon, D. (2011): Brave NUI world: Designing natural user interfaces for touch and gesture, Morgan Kaufmann/Elsevier, Burlington and MA.