

**Tobias Falk Nowack**

**Mensch-Technik-Interaktion mittels  
Freiraumgesten**

# Berichte aus der Biomechatronik

Herausgegeben von Prof. Dr. Hartmut Witte  
Fachgebiet Biomechatronik an der TU Ilmenau

Band 15

# **Mensch-Technik-Interaktion mittels Freiraumgesten**

Tobias Falk Nowack



Universitätsverlag Ilmenau  
2017

# Impressum

## **Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Angaben sind im Internet über <http://dnb.d-nb.de> abrufbar.

Diese Arbeit hat der Fakultät für Maschinenbau der Technischen Universität Ilmenau als Dissertation vorgelegen.

Tag der Einreichung: 21. Oktober 2016

1. Gutachter: Univ.-Prof. Dr.-Ing. habil. Prof. h. c. Peter Kurtz  
(Technische Universität Ilmenau, FG Arbeitswissenschaft)
2. Gutachter: Univ.-Prof. Dipl.-Ing. Dr. med. (habil.) Hartmut Witte  
(Technische Universität Ilmenau, FG Biomechatronik)
3. Gutachter: Prof. Dr.-Ing. Matthias Rötting  
(Technische Universität Berlin)

Tag der Verteidigung: 19. Juni 2017

Technische Universität Ilmenau/Universitätsbibliothek

### **Universitätsverlag Ilmenau**

Postfach 10 05 65

98684 Ilmenau

[www.tu-ilmenau.de/universitaetsverlag](http://www.tu-ilmenau.de/universitaetsverlag)

readbox unipress

in der readbox publishing GmbH

Am Hawerkamp 31

48155 Münster

<http://unipress.readbox.net>

ISSN 1865-9136 (Druckausgabe)

ISBN 978-3-86360-166-9 (Druckausgabe)

URN urn:nbn:de:gbv:ilm1-2017000379

---

Titelfoto: Dipl.-Biol. Helga Schulze | Bochum

# DANKE

Der größte Dank für die letzten Tage und Wochen gilt meiner Familie. Sie hat mir mit nahezu allem den Rücken freigehalten, so dass ich mich vollständig auf diese Arbeit konzentrieren konnte. Sie hat es aber auch verstanden, mir den Kopf zu waschen, wenn ich wieder mal nicht wusste, wie ich alles packen sollte und über den eigenen Anspruch gestolpert bin.

Darüber hinaus gilt der Dank meinem Doktorvater Univ.-Prof. Dr.-Ing. habil. Prof. h. c. Peter Kurtz für die Möglichkeit, in den letzten 10 Jahren an diesem und ähnlichen Themen forschen und dazu veröffentlichen zu dürfen.

Weiterhin möchte ich nicht Univ.-Prof. Dipl.-Ing. Dr. med. (habil.) Hartmut Witte vergessen, Danke zu sagen. Er hatte besonders den letzten Tagen der Fertigstellung durch seine Korrekturhinweise entscheidend Einfluss auf die nun vorliegende Endfassung.

Dank an alle meine Kollegen von den Fachgebieten Arbeitswissenschaft und Biomechatronik, besonders erwähnt seien Frau M.Sc. Sabine Wenzel und Herr Dipl.-Ing. Stefan Lutherdt, die in den letzten Wochen und Monaten etliche meiner Aufgaben übernommen oder über die langsame Bearbeitung anderer Aufgaben hinweggesehen haben. Frau Dipl.-Biol. Danja Voges möchte ich im Besonderen erwähnen, da ich ihrer Kreativität etliche Abbildungen der Arbeit verdanke.

Meinen Studenten, B.Sc. Stefan Jehring, M.Sc. Kisten Schürger und M.Sc. Yue Xiong und noch vielen anderen, die mit ihren Qualifikationsarbeiten und

den Diskussionen zu ihren Arbeiten ihren Beitrag an der Entwicklung der Gestendefinitionen hatten.

Und zu guter letzt meinen Eltern, Schwiegereltern und Freunden: Danke dass ihr an einen erfolgreichen Abschluss geglaubt und mich ermutigt oder eure Hilfe angeboten habt. Ich weiß, ich war nicht einfach und hatte kaum noch was anderes im Kopf, aber ich verspreche, das soll wieder anders werden. Und an meinen Vater: „**Ich schaff es doch!**“

# Zusammenfassung

Mit Industrie 4.0 gewinnt die Mensch-Technik-Interaktion deutlich an Stellenwert. Sollen Mensch und Maschine gemeinsam arbeiten, dann sind alternative Interaktionsformen zu entwickeln. Erfolgreiche Kommunikation bedeutet, dass beide Kommunikationspartner zum Kommunikationsende den gleichen Informationsstand haben. Die natürliche, menschliche Kommunikation findet auf verschiedenen Ebenen (z.B. visuell und akustisch) statt, welche zusammengehörig gedeutet werden. An der redebegleitenden Zeigegeste lässt sich das Zusammenspiel von gesprochenem Wort, ausgeführter Zeigegeste und kognitiver Leistung des Empfängers gut erkennen. Der „grüne Ordner da drüben“ fordert vom Empfänger neben der rein sensorischen Aufgabe des Hörens und Sehens die zusätzliche, wissensbasierte Verarbeitung von „grün“ und „Ordner“. Mit all diesen Informationen erreicht die gestenunterstützte Mensch-Mensch-Kommunikation eine Erkennungsrate von annähernd 100 %.

Sollen für die Mensch-Technik-Kommunikation „Zeigegesten“ verwendet werden, sind die zusätzlichen Kontext-Informationen z.Z. für die technische Auswertung noch nicht verfügbar. Daher wurden Beschreibungen der Zeigegeste entwickelt und evaluiert, mit denen bereits ohne Interpretation der Kontextinformation eine Gesamterkennungsrate von über 85 % realisiert werden kann.

Ausgehend von dem Nutzerverständnis - Zeigen wird als statische Geste verstanden - wurde eine technisch nutzbare Beschreibung entwickelt. Diese, nur die Körperpose des Zeigens auswertend, erreicht bei der Evaluation eine Erkennungsrate von ca. 50 %. Sie ließ sich bereits im Rahmen eines Forschungsprojektes erfolgreich in die Praxis übertragen. Auf Grundlage eines Phasenmodells der Gestenausführung wurden zwei weitere Beschreibungen entwickelt. Bei diesen wird der gesamte Gestenzyklus aus Einleitung, Hauptteil, Haltephase und Ausleitung betrachtet.

Mit den Daten der für die Evaluation eingesetzten Kinect® 2 konnten mit den erweiterten Phasendefinitionen Erkennungsraten von über 85 % realisiert werden.

Für eine Verbesserung der Erkennungsrate ist es zukünftig notwendig, weitere Gesten wie das „Heranwinken“ zu beschreiben, um sie vom „Zeigen“ abzugrenzen. Zusätzlich muss das Zeigeziel erkannt und anhand der Kontextinformation überprüft werden können.

# Abstract

Human-Machine-Interaction will be significant for future development in Human-Robot-Collaboration. Humans and robots should interact like colleagues. In any case the human has to control the robot. To realize interaction procedures, well known from interhuman communications like pointing, humans use visual and auditory sense as well as their cognitive skills. Human pointing may be supported by speech like „the green folder over there“. Humans know a „folder“ and can evaluate, if the folder is „green“. The words „over there“ let them know that the next movement, the next body posture will be a pointing gesture. With all these skills the detection rate will be nearly 100 %.

Using the pointing gesture for technically controlling a robot, this additional knowledge is not available yet. In this thesis three algorithms to detect pointing with a simple visual sensor like Kinect<sup>®</sup> 2 have been developed. Even without the additional knowledge the detection rate is up to 85 %. The test persons of an acceptance survey attest that pointing is looked at as a static body posture. The first developed algorithm detects pointing with a defined (static) body posture. A detection rate of about 50 % could be reached. The static algorithm has been implemented for a lifting and carrying robot during an industrial research project and works very well. Performing a gesture is a complete (dynamic) movement cycle. For the two other algorithms all phases, introduction, stroke, post-stroke hold and release have to be performed.

With the data of 15 test persons (302 gestures) a detection rate of 85 % could be reached by using those phase-definitions. To increase the detection rate it will be necessary to add other gesture detection algorithms (e.g. to detect waving). This could improve the decision to reject a pointing gesture. Detecting the pointing destination and to evaluate whether it is equivalent to the searched object will increase the accuracy additionally.



# Inhaltsverzeichnis

<b>Inhaltsverzeichnis</b>	<b>11</b>
<b>Abkürzungen</b>	<b>13</b>
<b>Formelzeichen</b>	<b>15</b>
<b>1 Einleitung</b>	<b>17</b>
<b>2 Propädeutik</b>	<b>21</b>
2.1 Die Geste . . . . .	22
2.2 Kommunikation . . . . .	23
2.3 Handlungsmodell . . . . .	29
2.4 Kognitive Leistungsfähigkeit des Menschen . . . . .	30
2.5 Kontext-Information . . . . .	36
2.6 Technisches System . . . . .	39
2.7 Akzeptanz . . . . .	45
2.8 Zusammenfassung . . . . .	52
<b>3 Gestenbeschreibung - Formen und Phasenmodell</b>	<b>55</b>
3.1 Nonverbale Ausdrucksformen, Sprachen . . . . .	56
3.2 Klassifizierung von Gesten . . . . .	61
3.3 Zeigegeste . . . . .	66
3.4 Gestenraum . . . . .	67
3.5 Phasenmodelle der Geste . . . . .	71
3.6 Nutzerverständnis . . . . .	73

<b>4</b>	<b>Parametrisierung der Zeigegeste</b>	<b>77</b>
4.1	Statische Definition . . . . .	78
4.2	Dynamische Beschreibung mittels linearer Regression . . . . .	83
4.3	Phasendefinition mittels zweier Winkel . . . . .	87
4.4	Phasendefinition mittels eines Winkels . . . . .	93
<b>5</b>	<b>Experiment</b>	<b>99</b>
5.1	Versuchsdesign . . . . .	99
5.2	Versuchsdurchführung . . . . .	101
5.3	Versuchsauswertung . . . . .	102
5.4	Kontrollexperiment zum Einfluss einer Drehung um die Körperlängsachse . . . . .	111
<b>6</b>	<b>Diskussion</b>	<b>117</b>
6.1	Phasendefinition . . . . .	117
6.2	Versuchsdesign . . . . .	119
6.3	Sonderfälle . . . . .	121
6.4	Winkelbetrachtung . . . . .	127
6.5	Fehlerdiskussion . . . . .	131
6.6	Zusammenführung von Online- / Offline-Version . . . . .	132
<b>7</b>	<b>Zusammenfassung</b>	<b>143</b>
<b>8</b>	<b>Ausblick</b>	<b>147</b>
8.1	Weiterentwicklung der Gestenbeschreibung . . . . .	147
8.2	Übertragung in die Praxis . . . . .	149
<b>A</b>	<b>Vollständige Optimierung</b>	<b>153</b>
	<b>Begriffsdefinitionen</b>	<b>159</b>
	<b>Abbildungsverzeichnis</b>	<b>163</b>
	<b>Tabellenverzeichnis</b>	<b>167</b>
	<b>Formelverzeichnis</b>	<b>170</b>
	<b>Literaturverzeichnis</b>	<b>171</b>
	<b>Verzeichnis der verwendeten Normen</b>	<b>177</b>

# Abkürzungen

<b>DGS</b>	Deutsche Gebärdensprache
<b>DMA</b>	Double Moving Averaging Filter
<b>fps</b>	frames per second (dt. Bilder pro Sekunde)
<b>MS</b>	Microsoft®
<b>ROI</b>	Region of Interest (dt. Bereich von Interesse)



# Formelzeichen

## Bestimmung des Gestenraumes

$\phi_{Gzm}$  minimaler Schulterwinkel des Menschen für den zum Zeigen genutzten Raum (männlich)

$\phi_{Gzw}$  minimaler Schulterwinkel des Menschen für den zum Zeigen genutzten Raum (weiblich)

$SH_m$  durchschnittliche Schulterhöhe (männlich)

$SH_w$  durchschnittliche Schulterhöhe (weiblich)

$HH_m$  durchschnittliche Höhe der Hand (Griffachse) über der Standfläche (männlich)

$HH_w$  durchschnittliche Höhe der Hand (Griffachse) über der Standfläche (weiblich)

$B_m$  durchschnittliche Gesäß-Beinlänge (männlich)

$B_w$  durchschnittliche Gesäß-Beinlänge (weiblich)

**Punkte** nach Definition Kinect® (siehe Abb. 2.14 auf Seite 44)

**Eg** Ellbogengelenkpunkt

**H** Hüftgelenkpunkt

**Hg** Handgelenkpunkt

**Sg** Schultergelenkpunkt

<b>WsB</b>	Punkt Wirbelsäule Basis
<b>WsS</b>	Punkt Wirbelsäule Schulter

**Vektoren**

$\overline{HS}$	Vektor Hand-Schulter, Hg $\rightarrow$ Sg
$\overline{T}$	Vektor Torso, WsB $\rightarrow$ WsS
$\overline{U}$	Vektor Unterarm, Hg $\rightarrow$ Eg

**Winkel**

$\Phi_{Eg}$	Ellbogenwinkel
$\Phi_{Sg}$	Schulterwinkel
$\Phi_{(\overline{HS}, \overline{T})}$	Winkel zwischen den Vektoren $\overline{HS}$ und $\overline{T}$
$\Phi_{(\overline{U}, \overline{T})}$	Winkel zwischen den Vektoren $\overline{U}$ und $\overline{T}$

**Winkelgeschwindigkeiten**

$\dot{\Phi}_{(\overline{HS}, \overline{T})}$	Winkelgeschwindigkeit zwischen den Vektoren $\overline{HS}$ und $\overline{T}$
$\dot{\Phi}_{(\overline{U}, \overline{T})}$	Winkelgeschwindigkeit zwischen den Vektoren $\overline{U}$ und $\overline{T}$

**Sonstige**

<b>Er</b>	Erkennungsrate
$Er_z$	Erkennungsrate „Zeigen“ <i>korrekt erkannt</i>
$Er_{nz}$	Erkennungsrate „nicht Zeigen“ <i>korrekt abgelehnt</i>
<b>f</b>	frame
<b>i</b>	Laufvariable im Zeitfenster
<b>px</b>	Pixel
$R^2$	Bestimmtheitsmaß
<b>t</b>	Zeit (time)
$t_H$	Äquivalent der notwendigen Haltezeit

# Kapitel 1

## Einleitung

Die für Industrie 4.0 notwendige Automatisierungstechnik soll durch Selbstkonfiguration intelligenter werden und dem Menschen bei zunehmend komplexer Arbeit in Selbstoptimierung assistieren und ihn so entlasten. Prozessen und Strukturen wird dabei mehr Wandlungsfähigkeit abverlangt. Wenn aber Regelprozesse in Chaos übergehen oder aus technischer, organisatorischer oder ökonomischer Sicht die Selbstfindung ausbleibt, dann wird es Szenarien geben, in den der Mensch dank seiner komplexen Wahrnehmung und Kognition unverändert gebraucht wird.

Freiraumgesten und im Besonderen das räumliche Zeigen, bieten sich für die Mensch-Technik-Interaktion im industriellen Umfeld der Industrie 4.0 an. In Bereichen der Logistik oder der Produktionsassistenz sind sie für den Nutzer eine natürliche und seit der Kleinkindzeit bekannte und eingeübte Interaktionsform und sind in der Mensch-Mensch-Kommunikation integriert. In der zwischenmenschlichen Kommunikation werden Gesten meist sprachbegleitend eingesetzt. Eine Zeigegeste unterstützt den Empfänger bei der räumlichen oder zeitlichen Verortung eines Objektes im Sinnkontext. Warum bietet sich diese Gestenform für die Mensch-Technik-Interaktion an? Weil Sie die räumliche Verortung von Objekten eindeutiger beschreibt, als der Mensch es sprachlich kann.

Exemplarisch soll mit dieser Arbeit die „Technische Erkennung der Zeigegeste“ beim gegenwärtigen Stand der Technik erforscht werden. Gleichsam soll an diesem Beispiel erläutert werden, welche Restriktionen die scheinbar einfache „Anweisung durch Zeigen“ in der Mensch-Technik-Interaktion erschweren.

Betrachtet wird diese spezielle Funktion der Zeigegeste im Kontext einer einfachen lagerlogistischen Aufgabenstellung. Einem Unterstützungssystem zum Heben und Tragen von Lasten sollen durch den Nutzer die Lagerorte angezeigt werden. Die Lagerlogistik mit ausschließlich automatischen Transportsystemen „kennt“ in der „Chaotischen Lagerhaltung“ den Lagerplatz eines Artikels ausschließlich anhand der Information im Warenwirtschaftssystem. Sind allerdings Menschen mit ihren besonderen Fähigkeiten des Beurteilens und Entscheidens (siehe dazu [VDI/VDE 3850-1, 2000]) im Lagerbereich tätig und werden durch das technische System unterstützt, dann ist die Lagerhaltung mit festen Lagerplätzen und einer durch den Menschen zu überblickende Lagerordnung vorzuziehen.

In solchen Lagerumgebungen ist der arbeitende Mensch zwei schwerwiegenden Belastungsgrößen ausgesetzt. Zum einen sind bereits zu bewegende Lasten mit Gewichten von 50 kg für den menschlichen Bewegungsapparat potentiell schädlich (und Lasten über 15 kg für Mitarbeiterinnen verboten, für Schwangere liegt die Obergrenze sogar nur bei 5 kg), zum anderen sind viele Lagerplätze nicht auf kurzen, ungefährlichen Wegen oder sogar nur mit Hilfsmitteln (z.B. Leiter) zu erreichen. Um derartige Lagerplätze aus einer ergonomisch günstigen Körperhaltung des aufrechten Stehens zu adressieren, bietet sich die Zeigegeste an.

Zeigen lernt der Mensch im Kindesalter als erste Geste und nutzt das Zeigen fortan intuitiv. „Eine einfache Zeigegeste zeigt eine Richtung an oder lokalisiert eine bestimmte Entität<sup>1</sup>“ [Fricke, 2007, S. 59]. Zeigen ist demnach das Deuten, Verweisen auf ein Objekt mit der Hand oder dem Finger. Die Ausführung dieser Geste kann individuell verschieden sein, und doch gibt es für alle Ausführungen grundlegende Gemeinsamkeiten, die es menschlichen Kommunikationspartnern ermöglichen, die intuitiv ausgeführte Geste auch eindeutig zu verstehen.

Unter den verfügbaren Sensoren ist ein System auszuwählen, das den Ausführenden erkennt und die von ihm individuell ausgeführte Zeigegeste deuten kann. Dafür bieten sich vor allem die bekannten, kostengünstigen, „markerlosen“, „optischen“ Systeme aus der Spielewelt, die Kinect® -Sensoren (incl. vergleichbarer Sensorsysteme z.B. von ASUS®) an. Ihre grundsätzliche Tauglichkeit für die „zuverlässige“ Erkennung der Zeigegeste muss evaluiert werden. Auf Basis aufgenommener und erkannter „Skelettpunkte“ werden verschiedene Ansätze zur Erkennung des „Zeigens“ entwickelt und in Probandenversuchen evaluiert.

---

<sup>1</sup> einen konkreten oder abstrakten Gegenstand

In dieser Arbeit sind für die technisch zuverlässige Erkennung der Freiraumgeste „Zeigen“ Algorithmen zu entwickeln und zu testen. Es sind Voraussetzungen und Bedingungen zu untersuchen, wie mit einem „Zeigen“ auf Objekte die eindeutige zuverlässige Erkennung im Praxiseinsatz funktionieren kann.

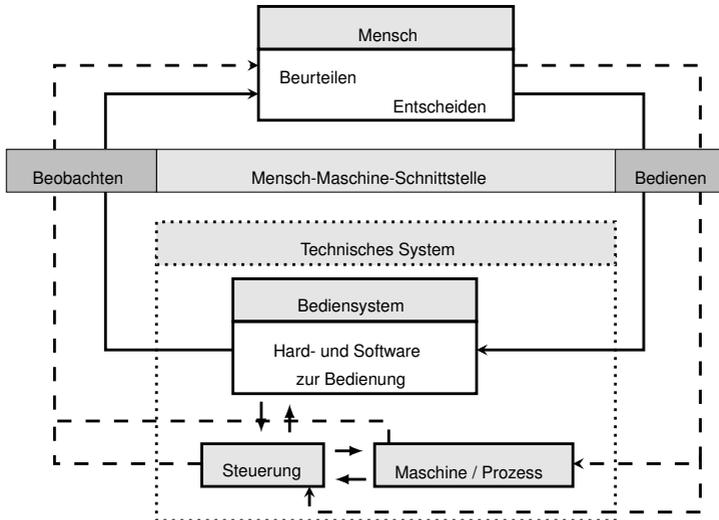


## Kapitel 2

# Propädeutik

Dieses Kapitel erläutert die grundlegenden Begriffe und einen Teil der Modelle, die für das weitere Verständnis notwendig sind. Für die beiden Interaktionspartner Mensch und Maschine (technische System) sind deren besonderen Fähigkeiten ausschlaggebend für die Systemkonzeption. Die VDI/VDE 3850-1 [2000] bildet diesen Zusammenhang mit dem Übersichtsmodell zur Mensch-Maschine-Schnittstelle (siehe Abb. 2.1 auf der nächsten Seite) ab.

„Der Mensch kann [...] insbesondere die Kontrolle in Situationen übernehmen, in denen Problemlösungsfähigkeit, Kreativität, Flexibilität und Bewertungen erforderlich sind“ [Herczeg, 2009, S. 25]. Auf der anderen Seite ist das „Technische System“, das Computer-Komponenten für die Interaktion, aber auch für die eigentliche Steuerung, enthält. Bei Herczeg sind die Aussagen zu den besonderen Fähigkeiten des „technischen Systems“ immer auf den Computer als technisches System bezogen: „Computer können somit insbesondere die Kontrolle in Situationen übernehmen, in denen umfangreiche, gut definierte, schnelle und systematische Analysen und Reaktionen erforderlich sind“ [Herczeg, 2009, S. 25]. Diese Aussage stellt aber die Situation für das „technische System“ bereits gut dar. Was die Maschine im „technischen System“ zudem auszeichnet, ist Ausdauer und Kraft bei gleichbleibender Genauigkeit (vergl. Kap. 2.6 auf Seite 39).



**Abbildung 2.1: Geltungsbereich der Richtlinie VDI/VDE 3850**

Darstellung nach [VDI/VDE 3850-1, 2000]

## 2.1 Die Geste

Der Titel der Arbeit spricht von „Interaktion mittels Freiraumgesten“. Bevor alle notwendigen Komponenten der Mensch-Technik-Interaktion vorgestellt und dann im Kapitel 3 auf Seite 55 auf die Geste im Detail eingegangen wird, soll der Begriff der Geste, im Besonderen der Zeigegeste im Kontext dieser Arbeit kurz vorgestellt werden.

„Gesten verraten viel über einen Menschen“ [Argyle und Trower, 1981, S. 19]. „Ein Sprichwort sagt, Neapolitaner könnten mit gefesselten Händen nicht sprechen.“ Also ist die Geste ein essentieller Bestandteil unserer menschlichen Kommunikation. Im Brockhaus [1988, S. 437] wird die „Geste [lat. gestus >Gebärdenspiel<] [als] zielgerichtete Ausdrucksbewegung des Körpers, bes. der Hände und des Kopfes: im Unterschied zur Ausdrucksbewegung des Gesichts (→ Mimik)“ definiert.

„Daß die hinweisende Gebärde [...] nicht nur die einfachste, sondern auch die ursprünglichste ist,“ schreibt Wundt [1900, S. 157]. Die „hinweisende Gebärde“, wie er sie beschreibt, kann mit der Zeigegeste gleichgesetzt werden. Wie später auch andere Autoren legt Wundt das hauptsächliche Augenmerk auf eine, die Aufmerksamkeit steuernde Wirkung der Geste, beschreibt aber

auch deren Informationsgehalt: „Wo der Gegenstand, auf den sich irgendeine Gebärdenmitteilung bezieht, im Sehbereich liegt, da ist die unmittelbare Richtung des Zeigefingers gegen ihn das einfachste, weil das sicherste und eindeutigste Mittel, die Aufmerksamkeit auf ihn zu lenken, ein Mittel, das in der Regel ohne Überlegung, aus dem unmittelbaren Trieb nach Mitteilung heraus angewandt wird“ [Wundt, 1900, S. 157]. Also ist die Zeigegeste schon bei Wundt eine intuitiv durchgeführte und intuitiv verstandene Geste.

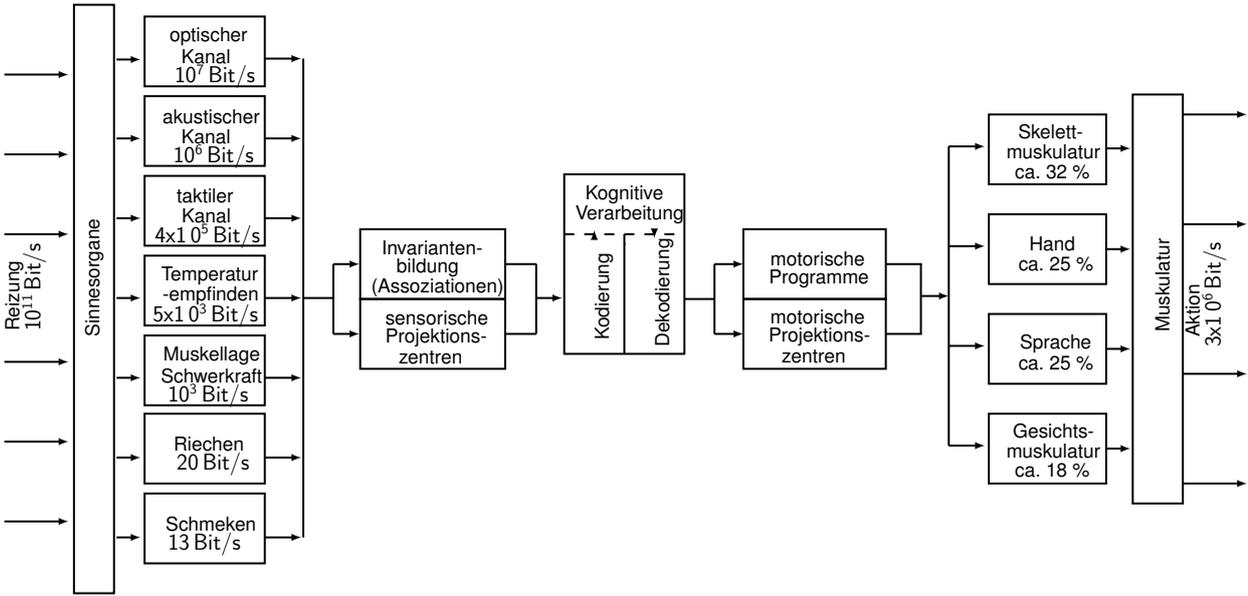
Wie Gesten allgemein und die Zeigegesten im Detail zu beschreiben sind, wird in Kapitel 3 auf Seite 55ff. vorgestellt. Wie diese Beschreibung technisch zu interpretieren ist, wird im Kapitel 4 auf Seite 77ff. behandelt. Die im Folgenden vorgestellten Themen werden prinzipiell im Hinblick auf die hinweisende Funktion der Zeigegeste analysiert und dargestellt.

## 2.2 Kommunikation

Kommunikation ist ein Grundbestandteil des menschlichem Zusammenlebens und aktuell ein wichtiges Standbein der technischen Umsetzung vernetzter Prozesse. Ein weiteres Feld der Kommunikation ist die Mensch-Technik-Interaktion (siehe Abb. 2.1 auf der vorherigen Seite), denn auch diese besteht aus dem Austausch von Informationen zwischen zwei Kommunikationspartnern, wie sie grundlegend im Modell von Shannon und Weaver (siehe Abb. 2.3 auf Seite 25) erläutert wird.

### Kommunikationsmodell

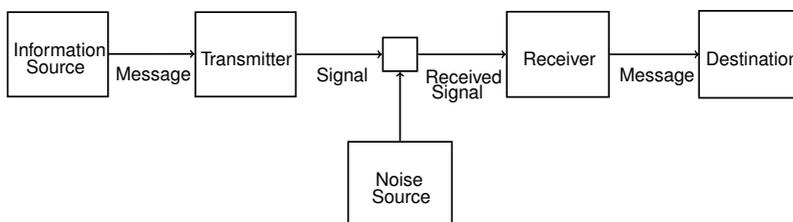
Die zwischenmenschliche Kommunikation basiert auf einer komplexen Wahrnehmung und Kognition. Sie wird auf mehreren Informationskanälen geführt. So ist die Sprache und damit der akustische Kanal üblicherweise für die menschliche Kommunikation gebräuchlich, das visuelle Sensorium des Menschen ist mit der höchsten Kanalkapazität ausgelegt (siehe Abb. 2.2 auf der nächsten Seite). Beide Informationskanäle werden bei der von Gesten unterstützten Kommunikation gemeinsam aktiv. Da die sprachbegleitende Geste den überwiegenden Inhalt über den akustischen Kanal überträgt, bleiben auf dem optischen Kanal Reserven, die der Mensch bewusst oder auch unbewusst nutzt, um zusätzliche Inhalte aufzunehmen (siehe Abb. 2.10 auf Seite 36). Diese Fähigkeit bleibt dem technischen System in der Kommunikation leider z.Z. noch versagt.



**Abbildung 2.2: Vereinfachtes Modell der Informationsaufnahme**

(Kanalkapazität der Menschlichen Sinnesorgane)

Darstellung modifiziert nach [Beelich und Schwede, 1991]



**Abbildung 2.3: Kommunikationsmodell von Shannon**

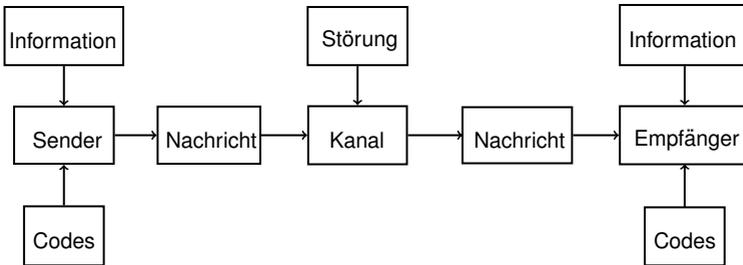
Darstellung nach [Shannon und Weaver, 1949]

Für die Kommunikation muss neben der Kanalkapazität der menschlichen Informationsverarbeitung bzw. der menschlichen Informationsaufnahme das Übertragungsverhalten von Informationen betrachtet werden. Hierzu bietet sich das Modell nach Shannon und Weaver [1949] (siehe Abb. 2.3) an, da es der „Mathematischen Theorie der Kommunikation“ entstammt. „Die Quelle wählt eine bestimmte Nachricht aus einem Satz möglicher Nachrichten (dieser Zusammenhang ist teilweise sehr wichtig [...]). Die ausgewählte Nachricht können Texte, Sprache, Bilder oder Musik sein. [...] Wenn ich mit dir spreche ist mein Gehirn die Quelle und deines die Senke, mein Sprachsystem der Sender und dein Ohr mit dem dazugehörigen Hörnerv der Empfänger“<sup>1</sup>.

Kommunikation besteht also aus dem Inhalt, der „Information“, deren Kodierung, der Übertragung als „Nachricht“, und dem Umkehrprozess des Dekodierens beim Empfänger. Die möglichen Störung wird von Shannon als Bestandteil der Kommunikation gesehen, die den Prozess der Übertragung charakterisiert. Für seine mathematische Theorie macht diese Betrachtung auch Sinn, da neben der verbalen Beschreibung der Information auch Information im Kontext seiner Arbeit unabhängig vom Inhalt betrachtet wird. „In der Tat sind zwei Nachrichten, von denen eine mit viel Bedeutung und die andere reiner Blödsinn ist, aus heutiger Sicht der Information betrachtet, exakt vergleichbar“<sup>2</sup>. Shannon betrachtet die Nachricht aus nachrichtentheoretischer Sicht und somit sind Nachrichten eine Anzahl aufeinanderfolgender Zeichen, die, wenn sie vergleichbar kodiert sind, den gleichen Informationsgehalt haben.

<sup>1</sup> Frei übersetzt nach [Shannon und Weaver, 1949, S. 7]

<sup>2</sup> Frei übersetzt nach [Shannon und Weaver, 1949, S. 8]



**Abbildung 2.4: Kommunikationsmodell nach Shannon und Weaver  
(Interpretation von Herczeg)**

Darstellung nach [Herczeg, 2009]

Diese Betrachtungsweise kann aber für die menschliche oder die Mensch-Technik Kommunikation nicht übernommen werden. Für den Menschen sind Inhalte der essentielle Bestandteil der Nachricht, und diese sollen so ungestört wie möglich übertragen werden. Die Modellinterpretation von Herczeg [2009] als Kommunikationsmodell der einfachen Mensch-Computer-Kommunikation bildet diese erweiterte Betrachtung sehr gut ab (siehe Abb. 2.4). Störungen auf dem analogen Kommunikationskanal, wie ihn Shannon noch kannte, werden auf modernen, digitale Daten übertragenden Kommunikationskanälen durch angepasste Codes, der Verwendung von Prüfsummen und weiterer Kontrollmechanismen möglichst gering gehalten. Herczeg erläutert zu seinem Modell, dass „je besser die verwendeten Codes übereinstimmen und je geringer der Kanal gestört ist, desto fehlerfreier wird die Nachricht übertragen. Gegenüber dem Modell von Shannon und Weaver werden hier ‚Nachrichten‘ als in qualitativ-bedeutungstragende Zeichen kodierte Informationen, bei der Übertragung betrachtet. Signale sind technisch-quantitative Ausprägungen der Nachrichten“ [Herczeg, 2009, S. 117].

Wenn man das Kommunikationsmodell von Shannon und Weaver in der Interpretation von Herczeg (siehe Abb. 2.4) auf die menschliche Kommunikation überträgt, sind die Schwachstellen wieder immanent. Besonders kritisch sind in diesem Fall

- die Kodierung (z.B. Sender spricht Englisch und Empfänger versteht nur Deutsch)
- sowie die klassische Störung auf dem eigentliche Kommunikationskanal (z.B. störender Industrielärm)

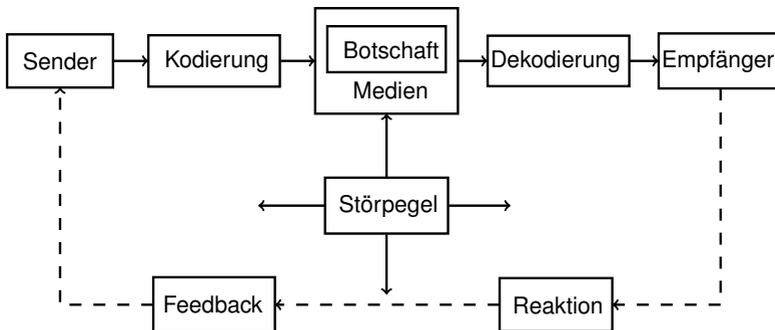
zu sehen. Wegen dieser möglichen Kommunikationsstörungen verwendet der Mensch zum Übertragen einer Nachricht in für ihn unklaren Kommunikationssituationen meist mehrere Kanäle der Informationsübertragung. Es werden vor allem „nonverbale Signale eingesetzt: Deuten, Gesten aller Art, Hilfestellungen, mimische Kommentare. In lärmenden Fabriken und in Aufnahmestudios, in denen man sich nicht durch Worte verständigen kann, nehmen sie teilweise den Platz der Sprache ein. Signale dieser Art sind jedoch nicht mit der Körpersprache identisch“ [Argyle und Trower, 1981, S. 32].

Daher wird das Kommunikationsmodell von Shannon und Weaver für die Mensch-Mensch-Kommunikation kontrovers diskutiert. So erläutert Bernsau [2003], dass für die zwischenmenschliche Kommunikation das informationstechnisch geprägte Modell ungeeignet sei. Basierend auf Krallmann und Ziemann sind: „Zwischenmenschliche Kommunikationsprozesse [...] keine Informationsübertragung zwischen Sender und Empfänger!“. Weiter stellt er aber auch heraus, dass „Auf der anderen Seite [...] im Laufe der Zeit der technische Begriff der Information gerade in der geistes- und sozialwissenschaftlichen Forschung und Lehre allgemein akzeptiert“ wird. „Und ich [Bernsau] muß zugeben, daß die Vorstellung, zwischenmenschliche Kommunikation erfolge linear zwischen Sender und Empfänger und Information sei objektiv existent, verführerisch eingängig ist“ [Bernsau, 2003]. Zum Schluss kommt er mit der Aussage, dass „sich unter dem Stichwort Kommunikationsmodelle bis auf den Saussureschen Kreislauf nur Modelle, die mehr oder weniger auf Shannon & Weaver basieren“, finden lassen.

Eine die Besonderheit der Mensch-Mensch-Kommunikation ist allerdings, dass die menschliche Kommunikation nicht als lineares Modell eine Kommunikationsbeziehung

Sender ⇒ Kanal ⇒ Empfänger

gesehen werden kann, sondern sich, wie von Kotler [2003] beschrieben, ein implizites Kreislaufmodell ergibt (siehe Abb. 2.5 auf der nächsten Seite). In seiner Erweiterung des Modells auf den Kommunikationsprozess der Marketingkommunikation unterscheidet er zwischen Wirkung, als „Reaktion der Empfänger auf die Botschaft“ und Feedback, „der Teil der Wirkung, der an den Sender zurück übermittelt wird.“ Die in dem Modell beschriebenen Rückkoppelschleifen können verbal und non-verbal ausgeführt werden.



**Abbildung 2.5: Kommunikationsmodell (Interpretation von Kotler)**

Darstellung nach [Kotler, 2003]

## Nonverbale Kommunikation

Der Mensch setzt intuitiv Gesten ein, um Kommunikationsstörungen zu umgehen. Wenn der Mensch sich sprachlich mit seinem Gegenüber nicht verständigen kann, so spricht er „mit Händen und Füßen“, er gestikuliert. Vergleichbar verhält er sich, wenn er sicher ist, dass der akustische Kanal nicht genutzt werden kann, z.B. der Einweiser eines LKW oder Flugzeuges.

Die menschliche Fähigkeit, in der Kommunikation alle Sinneskanäle für die Informationsaufnahme zu verwenden kann neben vielen Vorteilen aber auch große Nachteile aufweisen. Die Körpersprache sendet teilweise auch widersprüchliche Signale. Argyle und Trower [1981, S. 19] schreiben dazu bezogen auf die Gesten: „Sie lassen erkennen, ob er [der Mensch] wütend, ängstlich oder unsicher ist, ob er lügt oder sich freut. Gesten können Worte Lügen strafen oder ihnen Nachdruck verleihen.“ Auch hier wird auf eine, die sprachliche Kommunikation unterstützende Funktion der Geste verwiesen.

„Nehmen wir an, Marion möchte, daß Peter sich mehr um die Kinder kümmert. Belohnt Marion jeden Schritt, den er in diese Richtung unternimmt (durch Lächeln, Kopfnicken, Zärtlichkeiten, Blicke und Zustimmung), wird Peter sehr bald tun, was Marion wünscht. Aber Stirnrunzeln, Wegblicken oder andere Zeichen der Mißbilligung werden seine Bemühungen schnell wieder erlahmen lassen“ [Argyle und Trower, 1981, S. 36 f.].

Dieses Beispiel untersetzt das zwischenmenschliche Kommunikationsdilemma, das geprägt ist durch die Möglichkeit der Wahrnehmung von Information auf verschiedenen Kanälen. Vor allem ist der Mensch in seinem alltäglichen

Verhalten nicht trainiert, dieses Kommunikationsdilemma zu beachten, weil er intuitiv handelt.

Bestimmte Berufsgruppen z.B. Verkäufer sind trainiert, diese als „non-verbale-Kommunikation“ bezeichneten Zeichen und Verhaltensmuster gezielt einzusetzen und beim Gesprächspartner auch zu deuten. „Worte allein ergeben manchmal noch keinen Sinn. Deshalb benutzen wir nonverbale Methoden, um unsere Botschaften zu übermitteln. Wenn Worte versagen, helfen Gesten oft weiter“ [Argyle und Trower, 1981, S. 65].

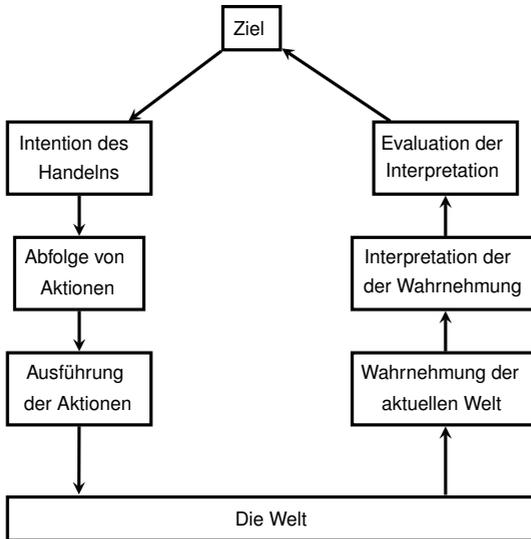
Und so wird die Zeigegeste auch gezielt eingesetzt, um redebegleitend zusätzliche Information zu übertragen und die Aussage zu präzisieren.

## 2.3 Handlungsmodell

Zum Verständnis des menschlichen Handelns haben verschiedene Autoren Handlungsmodelle entwickelt. Diese versuchen die Beziehung zwischen der Intention des Handelns und der Reaktion der manipulierten Welt zu beschreiben. Dazu werden Modelle verwendet, die wie z.B. durch die „sieben Stadien des Handelns“ von Norman [1988] beschrieben, einen Regelkreis abbilden (siehe Abb. 2.6 auf der nächsten Seite). Dieses Modell teilt das Handeln des Menschen in zwei unterschiedliche Aufgaben. Zum einen die Aufgabe der Wahrnehmung und Interpretation des Zustandes der Welt, und zum anderen die Intention und Handlungsplanung um die Welt zu verändern. Diese zwei Aspekte bestimmen jede Form einer menschlichen Handlung.

Beziehen wir dieses Modell jetzt auf die Mensch-Technik-Interaktion, so wird dieses Modell theoretisch zweimal durchlaufen. Der Mensch will mit dem technischen System interagieren, also mit dem System die Welt verändern oder die Welt des Systems verändern. Dazu wird er eine Bedienhandlung planen und durchführen. Damit die Veränderung der Welt geschehen kann, erwartet er von dem technischen System, dass es diese Bedienhandlung so interpretiert wie es seine Intention war. Das technische System muss also die Intention zurückgewinnen. Regelbasiert wird das technische System eine technische Handlung (eine die Welt verändernde Handlung) ausführen, die der Mensch wahrnimmt und überprüft. Auf Basis der wahrgenommenen Veränderung der Welt wird der Mensch den Zyklus erneut beginnen.

Alternative Modelle, die z.B. eine reine Prozessbeschreibung ohne Rückkopplungsschleife darstellen, wie das bei Herzog [2009] vorgestellte Prozessmodell nach Rasmussen (siehe Abb. 2.7 auf Seite 31), stellen keine vollständigen Handlungsmodelle, sondern eher Modelle der kognitiven Verarbeitung dar. Diese Modellvorstellungen sind für das Verständnis des menschlichen Handelns



**Abbildung 2.6: Sieben Stadien des Handelns**

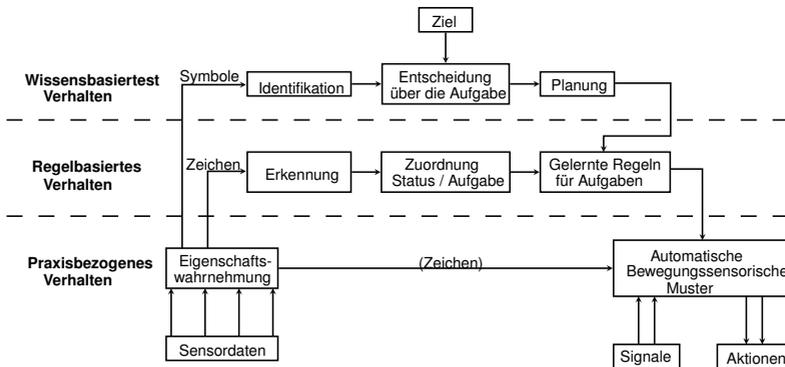
Darstellung frei übersetzt nach [Norman, 1988]

sehr hilfreich und erklären gleichzeitig, in welcher Ebene die Wahrnehmung eines technischen Systems angesiedelt ist.

Im weiteren Verlauf dieses Kapitels werden bestimmte Aspekte dieser Modellvorstellungen wieder aufgegriffen. Die technische Gestenerkennung, wie sie in dieser Arbeit beschrieben wird, wird sich in der Ebene des regelbasierten Verhaltens abspielen. Die Systemreaktion bzw. die Systementscheidung basiert auf den in der Arbeit vorgestellten Regeln (Algorithmen, siehe Kap. 4). Wünschenswert ist die Nutzung der obersten Ebene, das wissensbasierten Verhaltens, das durch Kenntnis über das Ziel oder Wissen über die Ausgangsobjekte die Erkennung deutlich verbessern würde, siehe dazu Kapitel 2.5 auf Seite 36 und Kapitel 6.6 auf Seite 132.

## 2.4 Kognitive Leistungsfähigkeit des Menschen

Nach Zimbardo et al. [1999, S. 13] umfasst der Begriff Kognition (lat. cognitio = Erkenntnis) „alle Prozesse und Strukturen, die traditionell mit dem Etikett »geistig« versehen wurden, also etwa die Prozesse des Wahrnehmens,



**Abbildung 2.7: Prozessmodell nach Rassmussen**

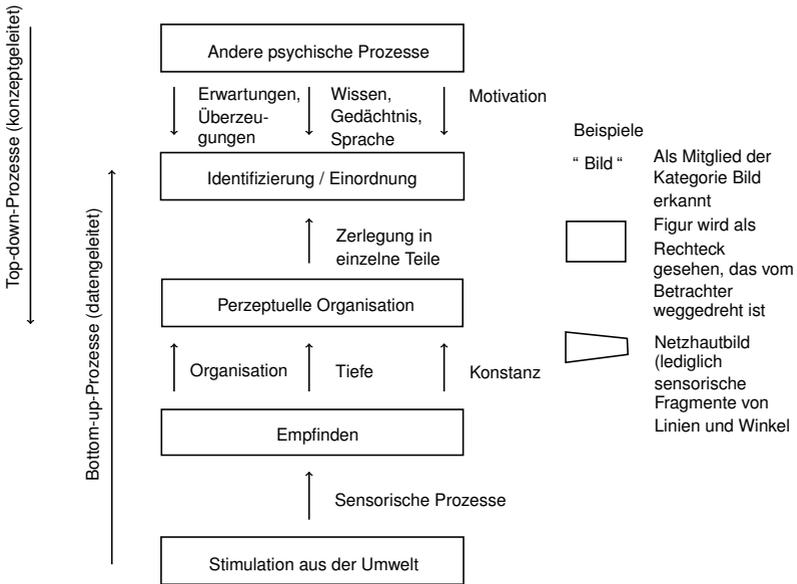
Darstellung frei übersetzt nach [Herczeg, 2009]

Schlußfolgerns, Erinnerns, Denkens, Problemlösens und Entscheidungen und die Strukturen des Gedächtnisses, die Begriffe und die Einstellungen“. Birbaumer [1993, S. 437] definiert die „kognitiven Funktionen [als] alle bewussten und nicht bewussten Vorgänge, die bei der ‚Verarbeitung‘ von organismusexterner oder -interner Informationen ablaufen, z.B. Verschlüsselung (Kodierung), Vergleich mit gespeicherten Informationen“.

## Wahrnehmung

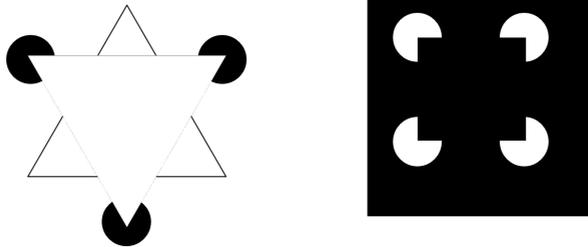
Für die Fragestellungen im Hinblick auf die Zeigegeste ist gerade der Prozess des Wahrnehmens, Erkennens und des Vergleichens mit gespeicherten Informationen entscheidend. So werden für die optische Wahrnehmung der Wirklichkeit in der Literatur zwei Prozessrichtungen vorgestellt, Top-Down und Bottom-Up. Beide Prozessrichtungen kann der Mensch nutzen, um die Geste, aber auch um das Zielobjekt zu erkennen.

Wie in Abbildung 2.8 auf der nächsten Seite dargestellt, geht der Bottom-up-Prozess von der visuell aufgenommenen Umgebung (datengeleitet), der Top-Down-Prozess von den erwarteten Objekten (konzeptgeleitet) aus. Der Mensch kann, abhängig von seiner Aufgabe, in einer ihm unbekanntem Umgebung ein beliebiges ihm bereits bekanntes Objekt erkennen, auch wenn er nur Teile des Objektes wirklich wahrnehmen kann (siehe Abb. 2.9 auf Seite 33). Alternativ,



**Abbildung 2.8: Stufen des Wahrnehmungsprozess**

Darstellung nach [Zimbardo et al., 1999]



**Abbildung 2.9: Optische Täuschung, Objekte können vom Menschen durch Fortsetzung von Bildteilen erkannt werden**

Darstellung nach [Dahm, 2006]

wenn ihm das Objekt das er erkennen soll bereits bekannt ist, kann er dieses auch bei ungenauen Lokalisierungshinweisen finden unabhängig davon, ob die notwendige Information verbal („im Regal hinter der Tür“) oder über eine Geste übermittelt werden.

## Lernen und Gedächtnis

Die Frage: „Wie wird Zeigen eigentlich definiert?“ muss im Anschluss noch detaillierter behandelt werden, obwohl der Mensch von klein auf über das „Zeigen“ Zielobjekte kennen lernt. Damit ist für ihn das „Zeigen“ eine unverzichtbare Geste im Rahmen seiner Entwicklung. Dazu schreibt Zimbardo et al. [1999, S. 477]: „Bereits Säuglinge setzen Zeigegesten (»pointing«) [...] ein, um ihre Bedürfnisse, Interessen und Präferenzen mitzuteilen“.

Auch von Wundt wird die Zeigegeste bei Kleinkindern beschrieben. „Unter den Ausdrucksbewegungen des Kindes kommt das Hindeuten auf die Gegenstände am frühesten und am selbständigsten zum Zweck der Mitteilung, also in der allgemeinen Bedeutung der Sprachgebärde vor“ [Wundt, 1900, S. 157].

McNeill [1992] untersetzt dies Aussagen damit, dass mit ungefähr zwölf Monaten ein Kind definiertes Zeigen beherrscht. Eine überzeugende Demonstration dieser ersten Form des Zeigens ist, wenn das Kind in die Richtung eines gewünschten Objektes zeigt und von diesem weg, zu einem Erwachsenen, der sich in einer anderen Richtung befindet, schaut. Diese Geste könnte auch

noch Teil der Reichbewegung sein, bindet den Erwachsenen aber deutlich in die Kommunikation mit ein.<sup>3</sup>

Zeigen ist eine Geste, die wir nicht mühevoll erlernen müssen, sondern sie ist Bestandteil unserer kindlichen Entwicklung. Lernen ist ein Prozess, der beim Menschen über verschiedene Sinnesmodalitäten abläuft (siehe Abb. 2.2 auf Seite 24). Die unterschiedlichen zu lernenden Inhalte werden genauso unterschiedlich im Gedächtnis abgelegt. Birnbaumer unterstreicht, das es zwei grundlegenden Formen des Gedächtnisses gibt:

- Das deklarative (explizite)
- und das prozedurale (implizite) Gedächtnis.

Das „deklarative (explizite) Gedächtnis ist für die bewusste Wiedergabe von Fakten und Ereignissen verantwortlich.[Das] prozedurale (implizite oder nicht-deklarative) Gedächtnis ist für die Wiedergabe von Fertigkeiten, Gewohnheiten, Bewegungsfolgen und Regeln sowie klassische Konditionierung verantwortlich. Prozedurales Lernen erfolgt im allgemeinen nicht bewusst, es besteht daher kein direkter Zugriff zum Gedächtnisinhalt“ [Birbaumer, 1993, S. 416].

Zeigen, als kindlich gelernte Geste, findet sich also im Bereich des impliziten Wissens. Somit wird die Geste von jedem Probanden individuell ausgeführt, abhängig davon, wie er diese Fertigkeit erlernt hat.

Gerade die intuitive Ausführung der Geste aus der individuellen Prägung soll nicht durch ein speziell zu erlernendes Gestenalphabet oder eine -sprache überlagert werden. Das diese Forderung im besonderen für die Zeigegeste gilt hat schon Wundt aufgezeigt: „Wo eine individuelle Gebärdensprache vollkommen neu sich ausbildet, wo etwa ein Taubstummer isoliert in hörender Umgebung aufwächst, da ist daher anfänglich die Hinweisung auf die Objekte fast die einzige überhaupt vorkommende Gebärde, und sie genügt für diesen Anfang um so mehr, weil sich bei dieser ersten Entwicklung das Interesse, das zur Mitteilung führt, nur solchen Gegenständen zuwendet, die der unmittelbaren Wahrnehmung zugänglich sind“ [Wundt, 1900, S. 157].

## Mentales Modell

Unter dem mentalen Modell wird allgemein das Modell eines technischen Systems verstanden, dass sich ein bestimmter Mensch von diesem technischen System gebildet hat. Dieses mentale Modell unterscheidet sich von Mensch zu

---

<sup>3</sup>Frei übersetzt nach [McNeill, 1992, S. 300]

Mensch. Wenn die Mensch-Technik-Interaktion so gestaltet ist, dass verschiedene Menschen ein vergleichbares mentales Modell für das zu bedienende technische System aufbauen, dann spricht man von einer gut gestalteten Mensch-Maschine-Schnittstelle.

Herczeg geht davon aus, dass diese Modelle „mehr oder weniger geeignete geistige Vorstellungen“ [Herczeg, 2009, S. 51] des Systemes sind. Unter den 15 Entwurfsprinzipien für die Entwicklung interaktiver Systeme von Preim nimmt die Forderung „Hilf Benutzern, ein mentales Modell zu entwickeln!“ [Preim, 1999, S. 55] den zweiten Platz ein. Das mentale Modell hat für Preim eine entscheidende Bedeutung, denn „wenn Menschen etwas lernen, versuchen sie, es zu verstehen, also ein Modell davon zu entwickeln, wie etwas funktioniert“ [Preim, 1999, S. 56].

Norman [1983] beschreibt sechs grundsätzliche Beobachtungen zu mentalen Modellen:

1. Mentale Modelle sind unvollständig.
2. Die Fähigkeit von Menschen, ihre Modelle zu nutzen ist sehr beschränkt.
3. Mentale Modelle sind instabil. Menschen vergessen Details eines Systems das sie verwenden, vor allem jene Details, die sie länger nicht genutzt haben.
4. Mentale Modelle haben keine festen Grenzen, ähnliche Geräte werden oft miteinander verwechselt.
5. Mentale Modelle sind unwissenschaftlich. [...]
6. Mentale Modelle sind unökonomisch. Der Mensch führt lieber körperliche Tätigkeiten aus und vermeidet eine komplexe geistige Planung.<sup>4</sup>

Eine intuitiv ausgeführte Zeigegeste, als kindlich erlernte Geste, ist unabhängig von dem technischen System mit einer Bedeutung, mit einem mentalen Modell belegt. Wie soll sich also für eine Gestenerkennung ein anderes mentales Modell entwickeln, ohne dass der Benutzer das alte Wissen vergessen und neue Syntax lernen soll?

Ziel dieser Arbeit ist, das bekannte „mentale Modell“ einer Zeigegeste aufzugreifen und für eine praktische Anwendung der Interaktion zwischen Mensch und technischem System nutzbar zu machen.

---

<sup>4</sup>Frei übersetzt nach [Norman, 1983, S. 8] mit Unterstützung von [Zühlke, 2004, S.28]

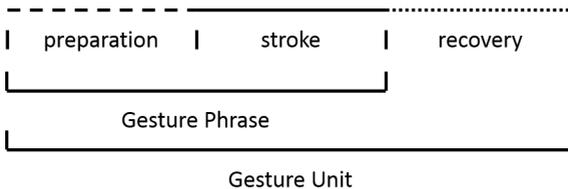
## 2.5 Kontext-Information

Kontext, vom lat. contextus: fortlaufen, verflochten bzw. contexto: zusammenweben, zusammensetzen, zusammenflechten, bezeichnet nach [Herczeg, 2009, S. 97] „generell Umgebungen, Einbettungen, Zusammensetzungen, Zusammenhänge, Verflechtungen und Vernetzungen bei der Bearbeitung von Aufgaben durch die betrachteten Benutzer“ (Zielgruppe).

Kontext-Information bzw. „Context“ wird von [Dey, 2001] definiert als: „Jede Information, die zur Charakterisierung der Situation eines Objektes (entity) genutzt wird. Objekte sind Personen, Orte oder aber auch Dinge, die zur Interaktion des Nutzers mit der Anwendung notwendig sind, einschließlich des Nutzers und der Anwendung selbst“<sup>5</sup>.

Da der Mensch beide im Vorfeld erläuterten Denkweisen auch iterativ für die Erkennung von Objekten beherrscht, nutzt er diese auch bei der Erkennung der Zeigegeste oder des Zielobjektes. Für den Top-Down-Prozess kann der Mensch jede zusätzliche Information verwenden, die er aus dem Kontext, der Situation gewinnt (siehe Abb. 2.8 auf Seite 32).

## Sprache als Informationsträger



**Abbildung 2.10: Phasen einer Geste nach Kendon**

Darstellung nach [Kendon, 2004]

Kendon [2004] untersuchte für sein Phasenmodell (siehe Abb. 2.10) vor allem die sprachbegleitenden Gesten, auch wenn ihm bewusst ist, dass wenn Sprache nicht verfügbar ist, die Gesten eine Sprache für sich selbst bilden<sup>6</sup>. Die sprachbegleitende Geste gewinnt ihre Kontextinformation überwiegend aus der Rede, zum Beispiel: „Der grüne Ordner dort oben.“ Der Mensch kann diese

<sup>5</sup>Frei übersetzt nach [Dey, 2001, S. 5]

<sup>6</sup>Frei übersetzt nach [Kendon, 2004, S. 3]

zusätzliche Information aufnehmen und in seine Auswertung der Geste integrieren. Ihm ist bekannt, wie ein Ordner aussieht und das Objekt „Ordner“ wurde vom Kommunikationspartner bereits präzisiert, dieses Objekt „Ordner“ hat die Farbe grün. Außerdem ist durch die Sprachinformation auch eine Information über die zu erwartende Geste kodiert. Bei dem zuvor genannten Beispiel wird ein Zeigen erwartet, sodass der angesprochene Kommunikationspartner die nächste Geste (konzeptgeleitet) entsprechend deuten kann.

Fricke geht darüber hinaus. Sie behauptet, dass die „redebegleitenden Gesten [...] nur unter Berücksichtigung der zugehörigen vokalen Äußerung interpretierbar.“ sind [Fricke, 2007, S. 150]. Zur Untersetzung ihrer Aussage bedient sie sich eines Beispiels von McNeil über einen Redner, der eine Szene aus einem Comic gestenhaft verbildlicht. Diese Geste ist bei McNeill [1992] unter den „Iconics“ eingeordnet.

In Kap. 3.2 auf Seite 61 wird diese Klassifikation von Gesten für die weitere Betrachtung eingeschränkt, obwohl nach Fricke die Zeigegeste „im Kendonschen Kontinuum keiner einheitlichen Klasse“ angehört, und „die konkreten Vorkommnisse [...] daher jeweils, je nach den vorliegenden Umständen, anders einzuordnen“ sind [Fricke, 2007, S. 152].

## Orientierung der Hand

Für das Gestenverständnis in der Mensch-Mensch-Kommunikation ist die Kontextinformation entscheidend. „Wenn sich Menschen unterhalten, ist es ihnen implicit möglich, Situationsinformation, also den Kontext zu nutzen um die Gesprächsbandbreite zu erhöhen“<sup>7</sup> schreibt dazu Dey [2001]

Bei der Aufzeichnung der Gestensequenzen, die intuitiv von den Probanden gewählt wurden, konnte beobachtet werden, dass sich die Grundhaltung der Probanden beim Zeigen und Heranwinken in der Grundorientierung der Handfläche deutlich unterscheiden. So zeigt beim Heranwinken die Handfläche meist nach oben, der Kommunikationspartner wird eingeladen, herzukommen. Beim „Zeigen“ aber zeigt die Handfläche nach unten und wird durch den Handrücken abgedeckt (siehe Abb. 2.11 auf der nächsten Seite). Auch diese, eher beiläufig wahrgenommene Information unterstützt den menschlichen Kommunikationspartner bei dem Verständnis der Geste.

---

<sup>7</sup>Frei übersetzt nach [Dey, 2001, S. 4]



**Abbildung 2.11: Ausrichtung der Handfläche beim Zeigen bzw. Heranwinken**

(a) Heranwinken (b) Zeigen

Baudel und Beaudouin-Lafon [1993] nutzen diese zusätzliche Information der Orientierung der Hand für die Erkennung einer Zeigegeste. „Wenn in die aktive Zone gezeigt wird, gilt das Kommando [Blättern] als erteilt, wenn die Orientierung der Handfläche nach rechts (Daumen nach unten), alle Finger gestreckt und die Hand von links nach rechts bewegt wird. Die Geste ist beendet durch Schließen der Hand oder durch Verlassen der aktiven Zone“<sup>8</sup>.

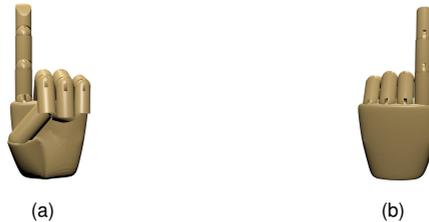
Fricke [2007] beschreibt dieses Verhalten bei einer weiteren Gestenform, die sie der „Zeigegeste als Emblem“ zuordnet. So ist nach Ihrer Definition auch die Achtung-Geste „ein Beispiel für eine emblematische Geste in Form einer Zeigegeste [...] Der Zeigefinger zeigt senkrecht nach oben, die anderen Finger sind gekrümmt, wobei der Daumen den Mittelfinger berührt. Die Handfläche ist dem Adressaten zugewandt. Der dieser Form zugeordnete Inhalt lässt sich mit dem Ausruf ‚Achtung!‘ paraphrasieren“ [Fricke, 2007, S. 153](siehe Abb. 2.12(a) auf der nächsten Seite). In der Praxis sieht man diese Zeigegeste aber auch, wenn sich zwei Kommunikationspartner nur kurzzeitig begegnen und dieses „Achtung!“ zu einer Grußgeste verkommt, ausgeführt von zwei Motorradfahrern, welche die Hand nicht vom Lenker lösen können oder auch von dem Fahrer eines PKW, der nur fix über das gegriffene Lenkrad grüßen will.

„Richtet man die Handfläche der Achtung-Geste anders aus, nämlich dem Gestikulierenden selbst zugewandt, erhält man eine Drohgebärde, die man mit ‚Wehe!‘ paraphrasieren könnte“ [Fricke, 2007, S. 154](siehe Abb. 2.12(b) auf der nächsten Seite).

Dieses Verständnis bzw. diese drei unterschiedlichen Bedeutungen der Achtung-Geste sind nicht in der eigentlichen Gestenbeschreibung hinterlegt, sondern im gelernten Verhalten des Menschen und werden implizit wahrgenommen. Als Kontextinformation und damit zur Unterstützung der Geste ist dieses Wissen von entscheidender Bedeutung, da es ebenso wie die Kommu-

<sup>8</sup>Frei übersetzt nach [Baudel und Beaudouin-Lafon, 1993, S. 32]

nikation durch Körpersprache unbewusst wahrgenommen wird und die verbale Kommunikation verstärkt bzw. die nonverbale Kommunikation bestätigt.



**Abbildung 2.12: emblematische Zeigegesten**

Darstellung nach Idee von [Fricke, 2007]

(a) Achtung! (b) Wehe!

## 2.6 Technisches System

Das mentale Modell beschreibt das Verständnis des Menschen über das technische System mit dem er interagieren will. Da jeder Mensch sein eigenes Modell zum Verständnis des technischen Systems entwickelt, ist eine allgemeine Definition des technischen Systems kaum auszumachen. Zühlke [2004, S. 3] unterstellt, dass „Ingenieure in ihrer Ausbildung sehr viel über Eigenschaften technischer Systeme lernen“. Daher ist es für ihn opportun, das technische System nicht näher zu definieren. [Johannsen, 1993, S. 21] verwendet „die Begriffe Maschine und technisches System [...] gleichbedeutend.“ Dazu übersetzt er, „dass jedes technische System allgemein in einen technischen Prozess, ein Automatisierungssystem [...] und ein Unterstützungssystem untergliedert wird“. Im Bezug auf die Automatisierung ergänzt er weiter, jede „Erhöhung des Automatisierungsgrades [verschiebt] nur die Schnittstelle zum Menschen im Sinne einer veränderten Funktionsteilung zwischen Mensch und Maschine“ [Johannsen, 1993, S. 22]. Darüber hinaus gilt die Erläuterung aus der VDI/VDE 3850-1 [2000], die vergleichbar zu Johannsen unter dem technischen System die drei Komponenten „Bediensystem“, „Steuerung“ und „Maschine/Prozess“ versteht.

Den menschlichen Fähigkeiten steht in der Mensch-Technik-Interaktion als Kommunikationspartner ein technisches System gegenüber. Ziel der Mensch-Technik-Interaktion ist es, die besonderen Fähig- und Fertigkeiten der beiden Interaktionspartner zur Lösung einer gemeinsamen Aufgabe zu nutzen. In der

VDI/VDE 3850-1 wird auf diese Verbindung schon im Geltungsbereich (siehe Abb. 2.1 auf Seite 22) und im Zweck hingewiesen: „Zweck dieser Richtlinie ist die Angabe von Regeln und Empfehlungen zur einheitlichen Gestaltung von Bediensystemen für Maschinen sowie Hinweise für deren Anwendung mit dem Ziel, die Eigenschaften des Bediensystems an die Fähigkeiten des Menschen anzupassen“ [VDI/VDE 3850-1, 2000, S. 3]. Das Bediensystem wird von der Richtlinie als „Gesamtheit aller wahrnehmbaren Komponenten einer technischen Einrichtung [...], die der Interaktion mit ihren Benutzern dient“ [VDI/VDE 3850-1, 2000, S. 6] definiert. Diese Definition umfasst auch die technischen Einrichtungen, die für die Interaktion mittels Freiraumgesten benötigt werden, nicht nur die Anzeige- und Bedienelemente einer klassischen bildschirmgestützten Variante.

Johannsen [1993, S. 21] erläutert den technischen Prozess näher, da dieser „ein physikalisch-technischer oder ein chemisch-technischer Vorgang mit je einem Material-, Energie- und/oder Informationsfluss am Eingang und Ausgang“ ist. „Material, Energie und/oder Information [gelangen] als Eingangsgröße [...] in den technischen Prozess, in dem sie verarbeitet oder umgewandelt werden, um dann als Ausgangsgrößen [...] durch den technischen Prozess verändert wieder herauszutreten.“

## **Besondere Fähigkeiten des technischen Systems**

So wird zwar deutlich, was das technische System leistet, aber nicht, wo sich die Leistungsfähigkeit des technischen Systems von den Fähigkeiten und Fertigkeiten des Menschen abhebt. Dazu müssen verschiedenartige technische Prozesse im Detail analysiert werden. Das technische System unterstützt die menschlichen Fähigkeiten bestmöglich und hilft, negative Belastungen zu vermeiden sowie Risiko-Situationen (Gefahren) zu entschärfen.

Am Beispiel der Hebe- und Trageaufgabe kann dieses Prinzip deutlich gemacht werden. Der Arbeitnehmer wird zu Schichtbeginn die Hebe- und Tragehilfe nicht einsetzen, da er noch ausgeruht ist und sich fit genug fühlt, die benötigte Last aufzunehmen. Mit zunehmender Ermüdung wird er die Hebehilfe einsetzen, da sie ihn entlastet oder die Fortführung der Arbeit überhaupt erst ermöglicht. Die besonderen Fähigkeiten dieses technischen Systems sind die aufzubringende Kraft sowie die entsprechende Ausdauer. Das System kann beliebig häufig die Last anheben und, sofern das technische System entsprechend ausgelegt ist, auch Lasten anheben und bewegen, die der Mensch nicht, oder zumindest nicht alleine, anheben könnte.

Betrachten wir nicht physische Prozesse sondern informatorische, findet man auch für diese Prozesse Beispiele, in denen das technische System seine Stärken ausspielen kann. So ist in Lagerbereichen nicht nur das Handling von Lasten ein durch ein technisches System gut zu unterstützende Tätigkeit, sondern auch die Lagerverwaltung. So kann ein informationstechnisches System die „Chaotische Lagerhaltung“ beherrschen, wenn zu jedem Teil der entsprechende Lagerort gespeichert wird. Wird ein Teil umgelagert, so wird auch der neue Lagerort sofort im technischen System abrufbar, während ein Mensch zuerst noch an dem bisherigen Lagerplatz suchen würde.

So lassen sich für alle technischen Prozesse die besonderen Fähigkeiten dieses technischen Systems darstellen. Was aber bei der Gesamtheit der technischen Systeme bisher noch nicht möglich ist, ist das Lösen von Aufgaben, für die das System nicht entwickelt oder zumindest angepasst wurde. So kann das Hebe- und Tragesystem in der „Chaotischen Lagerhaltung“ ein bekanntes Teil an einem beliebigen Lagerplatz aufnehmen und handhaben. Wurde das Teil umgelagert und der neue Lagerort nicht gespeichert, oder wurden die Aufnahmepunkte so modifiziert, dass der Greifer nicht mehr passt, ergeben sich für ein technisches System nicht lösbare Aufgaben.

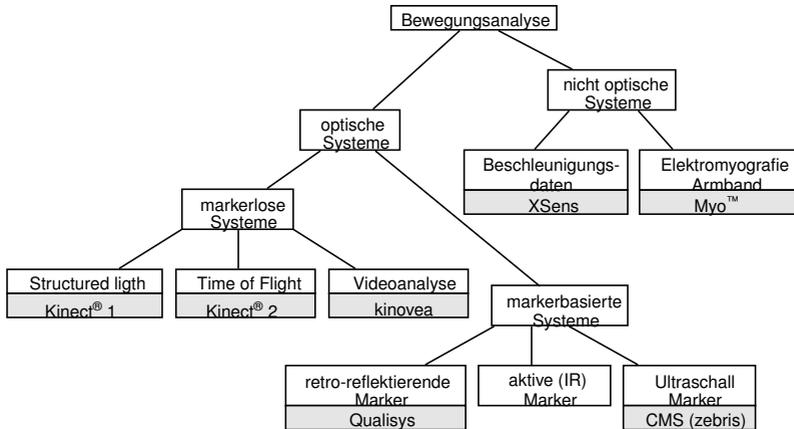
An dieser Aufgabe komplettiert sich der Systemverbund Mensch-Technik. Der Mensch „unterstützt“ das technische System beim Auffinden des Objektes und modifiziert den „Greifer“, damit das technische System wieder seine Vorteile wie Kraft und Ausdauer ausspielen kann.

## **Gestenaufnahmesysteme**

Soll dem technischen System (dem Hebe- und Tragesystem) der neue Lagerort mitgeteilt werden, können dazu verschiedene Eingabesysteme zum Einsatz kommen, z.B. die Eingabe der Bezeichnung des Lagerortes an einer Tastatur oder über das Scannen eines am Lagerort angebrachten Barcodes. Alternativ soll in dieser Arbeit das technische System mittels einer Freiraumgeste über den neuen Lagerort informiert werden.

Somit gilt es, die Systeme, welche eine technische Aufnahme von Freiraumgesten ermöglichen, gesondert zu betrachten. Diese Systeme können grundsätzlich in zwei Klassen unterteilt werden: die „optischen“ und die „nicht optischen“ Systeme (siehe Abb. 2.13 auf der nächsten Seite).

Die „nicht optischen“ Systeme gewinnen die zur Bewegungsanalyse notwendigen Daten auf Basis von am Körper getragenen Sensorsystemen. Dazu gehören Systeme, die z.B. inertielle Messeinheiten einsetzen, wie das System der Firma XSens. Dieses System nutzt 17 Beschleunigungs- und Drehraten-



**Abbildung 2.13: Bewegungserkennungssysteme**

Die beispielhaft genannten Systeme erheben keinen Anspruch auf Vollständigkeit und bilden die Marktsituation nur exemplarisch ab.

sensoren, um Bewegungen der unterschiedlichen Körperteile zu erfassen. Überwiegend werden diese Sensoren auch bei den „Datenhandschuhen“ verwendet. Ein weiteres, seit einigen Jahren auch kostengünstig zu erhaltendes „nicht optisches“ System basiert auf der Messung der Muskelaktivität (Elektromyografie, kurz EMG). Dazu trägt der Nutzer mit dem System Myo™ ein Armband, das die Muskelaktivität der Hand misst. Die nicht optischen Systeme scheiden für die weitere Betrachtung in dieser Arbeit aus, da der Nutzer zusätzliche Sensoren am Körper tragen muss.

Die „optischen“ Systeme lassen sich in zwei Untergruppen einteilen: die „markerbasierten“ und die „markerlosen“ Systeme. Auch die „markerbasierten“ Systeme wurden für die weitere Betrachtung ausgeschlossen, da der Aufbau der Systeme ungeeignet für den Praxiseinsatz ist.

Am Beispiel des Bewegungsanalyse-System Qualisys® soll diese Feststellung kurz erläutert werden. Das System ist mit einer Datenrate von bis zu 1000 frames per second (fps) sehr gut geeignet, um auch schnelle Bewegungen, wie sie bei der Gestenausführung vorkommen können, aufzuzeichnen. Außerdem bietet es durch die Verwendung von mehreren Kameras eine vollständige Erkennung aller Marker und bei geeigneter Position und Ausrichtung der Kameras keine Verdeckungen. Dieser Vorteil entwickelt sich für die Praxiseinführung zu einem Nachteil. Das System kann nur ortsfest betrieben

werden. Nach einem Ortswechsel muss das neue Setup aufwändig kalibriert werden. Darüber hinaus sind an einer Vielzahl relevanter Körperstellen des Nutzers retroreflektierende Marker anzubringen.

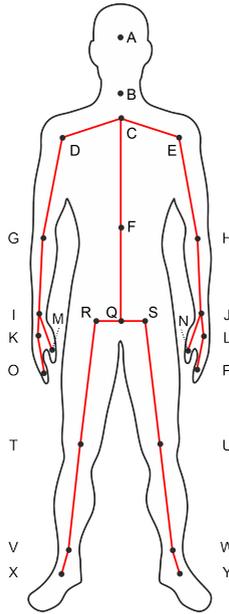
Das Ultraschallsystem der Firma zebri<sup>®</sup> wurde den „markerbasierten“, „optischen“ Systemen zugeordnet, da für das System vergleichbare Einschränkungen gelten.

Bleiben die „markerlosen“ „optischen“ Systeme. Diese Systeme haben mit der Markteinführung der Kinect<sup>®</sup> 1 2010 einen immense Nachfrage erlebt. Ein besonderer Vorteil dieser für den Spielesektor entwickelten Geräte ist neben dem Preis die sich rasch entwickelnde „Open Community“. Gerade im wissenschaftlichen Sektor stellt sie Bibliotheken für die „Skeletterkennung“ bereit und entwickelt diese ständig weiter. Nachdem die Kinect<sup>®</sup> 1 von der Kinect<sup>®</sup> 2 abgelöst wurde, wurde zeitgleich von Microsoft<sup>®</sup> eine Bibliothek zur Erkennung von 25 „Skelettpunkten“ bereitgestellt (siehe Abb. 2.14 auf der nächsten Seite). Während für andere Tiefensensoren diese Daten erst durch eigene Bildverarbeitungsalgorithmen gewonnen werden müssen, kann bei Nutzung der Kinect<sup>®</sup> sofort mit der Entwicklung der technischen Gestenbeschreibung begonnen werden. Da die Kinect<sup>®</sup> den Zugriff auf die Rohdaten ebenfalls zulässt, sind in der Community für die „Skeletterkennung“ alternative Open Source Lösungen verfügbar, die eine Anpassung der Algorithmen ermöglichen. Nachteilig ist jedoch, dass die von Microsoft<sup>®</sup> genutzten Modelle zur „Skeletterkennung“ und Berechnung nicht öffentlich bekannt sind.

„Markerlose“ Systeme bieten damit gerade für den Praxiseinsatz entscheidende Vorteile. So darf für ein System, das für eine Spielekonsole entwickelt wurde, das Einrichten des Sensors nicht besonders aufwändig sein. Darüber hinaus muss der Nutzer keine besondere Kleidung (mit Sensoren oder Markern) tragen. Nur so kann gewährleistet werden, dass sich in entsprechenden Nutzungskontexten die Nutzer schnell bei der Bedienung des technischen Systems ablösen können. Betrachtet man die Mensch-Technik-Interaktion unter den Gedanken „Industrie 4.0“ - Mensch und Technisches System interagieren gleichzeitig, gemeinsam in einem Wirkungsbereich - so ist eine System zur Aufnahme von Kommandos ohne zusätzliche Eingabelemente notwendig, da der Nutzer, in diesem Fall der Werker, z.B. bereits ein Werkzeug in der Hand hat.

## **Kinect<sup>®</sup> 2**

Ein paar Besonderheiten der Kinect<sup>®</sup> 2 und der eingesetzten Microsoft<sup>®</sup> Bibliotheken gilt es an dieser Stelle vorzustellen und für die weitere Arbeit festzulegen.



Kinect® 1		Kinect® 2
Kopf	A	Kopf
Nacken	B	Nacken
Schulter (rechts/links)	D/E	Wirbelsäule Schulter (WsS)
Torso	F	Schulter (rechts/links)
Ellbogen (Eg) (rechts/links)	G / H	Wirbelsäule Mitte
Hand (rechts/links)	I / J	Ellbogen (Eg) (rechts/links)
	K / L	Handgelenk (Hg) (rechts/links)
	M / N	Hand (rechts/links)
	O / P	Daumen (rechts/links)
	Q	Zeigefinger (rechts/links)
Hüfte (rechts/links)	R / S	Wirbelsäule Basis (WsB)
Knie (rechts/links)	T / U	Hüfte (rechts/links)
Fuß (rechts/links)	V / W	Knie (rechts/links)
	X / Y	Sprunggelenk (rechts/links)
		Fuß (rechts/links)

Abbildung 2.14: Stickfigure und Skelettpunkte von Kinect® 1 und 2 im Vergleich

- Das Koordinatensystem, das allen Berechnungen in der Arbeit zugrunde liegt, leitet sich von dem Koordinatensystem der Kinect® ab. Die Kinect® befindet sich in  $Z = 0$ . Die Tiefenkoordinate  $Z$  wird in Metern, als Abstand Sensor - Messpunkt angegeben.  
Die  $x$ - und  $y$ -Koordinaten beziehen sich auf die Bildebene und werden in Pixel (px) angegeben. Der Ursprung befindet sich konform mit den Vereinbarungen der Computergraphik links oben in der Bildebene. Daraus resultiert z.B., dass die  $y$ -Koordinate des Kopfes einer kleinen Person größer ist, als die  $y$  Koordinate einer großen Person.
- Der Begriff „Skelett“ bezeichnet in der vorliegenden Arbeit die auf Abb. 2.14 auf der vorherigen Seite dargestellte Stickfigure und nicht das anatomisch korrekte Skelett eines Menschen. Mit den angegebenen „Skelettpunkten“ verhält es sich gleichermaßen.

## 2.7 Akzeptanz

Akzeptanz ist die positive Einstellung des Nutzers zu einem System, die vielschichtige Ursachen haben kann. In der Usabilityforschung wird die Usability (Gebrauchstauglichkeit) durch die Beurteilungsgrößen „Effektivität, Effizienz und Zufriedenheit“ in einem definierten Nutzungskontext beschrieben (siehe Abb. 2.15 auf der nächsten Seite).

Die die Parameter „Effektivität“ und „Effizienz“ lassen sich nach der DIN EN ISO 9241-11 [1999] gut bestimmen. Sie sind wie folgt definiert:

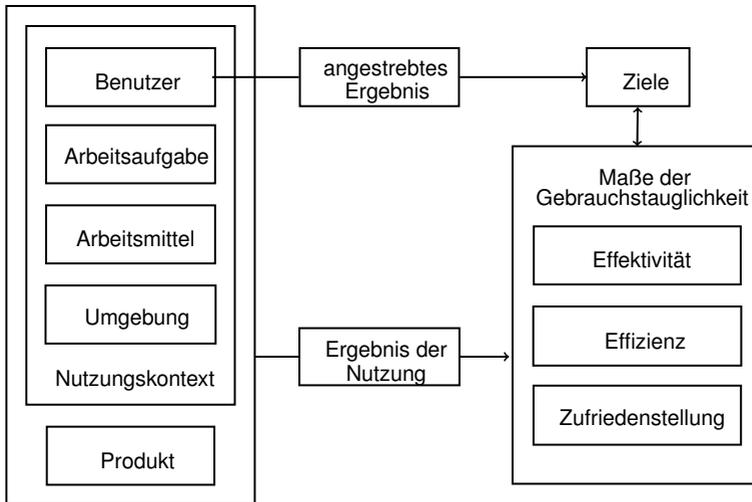
**Effektivität:** Die Genauigkeit und Vollständigkeit, mit der Benutzer ein bestimmtes Ziel erreichen.

**Effizienz:** Der im Verhältnis zur Genauigkeit und Vollständigkeit eingesetzte Aufwand, mit dem Benutzer ein bestimmtes Ziel erreichen [DIN EN ISO 9241-11, 1999, S. 4].

Diese Parameter lassen sich anhand des Nutzungskontextes einfach evaluieren. Dazu stellt man dem Probanden entsprechende Aufgaben und bestimmt die Fehlerrate bzw. die Zeit, als Indikator des Aufwandes, die der Proband benötigt um die Aufgabe zu lösen.

Dahingegen ist der Parameter „Zufriedenheit“ genauso subjektiv und schwer messbar wie die Akzeptanz.

**Zufriedenheit** Freiheit von Beeinträchtigungen und positive Einstellungen gegenüber der Nutzung des Produkts. [DIN EN ISO 9241-11, 1999, S. 4]



**Abbildung 2.15: Anwendungsrahmen für Gebrauchstauglichkeit**

Darstellung nach [DIN EN ISO 9241-11, 1999]

Schon allein aus der Definition lässt sich ableiten, dass die Zufriedenheit ein individuell subjektives Empfinden des Nutzers über die Nutzung eines Produktes charakterisiert. In ihrer Verbesserung sieht Zühlke [2004, S. 43] die überwiegenden Gründe für die Nutzerbeteiligung. Die Nutzerbeteiligung bringt

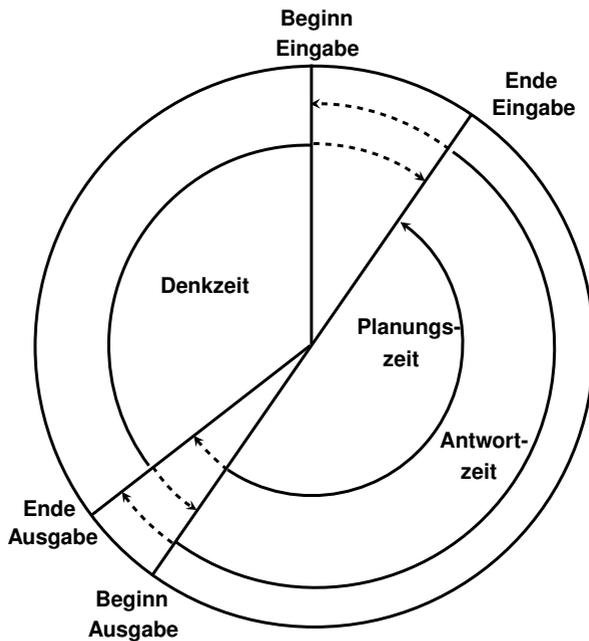
- **dem Hersteller** „höhere Akzeptanz durch Nutzerorientierung“, das ist das Produkt, das der Nutzer will und
- **dem Nutzer** „größere Identifikation bzw. höhere Akzeptanz“, das ist das Produkt, das für mich bzw. nach meinen Wünschen entwickelt wurde.

Hassenzahl et al. [2003, S. 187] schlagen zur Evaluation interaktiver Produkte neben der Bestimmung der Nutzungsqualität (Gebrauchstauglichkeit, „pragmatische Qualität“) auch die Analyse der „sogenannten hedonischen Qualitätsaspekte“ vor. Dazu bieten sie mit dem „AttrakDiff“ einen Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität an. Die Hedonische Qualität beruht „auf dem menschlichen Bedürfnis nach Stimulation und Identität, während bei Gebrauchstauglichkeit (bzw. pragmatischer Qualität) der Bedarf zur kontrollierten Manipulation der Umwelt im Vordergrund steht“ [Hassenzahl et al., 2003, S. 187]. Shneiderman und Plaisant [2010, S. 172 f.] geben Hinweise zu den Akzeptanzkriterien, die aber für jedes Projekt individuell

festzulegen sind. Sie lehnen die Kriterien der Gebrauchstauglichkeit als vage ab und geben folgende fünf Aspekte beispielhaft als messbare Kriterien vor:

- Die Zeit, die ein Nutzer benötigt, um eine spezifische Funktion zu erlernen.
- Die Bediengeschwindigkeit.
- Die Fehlerrate durch den Nutzer.
- Nutzerbindung durch langwierige Bearbeitung von Befehlen.
- Subjektive Nutzerzufriedenheit.<sup>9</sup>

## Zeitverhalten



**Abbildung 2.16: zeitliche Phasen eines Interaktionsschritts**

Darstellung nach [Herczeg, 2009]

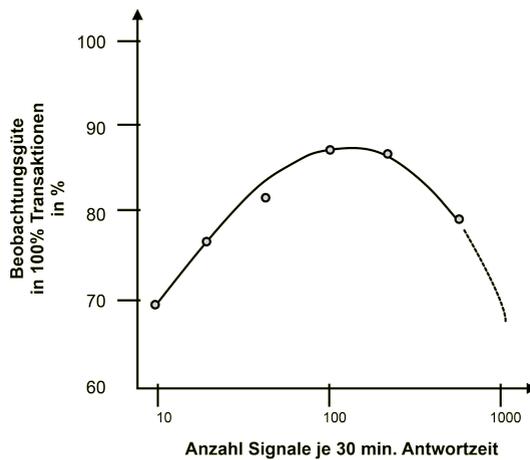
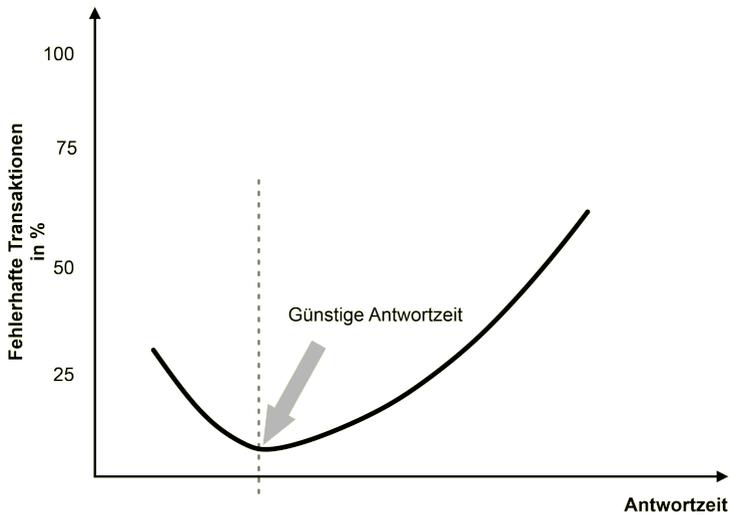
<sup>9</sup>Frei übersetzt nach [Shneiderman und Plaisant, 2010, S. 173]

Herczeg [2009, S. 145] sieht einen Aspekt der Akzeptanz in dem Zeitverhalten des Systems, vor allem, wenn die „wiederholte Interaktion mit Computersystemen [...] seitens der Benutzer Erfahrungen und Erwartungen [erzeugt], was die Zeitverläufe angeht. Treten im Verhalten interaktiver Systeme unerwartete Verzögerungen auf, so werden viele Benutzer verunsichert, verärgert oder gar frustriert, was sich direkt auf den Faktor der Zufriedenstellung der Benutzer auswirkt [...]. Neben Unzufriedenheit können Verzögerungen im Interaktionsverlauf auch zu einer höheren Fehlerrate der Systembenutzung führen.“

Diese erhöhte Fehlerrate bildet sich in der Usability in der Effektivität des Gerätes ab. Laut [Herczeg, 2009, S. 147] hat darüber hinaus das „menschliche Gedächtnis, insbesondere das Kurzzeitgedächtnis [...] einen erheblichen Einfluss auf das für Benutzer von interaktiven Systemen günstige Zeitverhalten [...]. Sind die Antwort- oder Ausgabezeiten zu lang, besteht die Gefahr, dass der Benutzer seine eigentlichen Arbeitsziele ganz oder teilweise vergisst. Die vom System gelieferten Resultate lassen sich dann nicht mehr oder nur noch unzulänglich mit den Intentionen und Zielen des Benutzers in einen Zusammenhang bringen.“

Da in der Mensch-Mensch-Kommunikation die Reaktion des Interaktions- bzw. Kommunikationspartners unmittelbar erfolgt, das heißt, dass die Antwortzeit sehr kurz ist und sich somit flüssig in die Kommunikation einfügt, wird der Zeitfaktor neben der Art und Verständlichkeit der Rückmeldung ein entscheidender Parameter für die Akzeptanz von Gestenbasierter Mensch-Technik-Interaktion sein.

Eine günstige Antwortzeit liegt nach [Herczeg, 2009, S. 149] als Faustregel bei 3-4 sec. da ansonsten die Fehlbedienungen signifikant ansteigen, wie eine Untersuchung von Shneidermann und Plaisant zeigt (siehe Abb. 2.17(a) auf der nächsten Seite). Die aufgezeigte Tendenz lässt sich durch die Beobachtungen von Schmidtke und Bernotat [1993] bestätigen (siehe Abb. 2.17(b) auf der nächsten Seite). Sie untersuchten bereits 1966 die Beobachtungsgüte in Abhängigkeit der auftretenden Ereignisse, eigentlich um Überforderungssituationen zu charakterisieren. Dabei zeigte sich aber auch, wenn die Signalfrequenz unter einen kritischen Wert von ca. 120 - 150 Signalen je 30 Minuten fällt, Fehler durch Monotonie-bedingte Ermüdung auftreten. 150 Signale je 30 Minuten entspricht einer durchschnittlichen Interaktionszeit von 12 Sekunden pro Ereignis, auf die auch [Shneiderman und Plaisant, 2010, S. 439] kommen (siehe Abb. 2.17(a) auf der nächsten Seite). Nach Herczeg [2009] (siehe Abb. 2.16 auf der vorherigen Seite) beträgt die Antwortzeit des Systems zwischen der Hälfte und Zweidrittel der gesamten Interaktionszeit, wobei die restliche Interaktionszeit dem Nutzer zur Verfügung steht.



**Abbildung 2.17: Zeitliche Abhängigkeit von Interaktionsfehlern**

(a) Fehlerhafter Interaktionen in Abhängigkeit von der Antwortzeit,  
Darstellung nach [Shneiderman und Plaisant, 2010]

(b) Signalfrequenz und Beobachtungsgüte, Darstellung nach [Schmidtke und Bernotat, 1993]

## **Zuverlässigkeit / Erwartungskonformität**

Neben der Antwortzeit, der Reaktionszeit des Systems, erwartet der Nutzer eines Interaktionssystems auch andere Formen einer Systemreaktion. Wie im Abschnitt 2.4 auf Seite 34 beschrieben, baut sich der Nutzer ein für ihn und seine Aufgabe passendes Systemmodell auf, welches die Systemreaktionen transparent beschreibt.

Aus der Mensch-Mensch-Kommunikation sind für die Zeigegeste bekannte Reaktionen, dass der Kommunikationspartner mit den Augen bzw. dem gesamten Kopf der Zeigegeste folgt und damit das Zielobjekt identifiziert. Die erfolgreiche Identifikation signalisiert der Kommunikationspartner durch z.B. ein Kopfnicken, während er die fehlgeschlagene Identifikation des Zielobjektes versuchen wird, durch Rückfragen zu präzisieren oder durch Schütteln des Kopfes das Zeigen erneut anzufordern (siehe Abb. 2.5 auf Seite 28).

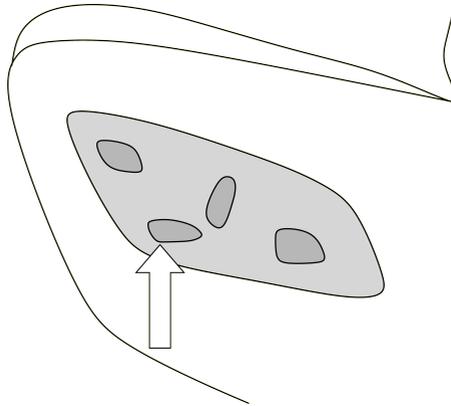
Diese Systemreaktionen erwartet der Nutzer auch in der Mensch-Technik-Kommunikation. Die Technik soll die eine erfolgreiche Identifikation des Zielobjektes bestätigen bzw. die fehlgeschlagene Identifikation signalisieren, damit die Geste erneut ausgeführt werden kann. Dabei erwartet der Nutzer aber auch, dass das technische System zuverlässig das richtige Zielobjekt identifiziert und mit zunehmender Nutzungsdauer die Zahl der fehlgeschlagenen Identifikationen abnimmt.

In einer Untersuchung zur Nutzung von Gesten auf einem Multitouch-Bildschirm haben Weyerstall und Saul [2013] ermittelt, dass für 92 % ihrer Probanden „intuitive Bedienung“ und „Schnelligkeit“ wichtig sind. Darüber hinaus sind 83% der Probanden die Fehlererkennung wichtig. „Am auffälligsten ist jedoch zu erwähnen, dass für alle Probanden, also 100 Prozent, die sichere Erkennung der einzelnen Gesten bedeutsam ist“ [Weyerstall und Saul, 2013, S. 104].

Eine 100 %ige Erkennung der Zeigegeste wird im Verlauf dieser Arbeit nicht zu gewährleisten sein. Das in der Arbeit vorgestellte System wird die aus der Auswahl des Gestenaufnahmesystems und den dazugehörigen Softwarebibliotheken resultierenden Fehler nicht beheben können. Die für die Akzeptanz wünschenswerte Lernfähigkeit des Systems bleibt z.Z. noch eine Wunschvorstellung.

## **Intuitive Bedienung**

Im Rahmen des Lernens, gerade des impliziten Lernens, wurde das Thema intuitive Ausführung einer Geste bereits angesprochen. Herczeg [2009, S. 197] schreibt dazu: „Das Kriterium, das im Zusammenhang mit der Gebrauchstaug-



**Abbildung 2.18: Intuitive Sitzverstellung**

Eigene Darstellung nach Idee von Norman [1988]

lichkeit von interaktiven Systemen wahrscheinlich am häufigsten genannt, aber am wenigsten geklärt wurde, ist die Intuitivität einer Benutzungsschnittstelle“. Wie auch von Weyerstall und Saul [2013] angesprochen fordern die Nutzer eine intuitive Bedienung.

Der Brockhaus [1988, Bd. 10, S. 582] kennt das substantivierte Adjektiv „Intuitivität“ nicht. Zu dem Adjektiv „intuitiv“ heist es: „auf Intuition beruhend“. Intuition ist „das unmittelbare, ganzheit[liche] Erkennen oder Erfahren von Sachverhalten im Gegensatz zu der u.a. durch Beweis, Erklärung, Definition vermittelten diskursiven Erkenntnis“ [Brockhaus, 1988, Bd. 10, S. 582].

[Herczeg, 2009, S. 197] schreibt dem Begriff im Kontext der Gestaltung „Interaktiver Systeme“ eine besondere Bedeutung zu, da in der Umgangssprache „Intuitiv“ so viel wie „was keiner Erklärung bedarf“ heißt. Ansonsten beruft er sich auf eine Definition von [Hurtienne et al., 2006, S. 28]: „Ein technisches System ist intuitiv benutzbar, wenn es durch nicht bewusste Anwendung von Vorwissen durch den Benutzer zu effektiver Interaktion führt.“ Eingeschränkt wird die Definition von [Hurtienne et al., 2006, S. 38] darauf, dass nur „Informationsverarbeitungsvorgänge des Menschen [...] als ‚intuitiv‘ bezeichnet werden [können]. Technik ist allenfalls benutzbar. Intuitivität ist ein Merkmal des Zusammenspiels von Mensch, Aufgabe und technischem System in einem konkreten Nutzungskontext“.

In seinem Buch „The design of every day things“ nähert sich Norman [1988] dem Thema Intuitivität so, wie der Begriff umgangssprachlich verwendet wird.

Er betrachtet Alltagsgegenstände oder -lösungen, die eingeführt und etabliert sind, und hinterfragt, was diese Lösung so gut macht, das sie sich gehalten hat. Dabei stellt er unter anderem die Hypothese auf, dass die Bedienung einer technischen Lösung, z.B. die elektrische Verstellung eines Autositzes wie in Bild 2.18 ( auf der vorherigen Seite), einfach zu verstehen und zu erlernen ist, wenn sie durch „natürliche Abbildungen“ dargestellt wird. „Das Bedienelement ist die Form des Sitzes selbst, die Abbildung ist unkompliziert. Um den Sitz vorne anzuheben, hebe auch den vorderen Teil der dargestellten Sitzfläche an.“<sup>10</sup>.

Intuitive Bedienung ist also gegeben, wenn die Bedienung für den Nutzer ohne zusätzlichen Lernaufwand sofort möglich ist. Die intuitive Nutzung des System im Rahmen der Usability-Untersuchungen ist ein subjektiver Faktor und kann bei den Akzeptanzuntersuchungen direkt der Zufriedenheit zugeordnet werden. Im Rahmen einer Akzeptanzuntersuchung gilt es Gesten, die einfach zu erlernen und ohne zusätzliche Erläuterung verstanden oder ausgeführt werden, für die Mensch-Technik-Interaktion mittels Freiraumgesten zu bestimmen und zu implementieren.

## 2.8 Zusammenfassung

In der Mensch-Technik-Interaktion verarbeiten beide Kommunikations- bzw. Interaktionspartner die gleiche Eingangsinformation. Sie unterscheiden sich in der Flexibilität bei der Verarbeitung der Information sowie bei der Gesamtsystemperformance. Zum Verständnis dieser Modellvorstellungen und zur Einführung einiger grundsätzlicher Begriffe wurden im Kapitel Propädeutik folgende Themenschwerpunkte behandelt:

- Mensch-Maschine-Schnittstelle allgemein nach VDI/VDE 3850;
- Die Geste als hinweisende Gebärde;
- Kommunikation auf Basis des Kommunikationsmodelles von Shannon und Weaver;
- Handlungsmodelle, unter anderem die sieben Stadien des Handelns von Norman;

---

<sup>10</sup>Frei übersetzt nach [Norman, 1988, S. 24]

- Was kann der Mensch, was kann die menschliche Informationsverarbeitung leisten;
- Welche zusätzliche Information (Kontext-Information) unterstützt den Menschen bei der Gestendeutung;
- Was kennzeichnet das technische System, und im Besonderen warum wurde die Kinect® als Sensorsystem für die Gestenaufnahme ausgewählt;
- Wie lässt sich Akzeptanz beschreiben und beim Nutzer hervorrufen.



## Kapitel 3

# Gestenbeschreibung Formen und Phasenmodell

„In ihrer ursprünglichen Bedeutung bezeichnet demnach die hinweisende Gebärde schlechthin den anwesenden Gegenstand, auf den sie die Aufmerksamkeit lenkt“[Wundt, 1900, S. 158].

So wie von Wundt geschrieben, ist die Intention einer Zeigegeste, den Empfänger auf einen gegenwärtigen Gegenstand hinzuweisen. „Eine einzelne Geste kann also Kommando und Parameter gleichzeitig sein.“<sup>1</sup> Wundt spricht von einer Gebärde, Baudel und Beaudouin-Lafon [1993] und viele andere Autoren sprechen knapp 100 Jahre später von Gesten. Was unterscheidet eine Geste von einer Gebärde? Welche Bedingungen müssen erfüllt sein, damit ein Mensch und darüber hinaus auch eine Maschine die Geste erkennen kann? Mit welchen Parametern lassen sich Gesten und die Umgebung des Ausführenden beschreiben? Und vor allem, was ist unter einer „intuitiven Zeigegeste“ zu verstehen?

Auf diese Fragestellungen sollen im folgende Kapitel erste Antworten gegeben werden, damit aus der menschlichen Auffassung der Zeigegeste die technische Erkennung entwickelt werden kann.

---

<sup>1</sup> Frei übersetzt nach [Baudel und Beaudouin-Lafon, 1993, S. 28]

### 3.1 Nonverbale Ausdrucksformen, Sprachen

„Die heute bei den südlichen Völkern Europas vorkommende Zeichensprache erscheint so als ein Überlebens der in der antiken Welt überhaupt lebendigeren Begleitung der Sprache durch die Gebärde, einer Erscheinung, die in der Pflege der Pantomime und in dem großen Wert sich ausspricht, den die Alten bei der Rede auf den Gestus legten“ [Wundt, 1900, S. 148].

Wundt spricht von Gebärden bzw. von einer Zeichensprache als Überbleibsel unserer Geschichte. Auch nach Kendon [2004] „hat das Interesse an der sichtbaren Handlung als Äußerung oder Geste eine lange Tradition in der westlichen Welt. Die Wissenschaft befasst sich seit der Antike mit dem Thema, doch seit Ende des sechzehnten Jahrhunderts wird das wissenschaftliche Interesse erneut sichtbar“<sup>2</sup>.

Es gilt also, sich der Gestendefinition sowohl über die nonverbalen (nicht wörtlichen) Ausdrucksformen als auch über die lautsprachlichen Alternativen zu nähern. In den folgenden Abschnitten werden daher verschiedene Sprach- bzw. Darstellungsformen vorgestellt und die Zeigegeste wird in diesen Kontext eingeordnet, aber auch gegenüber der Gebärde abgegrenzt.

#### Körpersprache - Mimik

Der Brockhaus [1988, Bd. 14, S. 617] versteht unter der Mimik ein „Mienen- [und] Gebärdenspiel; [die] Gesamtheit der Bewegungen im Gesicht als Ausdruck von Empfindungen.“ Laut Nöth [2000, S. 308] „umfaßt [Mimik] die Ausdrucksbewegungen des Gesichts sowie die aus ihnen resultierenden Haltungen der Gesichtszüge des Menschen“. „Beim Theater wird die Mimik von der Deklamation<sup>3</sup> und den körperlichen Ausdrucksbewegungen (Gebärde, Geste) unterschieden“ [Brockhaus, 1988, Bd. 8, S. 182].

Somit muss Körpersprache mehr als nur die Mimik umfassen. Nach Nöth [2000, S. 297] wird Körpersprache vielfach als „Synonym für nonverbale Kommunikation“ verstanden. Dieses Verständnis grenzt er dann aber weiter ein, da Körpersprache nur die Bereiche der nonverbalen Kommunikation umfasst, die mit der Physiologie des Körpers, der Körperhaltung, der nichtgestischen Körperbewegungen sowie der Körperorientierung im Raum zu tun hat. Argyle und Trower [1981, S. 30] bezeichnen sie als „rätselhaftere Sprache: die Körpersprache, die wir oft unbewußt benutzen, um die Mosaiksteinchen

<sup>2</sup>Frei übersetzt nach [Kendon, 2004, S. 2]

<sup>3</sup>theatralische Überbetonung

unseres Wesens auszudrücken - unsere Empfindungen, Einstellungen und Persönlichkeitsmerkmale“.

Da die Mimik durch das für die Umsetzung der entwickelten Erkennungsstrategien gewählte technische Aufnahmesystem nicht erkannt werden kann, steht diese Quelle zusätzlicher Kontextinformation für die Interpretation der Zeigegeste nicht zur Verfügung. Die Mimik aber auch die Körpersprache grenzen sich sehr deutlich von der in dieser Arbeit betrachteten Gestik ab und werden nicht weiter analysiert.

## Gebärde

Die Gebärde hingegen kann nicht so einfach aus der Betrachtung ausgeschlossen werden. Wie bereits gezeigt, wurde um die Jahrhundertwende des 20. Jahrhunderts noch nicht von Gesten, sondern nur von Gebärden gesprochen [Wundt, 1900]. Im Brockhaus [1988, Bd. 8, S. 182] wird die Gebärde aus Sichtweise der unterschiedlichen Forschungsrichtungen und Wissenschaftsdisziplinen heraus betrachtet.

Für die weitere Diskussion in dieser Arbeit ist nur die Sichtweise der Verhaltensforschung interessant. Die „Instinkt-G[ebärde als] angeborene, artspezifische Körperbewegung und -haltung, die der innerartlichen Kommunikation dient und instinktiv auch von den Artgenossen verstanden wird. Beispiele sind die Demutsgebärde, das Drohverhalten und das Imponiergehabe“ [Brockhaus, 1988, Bd. 8, S. 182]. Nöth [2000, S. 298] versucht Geste und Gebärde deutlicher voneinander abzugrenzen. So sind die „Gesten als ‚Ausdrucksbewegungen‘ im Gegensatz zu den Gebärden als ‚Zweckbewegungen‘ stärker konventionalisiert und in ihrer Ausdrucksform festgelegt“.

Aus Sicht der Verhaltensforschung werden auch die „künstlich oder konventionalisierte[n] G[ebärden]“ gesehen, denn sie „beruhen auf einer Übereinkunft über den Sinn dieser G[ebärde] innerhalb bestimmter Gruppen oder Berufe (z.B. bei Kräftfahrern, Jägern, Seeleuten oder Soldaten)“ [Brockhaus, 1988, Bd. 8, S. 182].

Diese konventionalisierten Gebärden können schon als Gesten angesehen werden, denn gerade wenn bezogen auf technische Berufe von Gebärden gesprochen wird, ist die Aufgabe der Gebärde, ein Kommando zu übertragen. An dieser Stelle ist es schwierig, einen Unterschied zwischen Gebärden und Gesten sauber zu definieren.

Für die folgende Arbeit soll davon ausgegangen werden, dass eine Geste immer ein Kommando überträgt, wie Baudel und Beaudouin-Lafon [1993, S. 28] es beschreiben. Diese Kommandos können explizit dargeboten werden, wie es

z.B. durch die Geste „Heranwinken“ dargestellt wird, aber auch implizit, wenn wie bei der Zeigegeste nur eine zusätzliche Information übermittelt wird z. B. „dort befindet sich ...“. Implizit drückt der Ausführende aber gleichsam mit aus, dass sein Kommunikationspartner doch bitte in die entsprechende Richtung schauen möge.

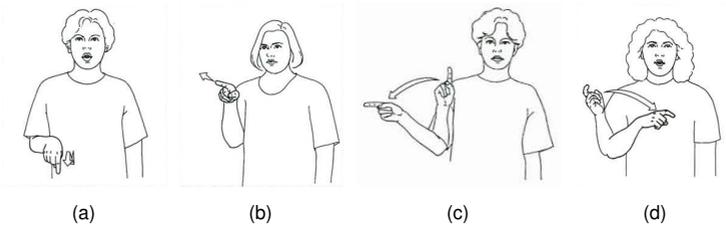
## Gebärdensprache

Auch wenn gerade der Ansatz vertreten wurde, dass bei der Gebärde die „Ausdrucksbewegung“ und nicht die „Zweckbewegung“ im Vordergrund steht, so hat sich die Gebärde in der Gebärdensprache nicht nur als Begriff für eine eigene Sprache entwickelt und erhalten. Eine Durchmischung oder gar eine Umbenennung der Gebärden- in Gestensprache ist in keiner Weise von der Gebärdensprachgemeinschaft gewünscht. Man fühlt sich als zusammengehörig und im positiven Sinne gemeint als „etwas Besonderes“.

„Man pflegt die Gebärdensprache als eine ‚Äußerung der Gedanken durch sichtbare, aber nicht hörbare Bewegungen‘ zu definieren und demnach der Gebärdenäußerung ihre Stellung mitten inne zwischen Schrift und Sprache anzuweisen. Gleich der ersteren stelle sie die Begriffe in sichtbaren Zeichen dar, während doch diese Zeichen, ähnlich den Sprachlauten, rasch vorübergehende Vorgänge seien. Sie erscheint so als eine Bilder- oder Zeichenschrift, die ihre Symbole mittels der flüchtigen Gebärde in die Luft zeichnet, statt auf ein solides, sie dauernd festhaltendes Material“ [Wundt, 1900, S. 136].

Bezogen auf gehörlose Kinder beschreibt Wundt [1900, S. 138] das Dilemma des Gehörlosen im hörenden Umfeld: „Die Unmöglichkeit, sich ihm [dem Erwachsenen] durch Laute verständlich zu machen, läßt dann von selbst zu einfachsten, zunächst fast ausschließlich hinweisenden Gebärden greifen“. Wobei er hier wieder auf die in Abschnitt 2.4 vorgestellte intuitive Funktionalität der Zeigegeste (Zeigegebärde) zurückkommt und diese bekräftigt.

Betrachtet man die Gebärden für die Ortsangaben (siehe Abb. 3.1 auf der nächsten Seite), muss im Kontext der vorliegenden Arbeit abgewogen werden, ob diese für die technische Erkennung durch das gewählte Sensorsystem geeignet sind. Die Gebärden der Gebärdensprache sind als Bestandteil einer Sprache vor allem auf den Fortgang der Kommunikation optimiert. Somit sind die einzelnen Bewegungen von der Ausdehnung ihrer Bewegungsräume sehr gering, damit der Gebärdende sich nicht sonderlich anstrengen muss. Darüber hinaus gilt „wenn man in der DGS (Deutsche Gebärdensprache) über Leute, Orte oder Gegenstände („Referenzobjekte“) spricht, die nicht in der unmittelbaren Umgebung sind, muß man diese Referenzobjekte an bestimmten Stellen



**Abbildung 3.1: Gebärden für Ortsangaben und Zeigebewegungen**

(a) Z-hier (b) Z-dort (c) Z-hier-BEWEGEN-Z-dorthin (d) Z-dort-BEWEGEN-Z-hierhin [Beecken et al., 1999]

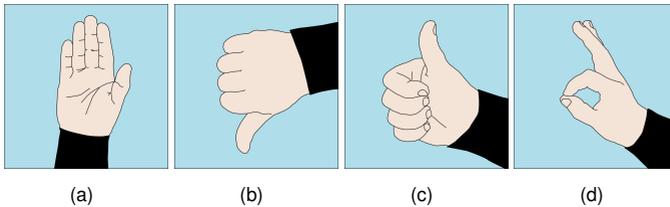
im Gebärdenraum vor sich ‚etablieren‘“ [Beecken et al., 1999, S. 90f.]. Ein Referenzobjekt, auf das in Kombination mit den vorgestellten Sprachelementen gezeigt wird, muss an dieser Stelle nicht sächlich vorhanden sein. Darüber hinaus wird eine Trennung der einzelnen Sprachelemente durch die enge, sowohl räumliche als auch zeitliche Verknüpfung, technisch eher nicht zu realisieren sein.

Grundsätzlich stellen die hier auf Abb. 3.1 gezeigten Gebärden für die meisten Menschen nicht die von ihnen intuitiv als Zeigen erkannten und ausgeführten Gesten dar, sodass sie erneut erlernt werden müssten. Für Nutzer aus der Gemeinschaft der Gebärdenden muss eine Zeigegeste sicher so erkannt werden, wie auch die entsprechende Gebärde ausgeführt wird, wobei durch die starke Formalisierung deren Erkennung mit einem entsprechend hochauflösendem technischen System sicher möglich ist.

## Zeichensprachen

Neben der der allgemeinen Gebärdensprache finden sich in verschiedenen Fachbereichen eigene Zeichensprachen, die abhängig von ihrem Verwendungszweck durch menschliche Gesten ausgeführt werden. Im Gegensatz zu der allgemeinen Deutschen Gebärdensprache (DGS) werden in diesen Zeichensprachen nur eingeschränkte Sets an Kommandos abgebildet. Die Zeichensprache bildet darüber hinaus keine Syntax ab und folgt auch keinem umfassenden sprachlichen Regelwerk, da ihre Bedeutung im Kontext und mit der Zielgruppe vor dem Gebrauch vereinbart und festgelegt ist.

Die Form der Handzeichen ist z.B. laut Arbeitsstättenverordnung §3a „eine koordinierte Bewegung und Stellung von Armen und Händen zur Anweisung



**Abbildung 3.2: Zeichen aus der Tauchersprache**

(a) Stopp (b) Gehe nach unten! (c) Gehe nach oben (d) Ok? [Southwood, 2011]

von Personen, die Tätigkeiten ausführen, die ein Risiko oder eine Gefährdung darstellen können“ [ArbStättV, 2004, S. 4].

Die für diese Arbeit im Rahmen von ersten Akzeptanzuntersuchungen ausgewählten Kommandos und ihrer gestische Repräsentation lehnte sich teilweise an die gängigen und bekannten Zeichensprachen aus den folgenden Berufsfeldern oder allgemeinen Lebensbereichen an (vergl. Kap. 3.6 auf Seite 74):

- Taucher
- Militär
- Verkehrssektor (Einweiser)
  - Flugverkehr
  - Straßenverkehr

Da die Kommunikations- und Kommandogesten aus diesen Bereichen streng formalisiert und festgelegt sind, wäre es abwegig, sie nicht auch in der Mensch-Technik-Interaktion einzusetzen. Ihre Erkennung gestaltet sich mit „pattern matching algorithms“<sup>4</sup> einfach, siehe z.B. die Taucherzeichen (siehe Abb. 3.2).

„Handzeichen sind gelebte Militärpraxis und viele Soldaten sind an ihren Einsatz gewöhnt. Die Handzeichen sind sogar schriftlich für z.B. die U.S. Armee oder die U.S. Marine dokumentiert.“<sup>5</sup> Die Zeichen, so berichtet Skinner et al. [2015] weiter, werden von Soldat zu Soldat weitergegeben. Diese Übermittlung verlangt, dass sich die Kameraden untereinander sehen können. Um mit der

<sup>4</sup>Mustererkennungsalgorithmen

<sup>5</sup>Frei übersetzt nach [Skinner et al., 2015, S. 446]

Gestekommunikation längere Strecken ohne Sichtverbindung überbrücken zu können, wird in dem von Skinner et al. [2015] vorgestellten Projekt das Handzeichen mittels Datenhandschuh erfasst und über Funkkommunikation weitergeleitet. Die Erkennung der Handzeichen bietet sich in dem beschriebenen Szenario mit der entsprechenden Hardware an, da die Aufnahmedistanz für einen optischen Sensor zu gering ist. Darüber hinaus bieten die Handzeichen durch die Standardisierung und den Bekanntheitsgrad eine geeignete, gut definierte Schnittstelle.

Für eine weitere Betrachtung im Kontext dieser Arbeit sind diese stark formalisierten Zeichensprachen für die Analyse der Zeigegesten nicht nutzbar, da der verwendete Gestensensor die Hand bzw. die einzelnen Finger nicht ausreichend abbildet, oder die Distanz Sensor-Hand sehr gering sein muss.

Weitere gebräuchliche Zeichensprachen finden sich z.B. im Straßenverkehr oder der Seefahrt. Die Zeichen in Form von Verkehrsschildern oder Flaggen sind mit den in dieser Arbeit betrachtenden Zeichensprachen nicht vergleichbar. Eine Geste aus dem Straßenverkehr soll aber für die Experimente als Kontrollgeste verwendet werden, „das Winken“.

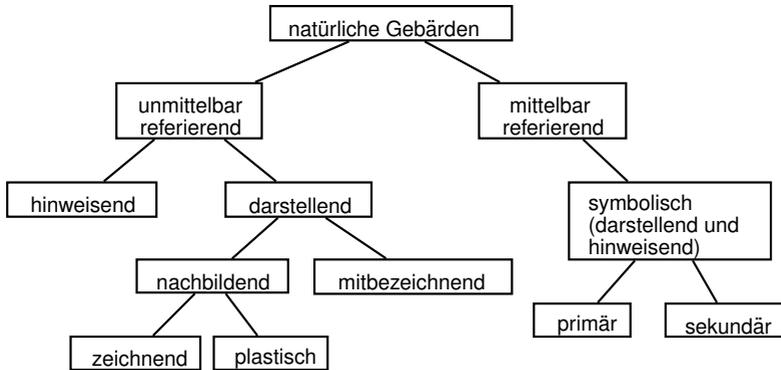
## 3.2 Klassifizierung von Gesten

Nachdem in Kapitel 2.1 auf Seite 22 die Geste als „zielgerichtete Ausdrucksbewegung des Körpers, bes[onders] der Hände“ [Brockhaus, 1988, Bd. 8, S. 437] eingeführt worden ist, ist sie in den vorherigen Abschnitten von anderen, der nonverbalen Kommunikation zugeordneten Begriffen abgegrenzt worden. Darüber hinaus wurden verschiedene Sprachformen erläutert, deren Informationsübermittlung mit den Händen durchgeführt wird und als Beschreibungsform der in der Arbeit ge- und untersuchten intuitiven Zeigegeste ausgeschlossen.

Es stellt sich die Frage nach der Unterscheidung unterschiedlichster intuitiv ausgeführter Gesten, denn Intuition kann ein breites Spektrum annehmen. Dazu bietet die Arbeit von Fricke [2007] einen Ansatz. Sie untersucht verschiedene Varianten zur Klassifikation von Gesten von denen einige auszugsweise hier vorgestellt werden.

### Wundt

Obwohl im vorherigen Kapitel die Geste von der Gebärde abgegrenzt wurde, war diese Abgrenzung zur Zeit von Wundt [1900] noch nicht so exakt definiert. Gab es für ihn nur Gebärden, so konnten diese verschiedenartige Bedeutungen



**Abbildung 3.3: Klassifikation natürlicher Gebärden nach Wundt**

Darstellung nach [Fricke, 2007]

und Beziehungen aufweisen. Für die Fragestellung der Ortsbeschreibenden Zeigegebärden sind nur die hinweisenden, unmittelbar referenzierenden, natürlichen Gebärden interessant (siehe Abb. 3.3). Laut Fricke [2007, S. 158] „unterscheidet Wundt [innerhalb der Gruppe der hinweisenden Gesten] implizit zwischen einem Zeigen auf Objekte, die selbst keine Zeichen sind, und einem Zeigen auf anwesende Objekte, die als Zeichen für etwas anderes stehen.“

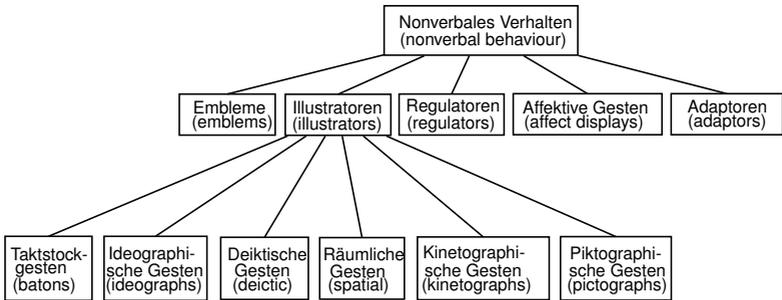
Diese detaillierte Abgrenzung der Zeigegebärde muss im Rahmen der vorliegenden Ausarbeitung nicht getroffen werden, da der wichtigste Aspekt des Zeigens für die technische Umsetzung die Ortsbestimmung von Objekten ist.

## Ekman

Ekman und Friesen [1969] stellt mit ihrer Arbeit zum „Nonverbalen Verhalten“ eine weitere Klassifikation bereit (siehe Abb. 3.4 auf der nächsten Seite). Das Nonverbale Verhalten umfasst aber mehr als die Gesten und Gebärden, auf die sich hier beschränkt werden soll. So werden die Begriffe:

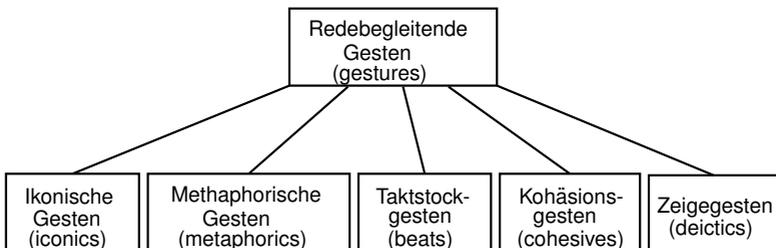
- Gebrauch „Usage“,
- Ursprung (Quelle) „Origin“ und
- Kodierung „coding“

für das nonverbale Verhalten definiert. Diese Begriffe erinnern an das Kommunikationsmodell von Shannon, das über das Zusammenspiel der verschiedenen



**Abbildung 3.4: Klassifikation redebegleitender Gesten nach Ekman/Friesen**

Darstellung nach [Fricke, 2007]



**Abbildung 3.5: Klassifikation redebegleitender Gesten nach McNeill**

Darstellung nach [Fricke, 2007]

Komponenten der Kommunikation berichtet. Für Ekman und Friesen [1969] ist „der Code [...] die Regel welche die Beziehung zwischen der Aussage selbst und dem, was sie signifikant macht, sei es intrinsisch<sup>6</sup> oder extrinsisch<sup>7</sup>, motiviert.“<sup>8</sup> Dieses „Coding“ wird von ihm weiter untersetzt und zu einer Klassifikation ausgeweitet. „Der deiktische Illustrator ist eine zeigende Bewegung und somit intrinsisch kodiert“<sup>9</sup>.

<sup>6</sup>von innen her kommend, aus eigenem Antrieb

<sup>7</sup>von außen motiviert

<sup>8</sup>Frei übersetzt nach [Ekman und Friesen, 1969, S. 60f.]

<sup>9</sup>Frei übersetzt nach [Ekman und Friesen, 1969, S. 69f.]

## McNeill

McNeill [1992] vereinfacht die Klassifikation der redegleitenden Gesten auf eine Ebene (siehe Abb. 3.5 auf der vorherigen Seite). Für ihn gibt es fünf Klassen, die Ikonischen, die Metaphorischen, die Taktstock-, die Kohäsions- und die Zeigegesten. McNeill [1992] gibt in seiner Übersicht alle fünf Kategorien an, schreibt aber selbst, dass „in seinem Ansatz [...] Gestenbewegungen nur in vier Hauptkategorien unterteilt [werden]: iconic, metaphoric, deictic (pointing) und beat (Taktstock-) Gesten“<sup>10</sup>.

Im Bezug auf technische Umsetzung einer Gestenerkennung wurden die vier von McNeill beschriebenen Gesten auf ihre Tauglichkeit untersucht:

- Die Taktstockgesten schließen bereits die Definition von McNeill [1992] selbst aus. „Beats definieren eine Bewegung ohne erkennbar Bedeutung“<sup>11</sup>. Wenn es für den Menschen mit all seinen kognitiven Fähigkeiten und seiner Erfahrung schwierig bis unmöglich erscheint, den Inhalt von Taktstockgesten zu deuten, wie soll diese Gestenform dann dem technischen System beigebracht werden?
- Die ikonische Geste hat nach McNeill [1992] „eine formale Beziehung zu dem semantischen Inhalt der Rede. Sie symbolisiert die Bedeutung, für die sie steht und damit gleich die Bedeutung der Rede“<sup>12</sup>. Eine trainierte Zeichensprache fällt in diese Kategorie und ist in dieser Arbeit für eine intuitiv nutzbare Geste bereits abgelehnt worden.
- Vergleichbar sind die Probleme für die technische Interpretation der metaphorischen Gesten. Sie definieren das Kommando (die Textbedeutung) nicht durch Symbole, sondern durch abstrakte Metaphern. Während die ikonischen Gesten erlernt werden könnten, und damit die Gestenerkennung deutlich vereinfacht würde, ist für das Verständnis von Metaphern abstraktes Denkvermögen entscheidend. Weicht man diese Definition auf und ergänzt die Methaper mit erlernten Elementen, so verliert sie ihren abstrakten Charakter.

Andererseits haben sich für Handgesten auf mobilen Eingabegeräten Metaphern, abstrakte Aktionen als sehr verständlich und gut reproduzierbar erwiesen, z.B. das Zoomen mit zwei Fingern oder das Umblättern durch Wischen als ob man ein Buch umblättert. Le Hong und Biesterfeldt

<sup>10</sup>Frei übersetzt nach [McNeill, 1992, S. 76]

<sup>11</sup>Frei übersetzt nach [McNeill, 1992, S. 80]

<sup>12</sup>Frei übersetzt nach [McNeill, 1992, S. 78]

[2010] führten eine Studie zu kulturellen Unterschieden und Gemeinsamkeiten gestenbasierter Bedienung von Multitouch-Oberflächen, wie sie Smartphones oder Tablets haben, durch. Dabei zeigte sich, dass das Kommando „Zoom“ auch international am häufigsten durch „Finger über dem Objekt spreizen“ dargestellt wird. Während bei Le Hong und Biesterfeldt [2010] die Geste „Nach links Wischen“ für das Kommando „Rückgängig“ im Vergleich am besten abschnitt (passives Anwendungswissen), konnte für das Kommando „Rückgängig machen“ diese Geste nicht bestätigt werden (aktives Gestaltungswissen). Bei Steinhorst et al. [2016] schneidet die Geste „Wischen nach rechts“ für das Kommando „Weiter“ allerdings schlecht ab. Er regt an, die abstrakte Metapher „Wischen“ im Sinne der Selbstbeschreibungsfähigkeit in Zukunft noch zu verbessern.

Trotz der hier exemplarisch gezeigten Studienergebnisse haben sich für diese Gesten mobile Multi-Touch-Umgebungen durchgesetzt. Weyerstall und Saul [2013] haben die Multi-Touch-Geste „Wischen“ einer gleichartigen berührungslosen Geste gegenübergestellt. Dabei stellte sich heraus, dass 58 % der Probanden die „Wischen-Geste“ bei der Multitouch-Bedienung intuitiv empfanden.

- Bleiben für die Erkennung von Freiraumgesten im ersten Schritt die als deiktisch bezeichneten Zeigegesten mit dem besonderen Bezug zur Umgebung. Mit der Zeigegeste adressiert der Mensch vor allem Gegenstände (Objekte), die ausgewählt werden sollen.

## Freeman

Eine ganz andere Gestenklassifizierung vertreten Freeman und Roth [1994]. Während alle vorgestellten Klassifikationen davon ausgehen, dass die Geste eine dynamische Bewegung ist, die Ihre Information in der Art der Bewegung kodiert, unterscheidet Freeman nur zwei Kategorien von (Hand-)Gesten. Zum einen statische Gesten, Handposen, die in einem Bild darstellbar sind, zum anderen dynamische Gesten, Bewegungsgesten, die in einer Abfolge von Bildern zu erkennen sind.<sup>13</sup> Wie aus der Erwähnung des notwendigen Umfangs an Bildern in der Unterscheidung ersichtlich, ist für Freeman die Frage der Möglichkeit zur technischen Auswertung entscheidend, weniger die Frage nach dem grundlegenden Gestenverständnis. Diese Fragestellung wird in dem Artikel durch die Information, dass der Algorithmus für die Gestenerkennung

<sup>13</sup>Frei übersetzt nach [Freeman und Roth, 1994, S. 1]

durch einen Computer schnell zu berechnen sein muss, verstärkt [Freeman und Roth, 1994]. Sie schlussfolgern daraus, dass ein einfacher Mustererkenner für Handgesten zu einem schnellen und damit nutzbaren System führt.

Dieser generellen Einteilung kann sich der Autor anschließen, da auch bzw. gerade für die Zeigegeste zu klären ist, ob „Zeigen“ als statische oder dynamische Geste aufzufassen ist.

### 3.3 Zeigegeste

Im Anschluss an die vorgestellten Klassifikationen zur Geste allgemein mit dem Verweis auf die deiktischen Gesten im Besonderen muss abschließend geklärt werden, was im Verlauf dieser Arbeit unter eine Zeigegeste zu verstehen ist.

„Eine einfache Zeigegeste zeigt entweder eine Richtung an oder sie lokalisiert eine bestimmte Entität<sup>14</sup> in Bezug auf einen Raumpunkt, der sich aus der Verlängerung der durch die Origo<sup>15</sup> und den Arm bzw. die Hand markierten Geraden ergibt“ [Fricke, 2007, S. 59].

Die einfache Zeigegeste, wie sie von Fricke beschrieben wird, entspricht der Definition, die im Kontext dieser Arbeit immer wieder zum Tragen kommt. Aber anders als in ihrer und den Arbeiten von McNeill und Kendon soll die technisch zu erkennende Zeigegeste nicht redebegleitend eingesetzt und interpretiert werden. „Redebegleitende Gesten sind spontane Bewegungen, insbesondere von Händen und Armen, die zu beobachten sind, wenn jemand spricht. Da sie nur unter Bezugnahme auf die begleitete Rede interpretierbar sind, werden sie als eng mit der Aktivität des Sprechens verbunden betrachtet“ [Fricke, 2007, S. 143]. Von dieser Definition gilt es sich vollständig zu lösen, denn wie Fricke die Geste als spontane Bewegung beschreibt, ist es für ein technisches System schwer, diese zu identifizieren und auszuwerten. Für die technische Gestenerkennung muss die Bewegung einer festen Regel entsprechen, die allerdings intuitiv vom Menschen ausgeführt wird.

Auch Fricke beschreibt diese festen Regeln. „Nehmen wir die typische Geste mit ausgestrecktem Arm und ausgestrecktem Zeigefinger. Dadurch, dass hier eine konventionalisierte Form vorliegt, [...] wird [sie] auch ohne begleitende Rede als hinweisende Geste verstanden. Andererseits ist [...] eine derart starke Konventionalisierung der Formseite nicht notwendig. Denn auch in abweichender Ausführung, beispielsweise mit einem anderen Finger oder mit der flachen Hand, ist die [...] Funktion gewährleistet“ [Fricke, 2007, S. 151].

<sup>14</sup>einen konkreten oder abstrakten Gegenstand

<sup>15</sup>„Deiktisches Zentrum“, ich, hier, jetzt

Die in dieser Arbeit vorgestellte technische Beschreibung der Zeigegeste orientiert sich an dieser Aussage von Fricke. So wird zunächst eine sehr stark konventionalisierte Beschreibung erarbeitet, die immer weiter vereinfacht und somit für die Ausführenden einfacher nutzbar wird. In der Arbeit von Baudel und Beaudouin-Lafon [1993] ist die Erkennung von Freiraumgesten technisch mit einem Datenhandschuh umgesetzt. Mit diesem kann Kontextinformation z.B. die Orientierung der Hand ausgewertet werden. „Das Kommando wird erkannt an der Ausrichtung der Handfläche (Daumen nach unten) sowie daran, dass alle Finger gestreckt“<sup>16</sup> sind. Diese Kontextinformation liegt aufgrund des vom Autor gewählten Gestensensors nicht vor und kann somit für die Definition der Zeigegeste nicht genutzt werden.

### 3.4 Gestenraum

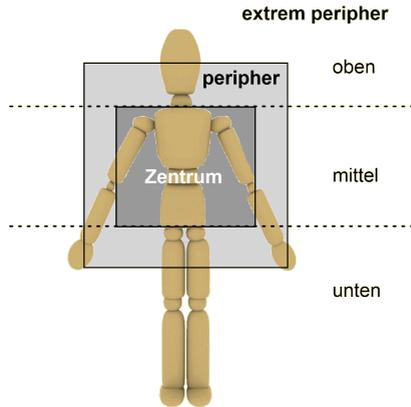
Neben den Fragen nach dem „Was“ und „Wie“ einer Zeigegeste ist auch die Frage nach dem „Wo“ zu stellen. Die Antwort auf diese Frage ist der „Gestenraum“.

Während Baudel und Beaudouin-Lafon [1993] von der „active zone“ sprechen, damit aber den Zielbereich der Zeigegeste meinen, muss für die vorliegende Arbeit eher der Handlungsraum des Menschen in Betracht gezogen werden. Dazu schreibt Herczeg [2009], der „Mensch sitzt oder steht am Ort seines Handelns und besitzt in dieser Lage einen begrenzten Handlungsraum, unmittelbar erkennbar durch den sogenannten Greifraum seiner Hände. In diesem Greif- und Handlungsraum können sich auch ganz andere Lösungen manifestieren.“ Denkt man ausgehend von Maus und Tastatur über „die Gestaltung und Nutzung des menschlichen Greifraumes neu, so kommen vielfältige anderen Lösungen zum Vorschein, die physikalische Gegenständlichkeit und computergestützte Interaktivität besitzen“, diese Gegenständlichkeit im Falle der Freiraumgeste aber nicht haben müssen. „Andere Geräte beginnen gerade den menschlichen Greif-, Handlungs- und Kommunikationsraum zu erobern“ [Herczeg, 2009, S. 104].

Aus Beobachtungen zur redebegleitenden Geste hat McNeill [1992] den Gestenraum in drei Bereiche unterteilt (siehe Abb. 3.6 auf der nächsten Seite). Der primär genutzte Raum ist zentral mittig vor dem Körper. Dieser schließt sich auf Grund des gewählten Erkennungssystems für die Gestenerkennung zunächst aus. Nach Analyse der Versuchsdaten stellte sich aber heraus, dass diese Einschränkung mit Verwendung der Kinect<sup>®</sup> 2 gegenüber anderen Pro-

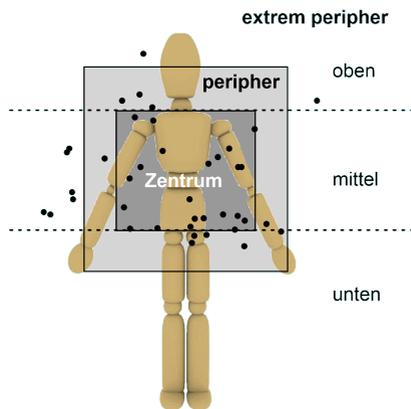
---

<sup>16</sup>Frei übersetzt nach [Baudel und Beaudouin-Lafon, 1993, S. 32]



**Abbildung 3.6: Typischer Gestenraum eines erwachsenen Sprechers**

Angepasste Darstellung nach [McNeill, 1992]



**Abbildung 3.7: Nutzung des Gestenraumes, auch des extrem entfernten, für deiktische Gesten**

Die Punkte entsprechen den Orten der Gestenausführung verschiedener Sprecher.

Angepasste Darstellung nach [McNeill, 1992]

bleme in den Hintergrund treten kann. Darüber hinaus definiert McNeill einen „peripheren“ und einen „extrem peripheren“ Raum.

Unterscheidet man bei der Analyse des genutzten Gestenraumes zwischen den einzelnen Gestenkategorien und betrachtet nur die deiktischen Gesten, so fällt auf, dass weder Gesten im unteren Bereich noch besonders viele im zentralen Bereich ausgeführt werden. Der genutzte Gestenraum beschränkt sich nach Untersuchungen von [McNeill, 1992] überwiegend auf den peripheren und extrem peripheren Raum (siehe Abb. 3.7 auf der vorherigen Seite).

$$SH_m = 1450 \text{ mm} \quad [\text{DIN 33402-2, 2005, S. 11}]$$

$$HH_m = 765 \text{ mm} \quad [\text{DIN 33402-2, 2005, S. 25}]$$

$$B_m = 1045 \text{ mm} \quad [\text{DIN 33402-2, 2005, S. 37}]$$

$$\phi_{GZm} = \arccos \left( \frac{SH_m - B_m}{SH_m - HH_m} \right) \quad (3.1)$$

$$\phi_{GZm} = \arccos \left( \frac{1450 - 1045}{1450 - 765} \right)$$

$$\phi_{GZm} = 53,7^\circ$$

#### **minimaler Schulterwinkel des Menschen für den zum Zeigen genutzten Raum (männlich) ( $\phi_{GZm}$ )**

mit durchschnittliche Schulterhöhe (männlich) ( $SH_m$ ), durchschnittliche Höhe der Hand (Griffachse) über der Standfläche (männlich) ( $HH_m$ ), durchschnittliche Gesäß-Beinlänge (männlich) ( $B_m$ )

Kombiniert man die Erkenntnis, dass sich Zeigen nur oberhalb der Hüfte abspielt, mit den Körpermaßen des Menschen (laut [DIN 33402-2, 2005]), so kann man den minimalen Schulterwinkel des Gestenraumes bestimmen. Dieser beträgt ungefähr  $56^\circ$  für Frauen (siehe Formel 3.2) und ca.  $54^\circ$  für Männer (siehe Formel 3.1).

$$SH_w = 1345 \text{ mm} \quad [\text{DIN 33402-2, 2005, S. 11}]$$

$$HH_w = 715 \text{ mm} \quad [\text{DIN 33402-2, 2005, S. 25}]$$

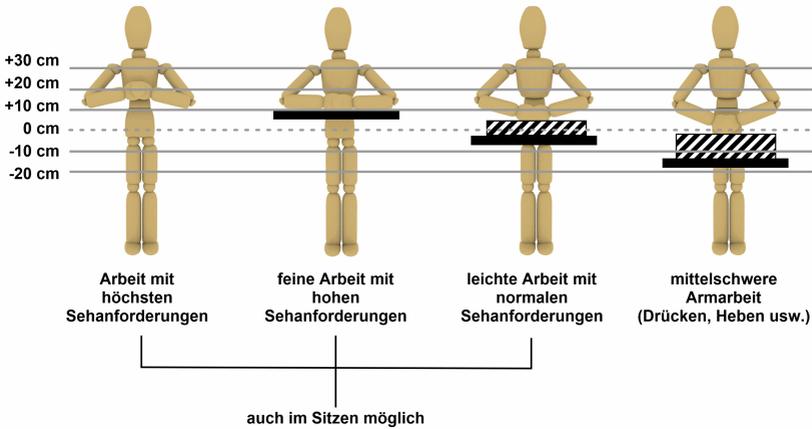
$$B_w = 990 \text{ mm} \quad [\text{DIN 33402-2, 2005, S. 37}]$$

$$\phi_{GZw} = \arccos \left( \frac{1345 - 990}{1345 - 715} \right) \quad (3.2)$$

$$\phi_{GZw} = 55,7^\circ$$

#### **minimaler Schulterwinkel des Menschen für den zum Zeigen genutzten Raum (weiblich) ( $\phi_{GZw}$ )**

mit durchschnittliche Schulterhöhe (weiblich) ( $SH_w$ ), durchschnittliche Höhe der Hand (Griffachse) über der Standfläche (weiblich) ( $HH_w$ ), durchschnittliche Gesäß-Beinlänge (weiblich) ( $B_w$ )



**Abbildung 3.8: Tätigkeitsabhängige Arbeitshöhe**

Modifizierte Darstellung nach [Schmauder und Spanner-Ulmer, 2014]

Der hier berechnete Schulterwinkel gilt als maximaler Wert für die Definition der Handhaltung während keine Geste, und somit auch keine Zeigegeste, ausgeführt wird. Bedenkt man weiter, dass diese Einschränkung nur für die deiktischen Gesten zutrifft, sollte der maximal mögliche Winkel nicht ausgeschöpft werden, weshalb in dieser Arbeit von einem Schulterwinkel von ca. 20° ausgegangen wird.

Der „Wirkraum des Hand-Arm-Systems kennzeichnet den Raumsektor, in dem der Mensch bei unbewegtem Oberkörper berühren, greifen und bewegen kann“ [Schmauder und Spanner-Ulmer, 2014, S. 117], Man erkennt, dass dieser, dem von McNeill bezeichneten Gestenraum entspricht. Schmauder bezeichnet die Räume nach Hettinger & Wobbe (1993) als „physiologisch kleiner Greifraum“, „physiologisch großer Greifraum“ und „Anatomisch maximaler Greifraum.“ Betrachtet man darüber hinaus die tätigkeitsabhängigen Arbeitshöhen (siehe Abb. 3.8), so fällt auf, dass die Arbeitshöhe mit der Arbeitsschwere immer weiter absinkt. Somit liegt für Tätigkeiten im Bereich der Lastenhandhabung, wie im Fall der Hebe- und Tragehilfe keine Überschneidung des Gesten- mit dem Wirkraum vor.

### 3.5 Phasenmodelle der Geste

Auch wenn die meisten vorgestellten Autoren von einer Unterscheidung in statische und dynamische Gesten absehen und für die gesamte Geste sogar widersprechen, so sind sie sich einig, dass eine Geste verschiedene Phasen durchläuft. Nach Baudel und Beaudouin-Lafon [1993] ist „jedes Gestenkommando durch eine Startposition, eine dynamische Phase sowie eine Endposition beschrieben.“<sup>17</sup> Sie gehen davon aus, dass „eine Geste von Natur aus kontinuierlich ist. Ein System, das die Geste interpretieren und in Kommandos übersetzen will, muss die kontinuierliche Bewegung in Segmente, diskrete Zeitabschnitte, zerlegen“<sup>18</sup>.



**Abbildung 3.9: Die drei Phasen sportlicher Bewegungsakte und ihre Relationen**

[Meinel und Schnabel, 2015]

So allgemein wie Baudel und Beaudouin-Lafon [1993] und Wilson et al. [1996] die Geste als dreigeteilte Bewegung darstellen, so begründet sich diese Dreiteilung nach Meinel und Schnabel [2015] verallgemeinert im „Bewegungsakt“. Genauer untersucht wurden bei Meinel die Beziehungen der Bewegungsakte „Vorbereitungsphase“, „Hauptphase“ und „Endphase“ (siehe Abb. 3.9). In dieser Hinsicht gilt es, für die weitere Entwicklung einer Gestenerkennung zu bedenken, dass jede Phase eine Ergebnisbeziehung zur vorhergehenden hat. Dieser Ansatz gilt im besonderen für die erarbeiteten „Phasendefinition mittels zweier Winkel“ und „Phasendefinition mittels eines Winkels“. Meinel und Schnabel [2015, S. 81] schreibt für die Endphase verallgemeinert: „Die [...] Funktion der Endphase ergibt sich aus dem Tatbestand, dass die Hauptphase meist zu einem Zeitpunkt beendet ist, zu dem sich der Körper noch in [...] einem labilen Gleichgewichtszustand befindet. Die Endphase führt im Ergebnis zu einem statischen Zustand.“

In dem von Kendon [2004] beschriebenen Phasenmodell (siehe Abb. 2.10 auf Seite 36) sind nur die Bewegungen, also die dynamischen Phasen einer

<sup>17</sup>Frei übersetzt nach [Baudel und Beaudouin-Lafon, 1993, S. 32]

<sup>18</sup>Frei übersetzt nach [Baudel und Beaudouin-Lafon, 1993, S. 29]

Geste berücksichtigt. Bedenkt man, dass die Arbeit sich auf die natürliche, redebegleitende Geste bezieht, lässt sich dieses Verständnis gut nachvollziehen. McNeill [1992] hat seine Definition einer redebegleitende Geste erweitert, da die Geste mit der Rede synchronisiert werden muss (siehe Abb. 3.10). Diese Synchronisierung übernehmen in der Definition von McNeill die optionalen, zusätzlichen Haltephasen (pre- bzw. post-stroke hold).

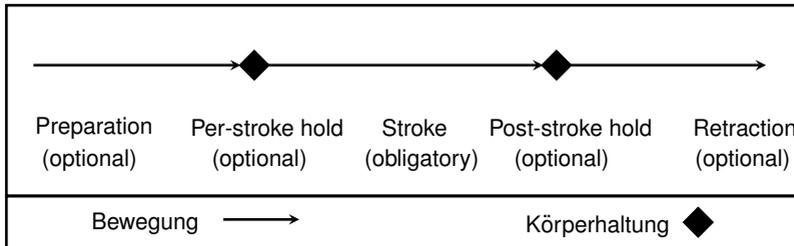


Abbildung 3.10: Phasenmodell nach [McNeill, 1992]

Betrachten wir alle Phasen in der Definition von [McNeill, 1992] im Einzelnen:

1. **Preparation, Vorbereitung** (optional), in der die Hand aus der Ruheposition zu einer Position im Gestenraum bewegt wird, an der die Hauptphase beginnen kann. Die Vorbereitungsphase korrespondiert mit den sprachlichen Teilen, die die Bedeutung der Geste ausdrücken.
2. **Pre-stroke hold** (optional) ist die Position und Handhaltung am Ende der Vorbereitung. Diese kann kurz gehalten werden, bis die Hauptphase beginnt. Ein pre-stroke hold tritt auf, wenn die Hauptphase aus irgendeinem Grund verzögert ist.
3. **Stroke, Hauptphase** (obligatorisch), ist der Höhepunkt der Gestenausführung. Es ist die Phase, welche die Bedeutung der Geste ausdrückt. Die Hauptphase ist synchronisiert mit der entsprechenden sprachlichen Information. Sie wird üblicherweise im zentralen Gestenraum ausgeführt. Der Gestenraum ist grob durch Taille, Schultern und Arme begrenzt, der Kopf wird gelegentlich mit in die Gestenausführung einbezogen.
4. **Post-stroke hold** (optional) ist die Endposition bzw. Haltung, welche eingenommen wird, wenn die Hauptphase abgeschlossen ist. Diese

kann mehr oder weniger kurz gehalten werden, bis der Rückzug beginnt. Ein post-stroke hold tritt auf, wenn die Synchronität mit der Sprache nicht gegeben ist.

5. **Retraction**, *Rückzug* (optional) ist die Rückkehr der Hand in eine Ruheposition.<sup>19</sup>

## 3.6 Nutzerverständnis

Zum Thema Nutzerverständnis gilt es grundsätzlich zwei Fragen zu besprechen:

- Wer ist der Nutzer eines technischen Systems? Also auf wen muss das System reagieren?
- Wie versteht bzw. interpretiert der Nutzer die Kommandos, die er dem technischen System übermitteln will? Welche Geste führt er aus, um das Kommando auszudrücken?

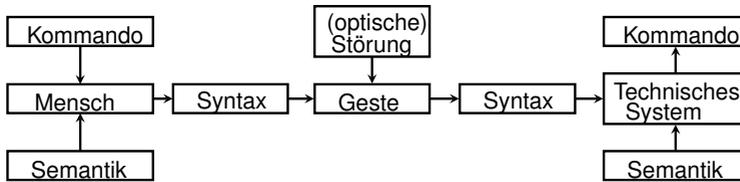
### Ausführender

Die Frage nach dem „Wer“ ist verhältnismäßig einfach. Johannsen [1993, S. 6] folgend zu beantworten. „Bei den Arbeitstätigkeiten des Menschen in Mensch-Maschine-Systemen denken wir meistens zunächst an die Tätigkeiten des Bedieners. Das ist derjenige Mensch, der [...] über Schnittstellensysteme in einem Wirkungsgefüge mit dem technischen System arbeitet“. Johannsen [1993] führt diesen Aspekt weiter aus und unterscheidet „Bediener“ und „Benutzer“ anhand des für die Nutzung notwendigen Wissens. Für die Mensch-Technik-Interaktion ist diese Frage aber auch praxisrelevant, wenn auch aus Sichtweise des technischen Systemes. Wenn das technische System mehrere Menschen erkennen kann, wer ist dann der „Ausführende“, der das System aktuell kommandieren darf?

Eine Antwort auf diese Fragestellung wird für die Übertragung auf ein Zusammenführung von Online- / Offline-Version (vergl. Kap. 6.6 auf Seite 132) näher betrachtet. Für die Parametrisierung der Zeigegeste sowie des Experimentes in dieser Arbeit wird davon ausgegangen bzw. dafür gesorgt, dass sich in der Erkennungszeit nur ein Nutzer im Sichtbereich des Gestensensors aufhält.

---

<sup>19</sup>Frei übersetzt nach [McNeill, 1992, S. 83]



**Abbildung 3.11: Modell der Gestenkommunikation**

für die Mensch-Technik-Interaktion, Weiterentwicklung nach [Shannon und Weaver, 1949]

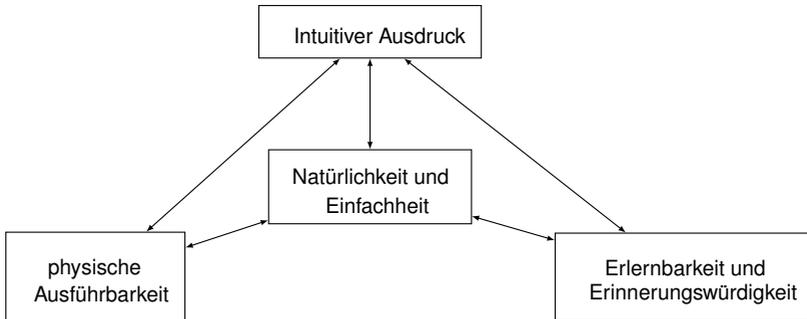
## Intuitives Gestenverständnis

„Ein technisches System ist im Rahmen des spezifischen Kontextes einer Aufgabenstellung in dem Maße intuitiv benutzbar, in dem der jeweilige Benutzer mit ihm durch unbewusste Anwendung von Vorwissen effektiv interagieren kann“ [Mohs et al., 2006, S. 80].

Nach dieser Definition von „intuitiv“ ist eine „intuitiv auszuführende Geste“ eine Geste, die nur unter unbewusster Anwendung von Vorwissen ausgeführt wird, sofern dem Nutzer nur das darzustellende Kommando genannt wird. Der Nutzer wird versuchen für das Kommando eine Geste zu benutzen, von der er ausgeht, dass das technische System diese so versteht, wie ein anderer Mensch sie verstehen würde (siehe Abb. 3.11). Nach Wundt [1900, S. 179f.] birgt dieser Ansatz aber auch Gefahren, „denn groß und klein werden zwar durch dieselben Bewegungen ausgedrückt wie oben und unten, aber diese Bewegungen beziehen sich nicht bloß auf die räumlichen Richtungen, sondern mehr noch auf das, was sich in ihnen ausdehnt.“ Wie dieser Aussage zu entnehmen ist, ist eine nicht formalisierte Ausführung einer Geste für den Kommunikationspartner schwer zu verstehen. Dieser Einstellung widerspricht McNeill [1992]: „Gesten sind spontane Kreationen des einzelnen Redners, einmalig und persönlich. Sie folgen generellen Prinzipien [...] aber in keiner Weise sind sie Elemente eines vorbestimmten Repertoires“<sup>20</sup>

Für die Entwicklung eines Algorithmus zur Erkennung einer Zeigegeste muss für eine Zeigegeste eine eben so stark formalisierte Gestensyntax gefunden werden, welche intuitiv vom Menschen entsprechend ausgeführt wird, sich dennoch gut auf Basis des gewählten Erkennungssystems interpretieren lässt. Dazu wurden sechs Kommandos ausgewählt und für jedes Kommando min-

<sup>20</sup>Frei übersetzt nach [McNeill, 1992, S. 1]



**Abbildung 3.12: Einflussfaktoren zur Entwicklung intuitiver Freiraumgesten**

destens eine dynamische und eine statisch Gestenvariante entwickelt (siehe Abb. 3.12).

Das Untersuchungskonzept sah vor, je drei als statisch bzw. dynamisch eingeschätzte Kommandos auszuwählen und zu untersuchen. Bei den Kommandos handelt es sich um:

- **Herkommen:** zwei dynamische und zwei statische Varianten
- **Heben:** zwei dynamische und zwei statische Varianten
- **Drehen:** eine dynamische und eine statische Variante
- **Anmelden am System:** zwei dynamische und zwei statische Varianten
- **Stopp:** zwei dynamische und zwei statische Varianten
- **Auswahl eines Objektes (Zeigen):** eine dynamische und eine statische Variante

Für jede Variante wurde eine Videoaufnahme angefertigt. Die Videoaufnahmen wurden zusammen mit einem online-Fragebogen zur Bestimmung der Intuitivität über das Portal UNIPARK im Internet zur Verfügung gestellt. Im Versuchszeitraum von sechs Wochen nahmen insgesamt 150 Personen an der Befragung teil, nur 90 beantworteten alle Fragen. Nur deren Fragebögen konnten ausgewertet werden. Es beteiligten sich Personen im Alter bis 60 Jahre, der Schwerpunkt liegt in der Altersgruppe von 21 - 30 Jahren (Studierende).

Für die Auswertung wurden die Items „Bekanntheit einer Geste als Kommunikationsmittel“, „Natürlichkeit einer Geste“, „Verwendung der Geste als Kommunikationsmittel bei Abwesenheit der Sprache“ sowie „Nachvollziehbarkeit des Zusammenhangs zwischen Kommando und Geste“ zu einem Intuitivitätswert zusammengeführt und mit den Aussagen zum allgemeinen Nutzerverständnis der Gestenausführung (statisch / dynamisch) und der Rangordnung der Bewertung der Videos zusammengefasst (siehe Tab. 3.1).

Tabelle 3.1: Auswertung der Untersuchung zum Intuitiven Gestenverständnis  
dt. nach [Nowack et al., 2015b]

	Dynamische Kommandos			Statische Kommandos		
	Herkommen	Heben	Drehen	Anmelden	Stopp	Auswählen
Platz 1 der Nutzerbewertung	dynamisch 79%	dynamisch 37%	dynamisch 72%	dynamisch 50%	statisch 91%	statisch 79%
Nutzerbewertung Intuitivitätswert	dynamisch (1,4)	statisch / statisch (2,0)	dynamisch (2,2)	dynamisch (1,4)	statisch (1,2)	statisch (1,3)
Nutzerverständnis des Kommandos	dynamisch	dynamisch	dynamisch	dynamisch	statisch	dynamisch

Für die weitere Betrachtung von Freiraumgesten wurden nach dieser Studie der Schwerpunkt auf das „statische Kommando“ Zeigen (Auswählen) gelegt. Das Kommando wird intuitiv als „statisch“ angesehen. Auf den ersten Platz bei der Videobewertung wurde eine statische Variante von den Probanden gewählt. Teilweise unterschieden sich die Videos zwischen der als statischen und dynamisch definierten Aufzeichnung nur in der Haltezeit, jedoch nicht in dem grundsätzlichen Bewegungsablauf. Auf die Frage nach ihrem Verständnis der Gestenausführung gaben die Probanden an, dass es sich eher um eine dynamische Ausführung handelt (vergl. Nowack et al. [2015b]).

Abgeleitet aus diesen Ergebnissen fielen viele basale Entscheidung zu den verschiedenen Varianten der Erkennung des Zeigens, wie sie im folgenden Kapitel vorgestellt werden.

Sollte neben dem „statischen Kommando“ „Zeigen“ auch zukünftig ein dynamisches untersucht werden, bietet sich das Kommando „Herkommen“ (Heranwinken) an, da es neben dem Zeigen die aussagekräftigsten Bewertungen erhalten hat. Die Geste „Heranwinken“ wird daher auch im Rahmen der weiteren Experimente als Kontrollgeste verwendet.

## Kapitel 4

# Parametrisierung der Zeigegeste

Im folgenden Kapitel werden basierend auf den vorgestellten theoretischen Ansätzen drei Varianten der mathematischen Beschreibung einer Zeigegeste vorgestellt. Wie in Kapitel 3 auf Seite 55 beschrieben, lässt sich die Geste nach Kendon durch drei Bewegungsphasen und zwei Warteposen definieren (siehe Abb. 3.10 auf Seite 72). Die hier vorgestellten mathematischen Beschreibungen basieren auf der Beschreibung des Zeigevektors in der Haltephase am Ende der Gestenausführung (post-stroke-hold) als „statische Beschreibung“ oder auf der Abfolge einzelner Bewegungen im Gesamtzyklus als „Phasendefinition“.

Für eine statische Beschreibung sprechen zwei Gründe. Zum einen ergibt sich diese aus dem Nutzerverständnis; zum anderen lässt sie sich mathematisch-technisch genau definieren, wie im Folgenden detailliert beschrieben wird. Von dieser Beschreibung ausgehend sind die dynamischen Definitionen (Phasendefinitionen) entstanden, da eine Geste allgemein als eine Bewegungsabfolge angesehen wird. In der vorgestellten Analyse zum Nutzerverständnis konnten diese zwei Interpretationen als handhabbar belegt werden (vergl. Kap. 3.6 auf Seite 73).

Da die Komplexität der menschlichen Gestenerkennung nicht vollständig in den mathematischen Definitionen abzubilden ist, müssen Einschränkungen in der Erkennung oder aber in der Ausführung gegenüber der natürlichen Hand-Armbewegung angenommen und akzeptiert werden. Dazu gehören:

- Einschränkung des nutzbaren Gestenraumes (vergl. Kap. 3.4 auf Seite 67).

- Vereinzeln der Gesten: die Gesten sind als Folge nacheinander auszuführen.

Darüber hinaus sollte der Nutzer in der Art der Gestenausführung möglichst nicht weiter eingeschränkt werden.

Während der Algorithmenentwicklung stellte sich heraus, dass unter einer dynamischen Beschreibung auch ein definierter sequenzieller Ablauf einzelner Zustände verstanden werden kann (Prozessdefinition, wird dann als „Phasen- definition“ bezeichnet). Die Bewegung muss nicht durch eine durchgängige, z.B. mittels Regression zu bestimmende, Funktion darstellbar sein.

Darüber hinaus kann der sequenzielle Ablauf verschiedener definierter Zustände die Zahl der notwendigen Datenpunkte zur Beschreibung und Erkennung einer Zeigegeste reduzieren. Details zu der Datenreduktion finden sich in der Beschreibung der Phasendefinition mittels eines Winkels (vergl. Kap. 4.4 auf Seite 93) sowie in der Versuchsauswertung (vergl. Kap. 5.3 auf Seite 102).

Der vorgestellte Ablauf entspricht dem Entwicklungsverlauf, von der einfach verbal zu beschreibenden, aber umständlich zu parametrisierenden statischen Definition mit niedriger Gesamterkennungsrate zur komplexen Phasendefinition. Komplex ist diese Definition, da der gesamte Gestenzyklus betrachtet wird. Von den zu berücksichtigenden Parametern ist die Phasendefinition mittels eines Winkels, der dazugehörigen Winkelgeschwindigkeit und einer Haltezeit hingegen die einfachste, aber zugleich jene, mit der besten Gesamterkennungsrate.

Unabhängig von der Gestenbeschreibung ist in der Arbeit als Winkel immer der eingeschlossene Winkel der zwei beschriebenen Vektoren im Raum gemeint. Berechnet wird der Winkel nach Formel 4.1 .

$$\cos \Phi = \frac{\bar{a} \cdot \bar{b}}{|\bar{a}| \cdot |\bar{b}|}$$

$$\Phi = \arccos \left( \frac{a_x \cdot b_x + a_y \cdot b_y + a_z \cdot b_z}{\sqrt{a_x^2 + a_y^2 + a_z^2} \cdot \sqrt{b_x^2 + b_y^2 + b_z^2}} \right) \quad (4.1)$$

**Eingeschlossener Winkel  $\Phi$  im Raum [Zeidler et al., 1996]**

## 4.1 Statische Definition

Die statische Beschreibung einer Geste setzt voraus, dass „der Ausführende“ eine bestimmte, als „Zeigen“ identifizierbare Körperhaltung (siehe Abb. 4.2

auf der nächsten Seite) einnimmt und diese für einen bestimmte „Haltezeit“ beibehalten muss.

Diese Körperhaltung leitet sich von der bei Fricke [2007] beschriebenen „typische Geste mit ausgestrecktem Arm und ausgestrecktem Zeigefinger“ ab. Durch die starke Konventionalisierung wird sie „auch ohne begleitende Rede als hinweisende Geste verstanden“ und bietet sich damit für die Erkennung durch ein technisches System besonders an.

Aus dieser Grundüberlegung ergeben sich alle Parameter, durch deren Festlegung die statische Definition beschrieben ist:

**Körperhaltung:** Wie lässt sich auf Basis der vom Gestenaufnahmesystem gelieferten Daten die Körperhaltung beschreiben?

**Haltezeit:** Wie lange muss der Ausführende die Körperhaltung beibehalten, damit sicher gestellt ist, dass sie nicht zufällig während anderer Bewegungsabläufe eingenommen wird?

Neben diesen beiden primären Fragestellungen ergeben sich weitere untergeordnete Fragestellungen im Rahmen der Nutzer- oder Zielerkennung, die für eine zuverlässige Gestenerkennung notwendig sind. Darunter fallen:

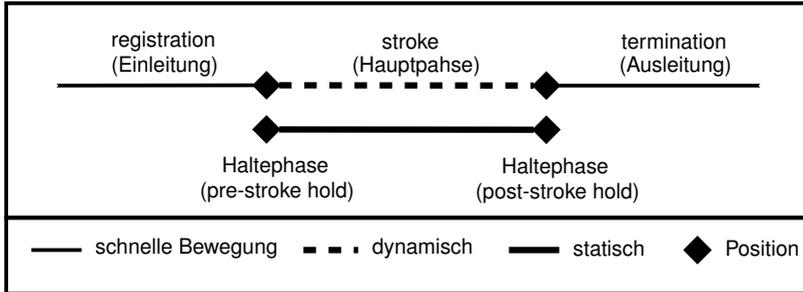
**Nutzererkennung:** Wer ist der Ausführende?

Dabei gilt es, zusätzlich die Frage der grundsätzlichen Nutzererkennung („Ist das ein Nutzer?“) sowie der Bediensicherheit („Darf dieser Nutzer das System bedienen?“) zu klären.

**Zielerkennung:** Wohin hat der Ausführende gezeitigt? Was ist das Zielobjekt?

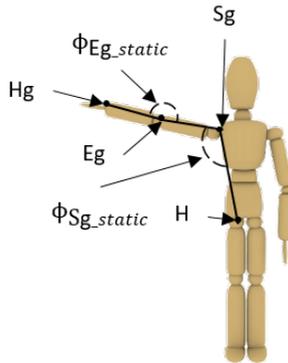
Antworten auf die untergeordneten Fragestellungen und aus ihnen abgeleitete Lösungsansätze werden mit der Frage der Übertragbarkeit auf ein Online-system (vergl. Kap. 6.6 auf Seite 132) diskutiert.

Das grundlegende Phasenmodell von Kendon [2004] (siehe Abb. 4.1 auf der nächsten Seite) basiert auf einem Mensch-zentrierten Ansatz der Gestenbeschreibung und -erkennung. Der Mensch durchläuft zur Gestenausführung immer die drei Bewegungsphasen Einleitung, Hauptphase, Ausleitung. Der Ansatz einer statische Beschreibung der Zeigegeste basiert auf der Hypothese, dass die Information zur Erkennung des Zeigens ausschließlich in der Haltephase gewonnen wird. Die für die Zeigegeste relevante Haltephase ist der post-stroke hold (Fortlaufend nur noch Haltephase genannt). Die menschliche Gestenausführung zeichnet sich durch einen vollständigen Gestenzyklus aus, der mindestens die drei Bewegungsphasen umfasst.



**Abbildung 4.1: Phasen einer Geste, erweitertes Modell**

nach [McNeill, 1992] [Pavlovic et al., 1997] [Kendon, 2004] [Walter et al., 2013]



**Abbildung 4.2: Statische Definition der Zeigegeste**

Handgelenkpunkt (Hg), Schultergelenkpunkt (Sg), Ellbogengelenkpunkt (Eg), Hüftgelenkpunkt (H),  
Ellbogenwinkel ( $\phi_{Eg}$ ), Schulterwinkel ( $\phi_{Sg}$ )

Die **Einleitung** beschreibt die Bewegung des zeigenden Armes in den Gestenraum hinein. Dazu muss der Arm die Ruheposition an beliebiger Stelle verlassen und den Gestenraum oberhalb der Brust erreichen. Im direkten Anschluss an die Einleitung wird die Bewegung der Hauptphase durchlaufen und erst in der Haltephase gestoppt, um die Zeigepose (siehe Abb. 4.2 auf der vorherigen Seite) einzunehmen und für ungefähr fünf Sekunden zu halten. Die Haltezeit der Zeigepose von fünf Sekunden ist eine aus der Beobachtung von Zeigegesten geschätzte Zeit, die durch weitere Untersuchungen präzisiert wird. Die Zeigepose definiert sich hauptsächlich durch den gestreckten Arm, Ellbogenwinkel ( $\Phi_{Eg}$ ) gleich  $180^\circ$ . Der Ellbogenwinkel berechnet sich aus den Vektoren, die bestimmt sind durch Handgelenkpunkt (Hg) Richtung Ellbogengelenkpunkt (Eg) sowie Schultergelenkpunkt (Sg) Richtung Schultergelenkpunkt (Sg).

Aufgrund von Abbildungsfehlern des Gestenaufnahmesystems (vergl. Kap. 6.5 auf Seite 131) sind die Winkel sowie die Haltezeit über Schwellwerte zu bestimmen. Daraus ergibt sich für den Ellbogenwinkel:

$$\Phi_{Eg\_static} > 160^\circ \quad (4.2)$$

#### **Zwangsbedingung für den Ellbogenwinkel $\Phi_{Eg}$**

während des statischen Zeigens

Eine weitere Bedingung neben dem gestreckten Arm ist, dass dieser nicht am Körper anliegen darf. Nur so kann geprüft werden, dass sich die Hand im Gestenraum befindet. Dieses Phänomen lässt sich durch den Schulterwinkel  $\Phi_{Sg}$  beschreiben. Der Schulterwinkel, definiert aus den Vektoren von Ellbogengelenkpunkt (Eg) Richtung Schultergelenkpunkt (Sg) sowie von Schultergelenkpunkt (Sg) Richtung Hüftgelenkpunkt (H), beschreibt bei Überschreiten einer definierten Schwelle das Verlassen des Arbeitsraumes, und damit eine Pose im Gestenraum. Bereits eine Abweichung von der Neutrallage um mehr als  $20^\circ$  ist vollkommen ausreichend (siehe Abb. 4.2 auf der vorherigen Seite).

$$\Phi_{Sg\_static} > 20^\circ \quad (4.3)$$

#### **Zwangsbedingung für den Schulterwinkel $\Phi_{Sg}$**

während des statischen Zeigens

Die als Zeigepose definierte Armhaltung muss von dem Bediener über einen Zeitraum von mindestens fünf Sekunden gehalten werden. Halten bedeutet, dass sich der Handgelenkpunkt (Hg) in dieser Zeit nicht „signifikant“ in der Position verändert. Neben den willkürlichen Bewegungen des Handgelenkpunktes (Hg) sind die mit der Microsoft® (MS) Kinect® bestimmten Skelettpunkte stark von unwillkürlichen Bewegungen des Handgelenkpunktes (Hg) in der Haltezeit überlagert. Hintergrund dieser verrauschten Skelettdaten ist zum Einen die bereits erwähnte Messunsicherheit des Gestenaufnahmesystems, zum Anderen aber auch der Haltetremor [Deuschl et al., 1998] des Menschen. Diese unwillkürlichen Ausgleichsbewegungen, insbesondere das hochfrequente Zittern des ausgestreckten Armes, bedingen auch für das Halten eine Schwellwertdefinition. Wie bei der Phasendefinition mittels eines Winkels können die Winkelverläufe zusätzlich durch Einsatz eines passenden Filters geglättet werden.

In den aufgezeichneten Daten wird die Zeit nicht mitgeführt, sondern nur die Bildnummer. Durch die verschiedenen eingesetzten Gestenaufnahmesysteme mit ihren unterschiedlichen Bildwiederholraten ergibt sich, dass für die mathematische Beschreibung die Zeitdauer in Bildern ausgedrückt wird. Bei der für die Evaluation (vergl. Kap. 5 auf Seite 99) eingesetzten Kinect® 2 ist die Bildwiederholrate mit 30 frames per second (dt. Bilder pro Sekunde, fps) angegeben, wodurch sich ein haltezeitäquivalenter Wert von 150 frames ergeben würde. In der Praxis hat sich gezeigt, dass 15 Bilder (oder 0,5 Sekunden) für die Gestenerkennung ausreichend sind. In dieser Zeit darf sich der Handpunkt nur um weniger als 50 mm, der Schulterwinkel ( $\Phi_{Sg}$ ) und Ellbogenwinkel ( $\Phi_{Eg}$ ) um weniger als  $5^\circ$  verändern.

Die Bestimmung des statischen Haltens wird mit einer Fensterfunktion (ähnlich einer Filterfunktion) durchgeführt. Diese Fensterfunktion wird mit der Laufvariablen  $i$  durch die Datensätze verschoben. Für die Evaluation des statischen Haltens werden abweichend nicht alle Datensätze innerhalb des Fensters betrachtet, sondern nur der Offset  $|\Delta\Phi_{(i,i-15)}|$  über die Fensterbreite von 15 Frames. Daraus ergibt sich für die Definition des statischen Haltens:

$$\begin{aligned} \Delta t &\geq 500 \text{ ms} \hat{=} 15 f \\ |\Delta\Phi_{Sg\_static}| &\leq 5^\circ \text{ mit } \Delta\Phi_{Sg\_static} = \Phi_{Sg(i)} - \Phi_{Sg(i-15)} \\ |\Delta\Phi_{Eg\_static}| &\leq 5^\circ \text{ mit } \Delta\Phi_{Eg\_static} = \Phi_{Eg(i)} - \Phi_{Eg(i-15)} \\ |\Delta Hg| &\leq 50 \text{ mm mit } \Delta Hg = Hg(i) - Hg(i-15) \end{aligned} \quad (4.4)$$

### Zwangsbedingung zur Definition des statischen Haltens

mit: Zeit ( $t$ ), Laufvariable im Zeitfenster ( $i$ ), frame ( $f$ )

Für die statische Beschreibung wird die **Ausleitung** nicht weiter betrachtet, da sie keine notwendige Information enthält. Während der Nutzer die Geste ausleitet, kann ein technisches System bereits das Zielobjekt suchen und identifizieren. Diese Überlegung ist nicht Bestandteil dieser Abhandlung.

## 4.2 Dynamische Beschreibung mittels linearer Regression

Die Untersuchungen zum Nutzerverständnis (vergl. Kap. 3.6 auf Seite 73) identifizierten die Zeigegeste als statische Geste. Dennoch werden im folgenden verschiedene dynamische Gestenbeschreibungen entwickelt, da eine vom Menschen ausgeführte Geste immer Bewegungen und damit dynamische Anteile enthält. Von den dynamischen Gestenbeschreibungen sind bei näherer Untersuchung einige wieder verworfen worden, z.B. die lineare Regression der Bewegung des Handgelenkpunktes.

Erst unter Vorgabe der Hypothese, dass eine definierte Abfolge von diskreten Zuständen (hier: Posen) auch eine dynamische Beschreibung eines Ausführungsprozesses ist, konnten zwei Beschreibungsansätze erarbeitet werden, die in verschiedenen Experimenten evaluiert wurden. Diese Form der dynamischen Beschreibung wird im folgenden als Phasendefinition bezeichnet.

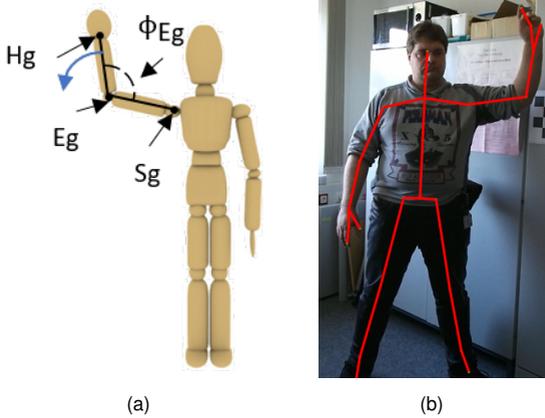
Der erste Ansatz zur dynamischen Bestimmung des „Zeigevektors“ auf Basis einer linearen Regression wird im Folgenden kurz vorgestellt. Dieser Ansatz basiert auf der Annahme, dass der Proband die Hand aus der Anmeldepose (siehe Abb. 4.3 auf der nächsten Seite) entlang einer virtuellen Geraden in Richtung des Zieles schiebt, so wie zum Beispiel ein Dart geworfen wird.

Dazu muss der Ellbogenwinkel ( $\Phi_{Eg}$ ) beginnend mit ungefähr  $90^\circ$  (Armhaltung der Anmeldepose, siehe Abb. 4.3(b) auf der nächsten Seite) monoton bis auf mind.  $160^\circ$  ansteigen (siehe Abb. 4.4 auf der nächsten Seite, Frame 2893 bis 2906). Diese Änderung des Ellenbogenwinkels ( $\Phi_{Eg}$ ) wird zur Identifikation einer beliebigen Geste, nicht aber zur endgültigen Identifikation des Zeigens verwendet.

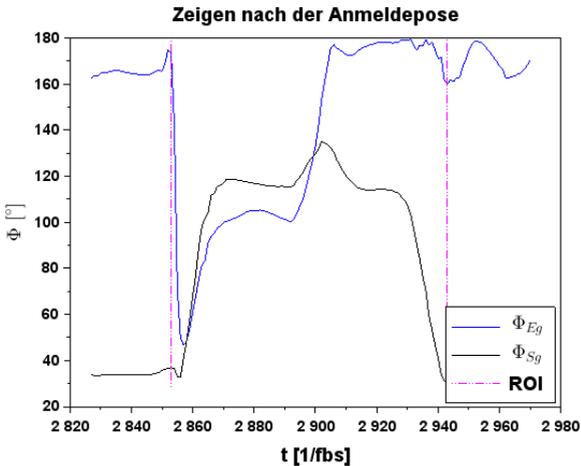
$$\Phi_{Eg(i-1)} \leq \Phi_{Eg(i)} \quad (4.5)$$

**Zwangsbedingung der Definition des Ellenbogenwinkles  $\Phi_{Eg}$**   
für dynamischen Zeigen mittels Linearer Regression)

mit: Laufvariable im Zeitfenster (i)

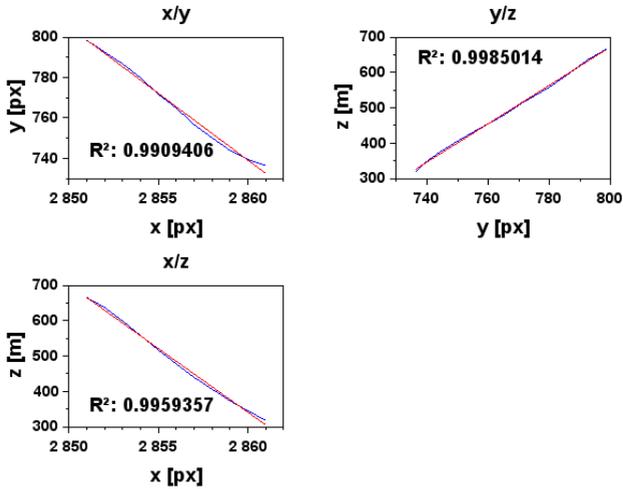


**Abbildung 4.3: Anmelderpose (pre-stroke hold)**  
 Handgelenk (Hg), Schultergelenk (Sg), Ellbogengelenk (Eg), Ellbogenwinkel ( $\Phi_{Eg}$ )  
 (a) postulierte Körperhaltung (b) eigene Aufnahme



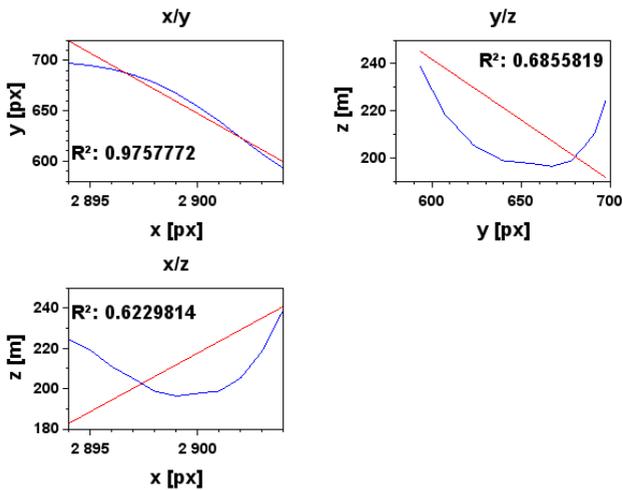
**Abbildung 4.4: Charakteristische Änderung des Ellbogen- ( $\Phi_{Eg}$ ) und des Schulterwinkels ( $\Phi_{Sg}$ ) im Zeitverlauf**

Dynamische Beschreibung der Zeigegeste mittels linearer Regression  
 mit  $t$  in [1/fps] bei 30 fps, Region of Interest (ROI)



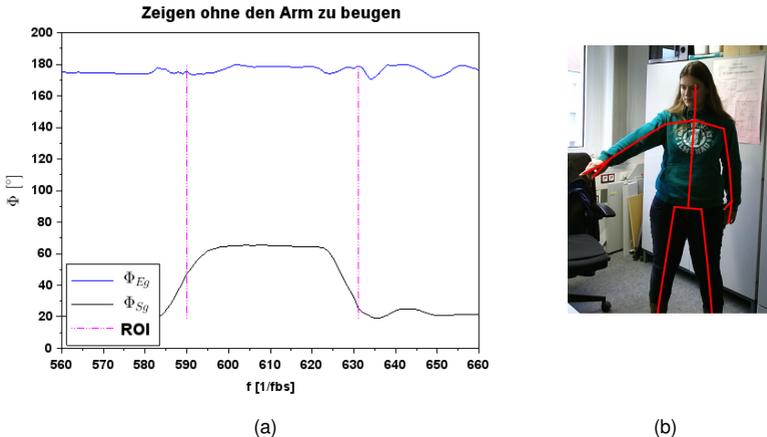
**Abbildung 4.5: Erwartete Bewegungsbahn des Handgelenkpunktes in allen Ebenen (Hg, Kartesische Koordinaten)**

Dynamische Beschreibung der Zeigegeste mittels linearer Regression



**Abbildung 4.6: Lineare Regression der Bewegungsbahn des Handgelenkpunktes in allen Ebenen (Hg), Kartesische Koordinaten**

Dynamische Beschreibung der Zeigegeste mittels linearer Regression



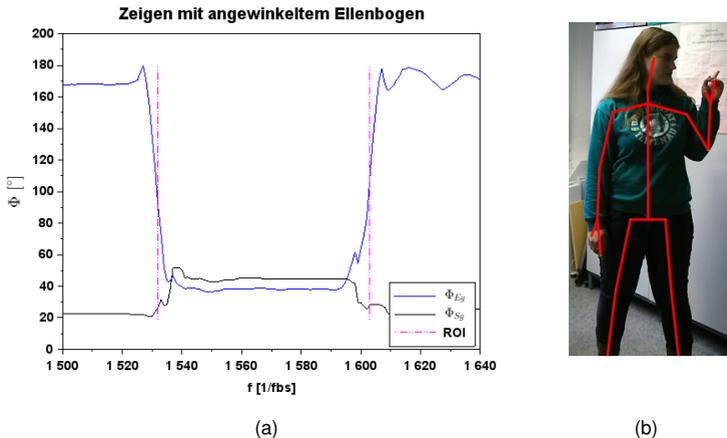
**Abbildung 4.7: Zeigen ohne den Arm zu beugen**

Winkelverlauf, der nicht mittels der lineare Regression identifiziert werden kann.  
mit Ellbogenwinkel ( $\Phi_{Eg}$ ), Schulterwinkel ( $\Phi_{Sg}$ ),  $t$  in [1/fps] bei 30 fps, Region of Interest (ROI)

(a) Winkelverlauf (b) eigene Aufnahme

Forderung: zur Eingrenzung der Geste und zum Erkennen des Zeigens soll sich bei der Betrachtung des Handgelenkpunkts über den Zeitraum der Zeigebewegung in allen drei Koordinatenebenen eine lineare Regression mit einem Bestimmtheitsmaß ( $R^2$ ) über 0,95 ermitteln lassen (siehe Abb. 4.5 auf der vorherigen Seite). Zur Ermittlung dieses Datenbereiches bietet sich der Einsatz einer Fensterfunktion mit variablen Fensterbreite und Startposition an. Im Zeitbereich des untersuchten Gestenzyklus lies sich eine Sequenz von 10 frames identifizieren, in der die Regression in allen Raumebenen im geforderten Maß liegt. Diese Sequenz befindet sich aber nicht in dem Zeitbereich, der durch die Bedingungen zum Verlassen der Ruhephase eingegrenzt wurde. Die beste lineare Annäherung in diesem Zeitfenster liegt bei  $R^2 = 0,97$  in der x/y Ebene (siehe Abb. 4.6 auf der vorherigen Seite). Für die x/z Ebene und die y/z Ebene beträgt  $R^2$  dann weniger als 0,7.

Schon allein aus dieser Erkenntnis heraus muss diese Variante verworfen werden. Darüber hinaus zeigte sich, dass viele Probanden bei der spontanen Ausführung der Zeigegeste einen alternativen Bewegungsablauf bevorzugen, vor allem wenn keine Anmeldepose der Gestenausführung zwingend vorgeschaltet wird. Wie bei [Nowack et al., 2016] bereits dargestellt, sind die beiden häufig ausgeführten Alternativen bei der Gestenausführung:



**Abbildung 4.8: Zeigen mit angewinkeltem Ellenbogen**  
 Winkelverlauf, der nicht mittels der lineare Regression identifiziert werden kann.  
 mit Ellbogenwinkel ( $\Phi_{Eg}$ ), Schulterwinkel ( $\Phi_{Sg}$ ),  $t$  in [1/fps] bei 30 fps, Region of Interest (ROI)  
 (a) Winkelverlauf (b) eigene Aufnahme

- Die direkte Bewegung des gestreckten Armes in die Zeigepose (siehe Abb. 4.7 auf der vorherigen Seite).
- Das Zeigen mit dem angewinkelten Arm (siehe Abb. 4.8).

### 4.3 Phasendefinition mittels zweier Winkel

Unter der Annahme der Hypothese, dass ein dynamischer Prozess auch als Abfolge diskreter Zustände angesehen werden kann, gilt es, die folgenden Beschreibungsmodelle zu betrachten, welche auf einer erweiterten Abfolge des Phasenmodelles von Kendon (siehe Abb. 4.1 auf Seite 80) basieren. Dem Phasenmodell wird eine zusätzlicher Phase vor- bzw. nachgeschaltet. Es handelt sich um die **Ruhephase** oder der „rest state“ wie auch für die „tri-phasic gesture“ von Wilson et al. [1996] beschrieben. Derartige Definitionen werden im folgenden als Phasendefinitionen bezeichnet.

Der Ausgangszustand, die **Ruhephase**, wird definiert durch die sich parallel zum Vektor Torso ( $\vec{T}$ ) (Punkt Wirbelsäule Basis (WsB)  $\rightarrow$  Punkt Wirbelsäule Schulter (WsS)) befindlichen Arme. Die Parallelität bedingt, dass sich die Handgelenkpunkt (Hg) außerhalb des Gestenraumes befinden. Die Paral-

lelitätsbedingung kann auf Basis eines virtuellen Schulterwinkels ( $\Phi_{(\overline{HS}, \overline{T})}$ ) überprüft bzw. berechnet werden (siehe Abb. 4.9(a) auf der nächsten Seite). Der virtuelle Schulterwinkel ( $\Phi_{(\overline{HS}, \overline{T})}$ ) wird gebildet aus der Bezugsachse, „Vektor Torso ( $\overline{T}$ )“ (WsS  $\rightarrow$  Punkt Wirbelsäule Basis (WsB)) und dem virtuellen Arm, dem Vektor Hand-Schulter ( $\overline{HS}$ ).

Die Abbildung 4.9(b) auf der nächsten Seite stellt neben dem Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ) (blau) noch die Winkelgeschwindigkeit ( $\dot{\Phi}_{(\overline{U}, \overline{T})}$ ) eines weiteren virtuellen Winkels zwischen dem Vektor Unterarm ( $\overline{U}$ ) und Vektor Torso ( $\overline{T}$ ) über den Einzelbildern (Frames) in grün dar. Diese Winkelgeschwindigkeit wird für die der Erkennung der Geste (Beginn und Ende) verwendet, da egal wie der Ausführende eine Zeigebewegung ausführt, die Hand und damit der Unterarm bewegt wird. Selbst wenn der Ausführende mit angewinkeltem Arm ( $\phi_{Eg} \approx 90^\circ$ ) zeigt, muss sich die Hand im Gestenraum befinden (siehe Abb. 4.8(b) auf der vorherigen Seite).

Basierend auf der Annahme, dass in der Ruhephase der Arm gestreckt ist, kann der Arm durch den direkten Vektor Hand-Schulter ( $\overline{HS}$ ) (Handgelenkpunkt (Hg)  $\rightarrow$  Schultergelenkpunkt (Sg)) approximiert werden. Dazu muss davon ausgegangen werden, dass die Vektor Hand-Schulter ( $\overline{HS}$ ) bedingt durch die Anatomie des Menschen nicht exakt parallel zur „Vektor Torso ( $\overline{T}$ )“ ausgerichtet sein kann. Daraus folgt auch für die Parallelitätsbedingung (Ruhephase), dass diese durch einen Schwellwert angenähert wird. Der Schwellwert wird auf  $11^\circ$  festgelegt.

$$\Phi_{(\overline{HS}, \overline{T})} \leq 11^\circ$$

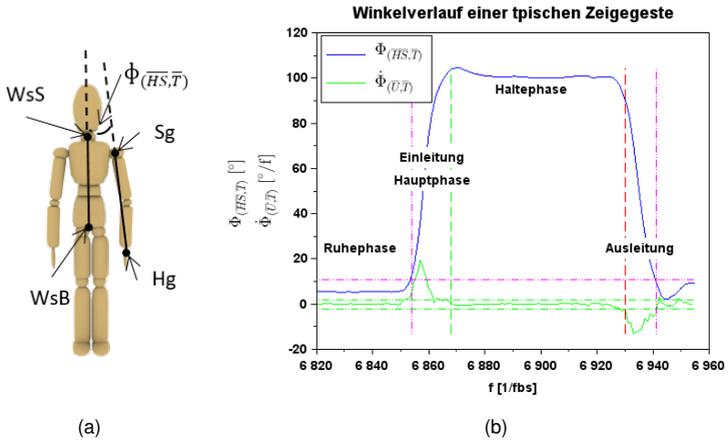
mit  $\Phi_{(\overline{HS}, \overline{T})} = \angle(\overline{HS}, \overline{T})$  (4.6)

**Zwangsbedingungen zur Definition der Ruhephase (Parallelitätsbedingung)**  
**(siehe Abb. 4.9(a) auf der nächsten Seite)**

in der Phasendefinition mittels zweier Winkel

mit: Vektor Hand-Schulter ( $\overline{HS}$ ), Vektor Torso ( $\overline{T}$ ), Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ )

Über die Winkelgeschwindigkeit zwischen den Vektoren  $\overline{U}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{U}, \overline{T})}$ ) in Verbindung mit der Auflösung der Parallelitätsbedingung, Formel 4.6 definiert sich der **Beginn der Gestenausführung** (siehe Abb. 4.10(a) auf der nächsten Seite) wie folgt:



**Abbildung 4.9: Winkelverlauf einer typischen Zeigegeste, incl. Ruhephase**

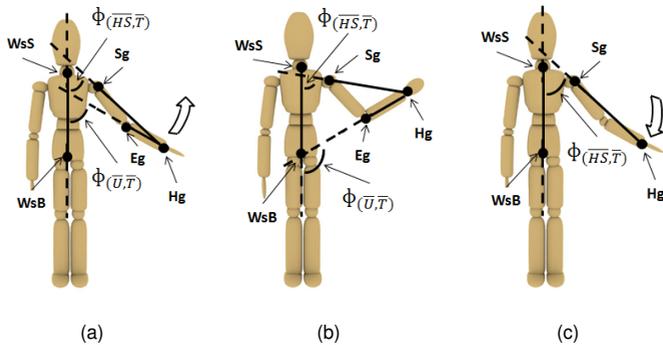
(Phasendefinition mittels zweier Winkel),

mit Punkt Wirbelsäule Schulter (WsS), Punkt Wirbelsäule Basis (WsB), Schultergelenkpunkt (Sg), Handgelenkpunkt (Hg), Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(HS, T)}$ ),

Winkelgeschwindigkeit zwischen den Vektoren  $\overline{U}$  und  $\overline{T}$  ( $\dot{\Phi}_{(U, T)}$ ), frame (f), t in [1/fps] bei 30 fps

(a) Körperhaltung in der Ruhephase (b) Winkelverlauf einer typischen Zeigegeste

Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(HS, T)}$ ) (blau), Winkelgeschwindigkeit zwischen den Vektoren  $\overline{U}$  und  $\overline{T}$  ( $\dot{\Phi}_{(U, T)}$ ) (grün)



**Abbildung 4.10: Körperhaltung in den verschiedenen Gesteinsphasen**

in der Phasendefinition mittels zweier Winkel

(a) Beginn Gesteinsphase (b) Haltephase (c) Ausleitung

$$\begin{aligned} \dot{\Phi}_{(\bar{U}, \bar{T})} &\geq 2^\circ / f \\ \text{mit } \Phi_{(\bar{U}, \bar{T})} &= \angle(\bar{U}, \bar{T}) \\ \text{und } \Phi_{(\overline{HS}, \bar{T})} &\leq 11^\circ \end{aligned} \quad (4.7)$$

**Zwangsbedingung zur Definition Beginn der Gestenausführung (siehe Abb. 4.10(a) auf der vorherigen Seite)**

in der Phasendefinition mittels zweier Winkel

mit: Vektor Unterarm ( $\bar{U}$ ), Vektor Torso ( $\bar{T}$ ), Winkel zwischen den Vektoren  $\bar{U}$  und  $\bar{T}$  ( $\Phi_{(\bar{U}, \bar{T})}$ ), frame (f)

Eine detaillierte Beschreibung / Parametrisierung der **Einleitung** und der **Hauptphase** ist zur Erkennung der Zeigegeste nicht notwendig. Diese wird, wie bereits bei der ausschließlich statischen Beschreibung erläutert, hauptsächlich während der Haltephase beschrieben. Dem Beginn der Gestenausführung muss daher die Haltephase entsprechend der Formel 4.8 folgen (siehe Abb. 4.9(b) auf der vorherigen Seite). Es darf zwischendurch kein weiteres der im folgenden beschriebenen Gestenelemente durchlaufen bzw. erreicht werden, siehe hierzu auch das Ablaufdiagramm der dynamischen Gestenausführung (siehe Abb. 4.11 auf Seite 92).

Dieser konzeptionelle Ansatz ermöglicht zukünftige Erweiterungen, z. B. die Erkennung weiterer Gesten. Diese können nach dem Beginn der Gestenausführung durch ein Bewegungsmuster beschrieben werden, welches alternativ zur Zeigegeste vom technischen System erkannt werden kann, dazu mehr im Ausblick (vergl. Kap. 8 auf Seite 147).

Auch in der „Phasendefinition mittels zweier Winkel“ ist das Zeigen eine statische Körperpose (siehe Abb. 4.10(b) auf der vorherigen Seite), die durch den Winkel zwischen den Vektoren  $\overline{HS}$  und  $\bar{T}$  ( $\Phi_{(\overline{HS}, \bar{T})}$ ), die Winkelgeschwindigkeit zwischen den Vektoren  $\bar{U}$  und  $\bar{T}$  ( $\dot{\Phi}_{(\bar{U}, \bar{T})}$ ) und eine Haltezeit definiert ist. Durch den festen Ablauf und die noch zu definierende **Ausleitung** kann das Äquivalent der notwendigen Haltezeit ( $t_H$ ) auf 5 frames  $f$  reduziert werden.

$$\begin{aligned} \Phi_{(\overline{HS}, \bar{T})} &\leq 11^\circ \\ \text{und } |\dot{\Phi}_{(\bar{U}, \bar{T})}| &\geq 2^\circ / f \\ \text{über } t_H &> 5 f \end{aligned} \quad (4.8)$$

**Zwangsbedingung zur Definition Zeigen in der Haltephase (siehe Abb. 4.10(b) auf der vorherigen Seite)**

in der Phasendefinition mittels zweier Winkel

mit: Äquivalent der notwendigen Haltezeit ( $t_H$ )

Die **Ausleitung**, welche der Haltephase unmittelbar folgt, wird charakterisiert durch einen monoton fallenden Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ) und abgeschlossen durch das Erreichen der Ruhephase gemäß Formel 4.6 auf Seite 88 . Erst mit Erreichen der Ruhephase, das heißt mit erfolgreichem Durchlaufen des gesamten Gestenzyklus, kann die Zeigegeste erfolgreich erkannt werden. Diese Beobachtung bestätigt auch McNeill [1992]. Er beschreibt, dass die Bewegung der Hände bei 70 % der ausgeführten Gesten außerhalb des Gestenraumes endet, um die Hände für eine weitere Geste einsetzen zu können.

$$\dot{\Phi}_{(\overline{HS}, \overline{T})} \leq 0^\circ / f \quad (4.9)$$

**Zwangsbedingung der Definition Ausleitung (siehe Abb. 4.10(c) auf Seite 89)**

in der Phasendefinition mittels zweier Winkel

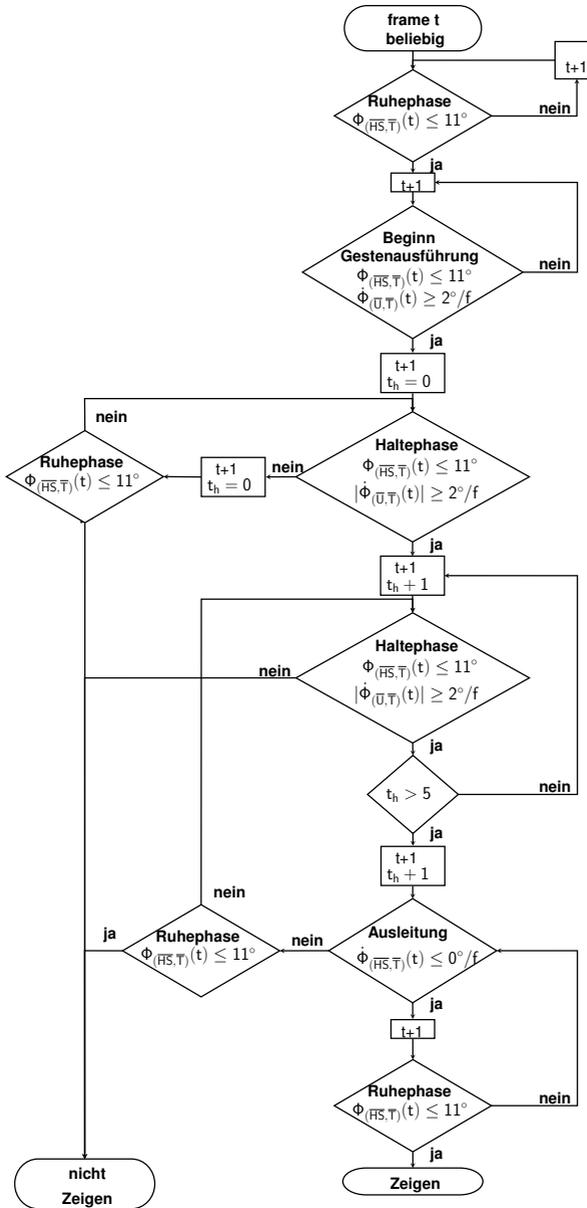


Abbildung 4.11: Ablaufdiagramm der Phasendefinition mittels zweier Winkel

## 4.4 Phasendefinition mittels eines Winkels

In der vorgestellten Phasensdefinition mittels zweier Winkel dient der Winkel zwischen den Vektoren  $\bar{U}$  und  $\bar{T}$  ( $\Phi_{(\bar{U}, \bar{T})}$ ) bzw. die Winkelgeschwindigkeit zwischen den Vektoren  $\bar{U}$  und  $\bar{T}$  ( $\dot{\Phi}_{(\bar{U}, \bar{T})}$ ) zur Festlegung, ob der Unterarm bewegt wird. Damit wird entschieden, ob eine Geste ausgeführt wird. Für den Vektor Unterarm ( $\bar{U}$ ) sowie für den Vektor Hand-Schulter ( $\overline{HS}$ ) ist während der hier betrachteten Bewegung das bestimmende Element der Handgelenkpunkt (Hg). Daher ist es zur Phasenbeschreibung einer Zeigegeste ausreichend, den Vektor Hand-Schulter ( $\overline{HS}$ ) und den Winkel zwischen den Vektoren  $\overline{HS}$  und  $\bar{T}$  ( $\Phi_{(\overline{HS}, \bar{T})}$ ) zu verwenden. Für die einzelnen Phasen der Gestenbeschreibung ergibt sich daraus:

In der **Ruhephase** ist der Winkel zwischen den Vektoren  $\overline{HS}$  und  $\bar{T}$  ( $\Phi_{(\overline{HS}, \bar{T})}$ ) kleiner  $21^\circ$ .

$$\begin{aligned} \Phi_{(\overline{HS}, \bar{T})} &< 21^\circ \\ \text{mit } \Phi_{(\overline{HS}, \bar{T})} &= \angle(\overline{HS}, \bar{T}) \end{aligned} \quad (4.10)$$

### Zwangsbedingung der Definition Ruhephase (Parallelitätsbedingung)

Phasendefinition mittels eines Winkels

mit Vektor Hand-Schulter ( $\overline{HS}$ ), Vektor Torso ( $\bar{T}$ ), Winkel zwischen den Vektoren  $\overline{HS}$  und  $\bar{T}$  ( $\Phi_{(\overline{HS}, \bar{T})}$ )

Damit durch den Tremor des Ausführenden (siehe Abb. 5.6 auf Seite 107) die Schwellwerte der Winkelerkennung und der Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\bar{T}$  ( $\dot{\Phi}_{(\overline{HS}, \bar{T})}$ ) nicht zu früh über- bzw. unterschritten werden, wird der Zeitverlauf des Winkels Hand-Schulter - Torso ( $\Phi_{(\overline{HS}, \bar{T})}$ ) mit einem „Double Moving Average Filter“ entsprechend der Empfehlung im Microsoft® Whitepaper [Azimi, 2016] geglättet.

Der Übergang aus der Ruhephase in die **Einleitung (Gestenausführung)** wird definiert über die Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\bar{T}$  ( $\dot{\Phi}_{(\overline{HS}, \bar{T})}$ ). Ist die Winkelgeschwindigkeit größer als  $1^\circ$ , ist die Ruhephase beendet und ein gewolltes Anheben, also in den Gestenraum bringen des Handgelenkpunkts (Hg), eingeleitet.

$$\dot{\Phi}_{(\overline{HS}, \overline{T})} > 1^\circ/f \quad (4.11)$$

### Zwangsbedingungen der Definition Beginn der Gestenphase

Phasendefinition mittels eines Winkels

mit Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ )

Der Einleitung muss die Haltephase folgen. Wird nach der Einleitung der Gestenraum wieder verlassen, so bricht die Gestenerkennung ab und sucht nach einem erneuten Beginn der Gestenausführung (siehe Ablaufdiagramm, Abb. 4.13). Verbleibt der Handgelenkpunkt im Gestenraum, das heißt der Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ) ist größer  $21^\circ$ , kann es sich um die Haltephase oder eine andere, zukünftig noch zu definierende Gestenphase handeln. Für die Zeigegeste ist die folgende Haltephase durch das Einstellen der Bewegung, Betrag von Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ ) kleiner  $1\%/f$ , definiert. Handelt es sich um eine andere Bewegung, bricht die Erkennung mit der Meldung „nicht Zeigen“ ab.

$$\begin{aligned} \Phi_{(\overline{HS}, \overline{T})} &> 21^\circ \\ |\dot{\Phi}_{(\overline{HS}, \overline{T})}| &< 1^\circ/f \end{aligned} \quad (4.12)$$

### Zwangsbedingung der Definition Beginn Haltephase

Phasendefinition mittels eines Winkels

Komplementär zum Beginn der Haltephase ist deren Ende definiert. Dazu wird nur die Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ ) betrachtet, da sich während der Haltephase der Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ) nicht verändern kann.

$$|\dot{\Phi}_{(\overline{HS}, \overline{T})}| > 1^\circ/f \quad (4.13)$$

### Zwangsbedingung der Definition des Endes der Haltephase

Phasendefinition mittels eines Winkels

Liegen zwischen Beginn und Ende der Haltephase weniger als 6 frames, so bricht die Gestenerkennung ab und sucht nach dem Wiedereintritt in die Ruhephase.

Zu diesem Zeitpunkt könnte für eine potentielle Erweiterung die Körperpose mit anderen Gestenbeschreibungen untersucht werden. Für dynamische Gesten, wie sie von Freeman und Roth [1994] diskutiert werden, sollte die alternative Suche allerdings bereits nach der Einleitung der Gestenphase beginnen.

Wie auch bei der „Phasendefinition mittels zweier Winkel“ muss unmittelbar auf die Haltephase eine Ausleitung erfolgen. Die Hand kehrt auf direktem Weg in die Ruhelage zurück, der Zeitverlauf des Winkles Hand-Schulter -Torso ( $\Phi_{(\overline{HS}, \overline{T})}$ ) ist streng monoton fallend.

$$\dot{\Phi}_{(\overline{HS}, \overline{T})} < 0^\circ / f \quad (4.14)$$

#### **Zwangsbedingung zur Definition der Ausleitung**

Phasendefinition mittels eines Winkels

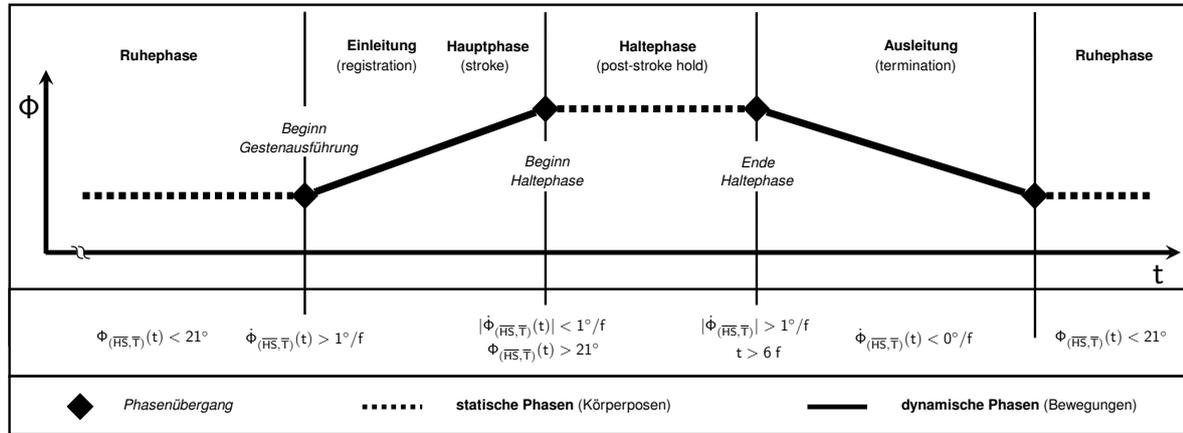


Abbildung 4.12: Phasen der Gestenerkennung

Phasenmodell mittels eines Winkels

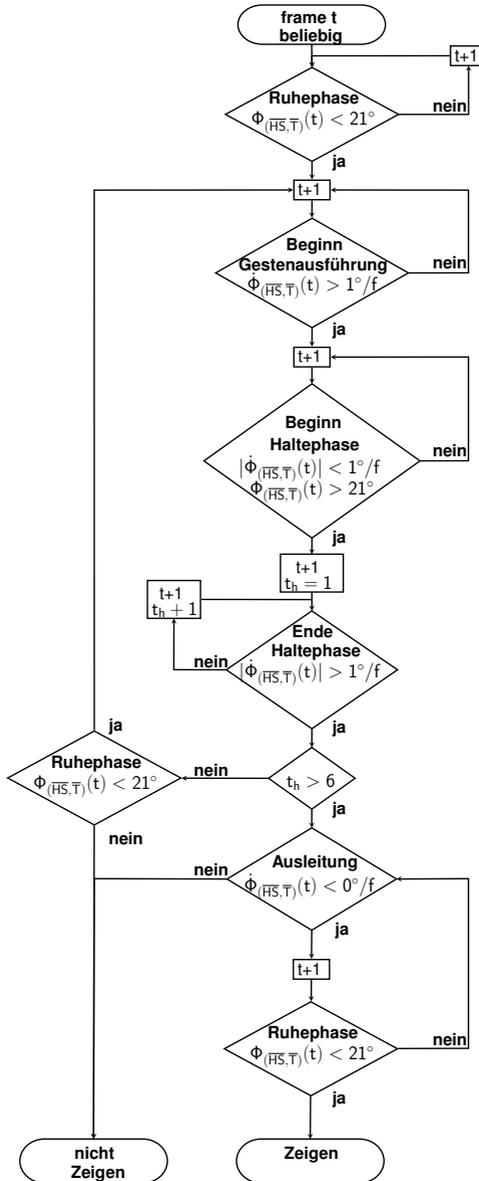


Abbildung 4.13: Ablaufdiagramm der Phasendefinition mittels eines Winkels



## Kapitel 5

# Experiment

Zur Parametrierung der im vorherigen Kapitel dargestellten mathematischen Beschreibungen einer Zeigegeste wurden verschiedene Versuche durchgeführt. Diese werden als eine zusammenhängende Versuchsreihe beschrieben, deren Daten zur Absicherung der theoretisch ermittelten Ansätze und zur Überprüfung der eingesetzten Schwellwerte herangezogen werden.

Es werden die statische Definition sowie die „Phasendefinition mittels zweier Winkel“ und die „Phasendefinition mittels eines Winkels“ untersucht. Dazu wurden bei den Versuchen sowohl die „Skelettdaten“ als auch die Bilder aufgezeichnet und anschließend mit einer Scilabumsetzung der verschiedenen Definitionen validiert.

### 5.1 Versuchsdesign

Es wurden Probandenversuche mit 15 Versuchspersonen, 8 weiblich, 7 männlich durchgeführt. Es handelte sich überwiegend um Studierende und Mitarbeiter der TU Ilmenau. Zwei Probanden kamen aus einem Industrieunternehmen und fungieren dort als Maschinenbediener und Schichtleiter.

Für die Versuche wurden in einem Laborraum mehre Zielmarkierungen an den Wänden angebracht, so dass sich die Zeigeziele im Raum vor der Frontalebene des Proband befinden. Für den Probanden befanden sich die Zeigeziele im Bereich von ca.  $\pm 90^\circ$  um seine Längsachse und  $-60/ + 90^\circ$  um seine Querachse. Ihre exakte Position hat für die Versuche keine direkte Auswirkung.



**Abbildung 5.1: Aufzeichnungssoftware für die Probandenversuche**

Aufgezeichnet wurden die Versuchsergebnisse mit einer Software [Wenzel et al., 2016] (siehe Abb. 5.1), die auf einem PC die, mit der Microsoft Kinect<sup>®</sup> 2 erfassten Koordinaten und den Status von 25 Skelettpunkten (siehe Abb. 2.14 auf Seite 44) mit einer Framerate von 30 fps (frames per second) aufzeichnen kann. Diese Software bietet dem Versuchsleiter die Möglichkeit, nach Ausführung einer Geste eine „Marker“ in den Daten zu setzen und so die einzelnen Gestensequenzen voneinander zu trennen.

Für die Versuchsdurchführung wurde die Kinect<sup>®</sup> 2 auf einem Stativ in ca. 1,4 m Höhe und in einem Abstand von zwei bis drei Metern vor dem Probanden aufgebaut. Auch diese Maße wurden bewusst nicht exakt eingemessen, da die Gestenerkennung später unabhängig von der Umgebung und den Aufnahmebedingungen funktionieren soll.

Folgende Anforderungen an die Position von Probanden und Versuchsleiter wurden gestellt:

- Zwischen Sensor und Proband war freie Sicht zu gewährleisten.
- Der Proband sollte annähernd frontal vor dem Sensorsystem stehen.

- Der Proband sollte während der Versuche vollständig von der Skeletterkennung erfasst werden.
- Der Versuchsleiter durfte während der Aufzeichnung nicht im Bild erscheinen.  
(Die Kinect® 2 kann mit der verwendeten Microsoft-Bibliothek bis zu sechs Personen im Bild erkennen. Diese zusätzlichen Daten müssten händisch aus den Aufzeichnungen entfernt werden.)

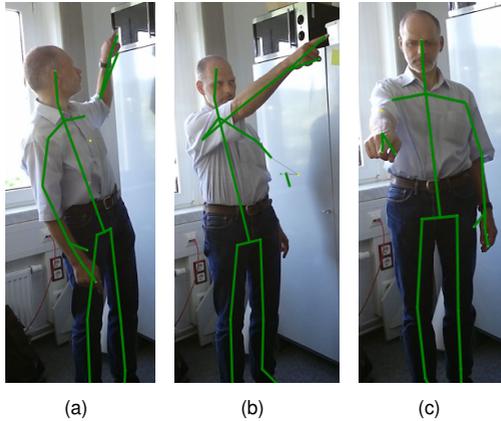
## 5.2 Versuchsdurchführung

Auf Anweisung des Versuchsleiters führten die Probanden in zwei Durchläufen definierte Gesten aus. Dabei wurde sichergestellt, dass die Probanden vor der Versuchsdurchführung nur begrenzt über die Auswertungsstrategie des Versuches instruiert wurden. Damit sollte erreicht werden, dass die Probanden die Gesten so ausführen, wie sie diese intuitiv ausführen würden. Die Probanden wurden lediglich instruiert, dass zwischen den einzelnen Gesten jeweils die Arme den Gestenraum wieder zu verlassen hatten.

Im „Versuchsteil 1“ (Gestenausführung mit definierter Körperseite) sollten die Probanden vier Gesten ausführen, jeweils zwei Mal hintereinander:

- Zeigen rechts,
- Zeigen links,
- Heranwinken rechts,
- Heranwinken links.

Im „Versuchsteil 2“ (Gestenausführung mit definierten Zielen bzw. Aussagen) wurden die Probanden zu gezielten Zeigebewegungen auf die zuvor vorgestellten Zielmarkierungen bzw. auf entsprechende Gegenstände im Versuchsraum aufgefordert. So konnte sichergestellt werden, dass die Probanden Zeigegesten in unterschiedliche Richtung ausführten und eine Auswertung der aufgenommenen Daten im 3D Raum erfolgen musste. Die Auswahl der Zeigeeziele und deren Reihenfolge war dem jeweiligen Versuchsleiter überlassen und änderte sich zwischen den einzelnen Versuchen zufällig. Auch in diesem Versuchsteil sollten die Probanden mit den Händen nach der Gestenausführung wieder den Gestenraum verlassen. Außerdem wurden zwischen einzelnen Zeigegesten die Probanden aufgefordert, andere Gesten wie „Klatschen“ oder zum „Abschied Winken“ (Kontrollgesten) auszuführen.



**Abbildung 5.2: Erkennungsprobleme in Abhängigkeit von der Zeigegeste**

(a) Oberkörper verdreht (b) Zeigen über den Oberkörper (c) Arm durch Hand verdeckt

Die Probanden konnten die Gesten mit der von ihnen bevorzugten Hand ausgeführt. Dadurch kam es teilweise zu Verdeckungen, da z.B. über den eigenen Körper hinweg gezeigt wurde (siehe Abb. 5.2(b)). Auch wurde von den Probanden der Oberkörper verdreht, sodass einzelne Skelettpunkte nicht mehr erkannt werden konnten (siehe Abb. 5.2(b) bzw. 5.2(c)).

Der Zeitliche Ablauf der Gesten im Versuchsteil 2 war wie folgt:

- vier verschiedene Zeigeeziele (Zeigen),
- Klatschen (nicht Zeigen),
- zwei verschiedene Zeigeeziele (Zeigen),
- Zeigen auf sich selbst oder am Kopf kratzen (nicht Zeigen),
- drei verschiedene Zeigeeziele (Zeigen),
- zum Abschied winken (nicht Zeigen).

### 5.3 Versuchsauswertung

Im Rahmen der Probandenversuche wurden 302 Gestensequenzen aufgezeichnet. Davon haben die Probanden 137 mal mit dem rechten und 58 mal mit

dem linken Arm gezeigt sowie 107 andere Kontrollgesten ausgeführt. Für die Auswertung wurden fünf Kategorien unterschieden:

- Es wurde eine „Zeigegeste“ ausgeführt und vom Algorithmus korrekt als „Zeigen“ erkannt (True Positive<sup>1</sup>).
- Es wurde eine „Zeigegeste“ ausgeführt und vom Alogrithmus als „nicht Zeigen“ abgelehnt (False Positive).
- Es wurde eine „Kontrollgeste“ ausgeführt und vom Algorithmus als „Zeigen“ erkannt (False Negative).
- Es wurde eine „Kontrollgeste“ ausgeführt und vom Algorithmus korrekt als „nicht Zeigen“ abgelehnt (True Negative).
- Sonderfälle, z.B.: Geste wurde mit rechts ausgeführt aber links erkannt, oder die Gestenausführung wurde zeitgleich rechts und links erkannt. Die erkannten Sonderfälle gehen als „Geste wurde nicht erkannt“ in die Bestimmung der Gesamterkennungsleistung des Algorithmusses ein. Die Diskussion der Sonderfälle folgt in Kap. 6.3 auf Seite 121.

Durch eine schrittweise Optimierung der Schwellwerte konnten abschließend mit den bereits in Kap. 4 auf Seite 77 vorgestellten Algorithmen folgende Erkennungsraten erzielt werden:

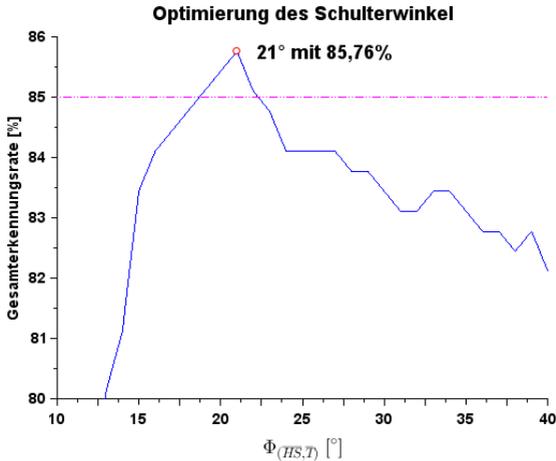
- zeigen Statisch: 49,0 %
- Phasendefinition mittels zweier Winkel: 63,9 %
- Phasendefinition mittels eines Winkels: 85,7 %

Die Optimierung der Parameter wurde nur für die „Phasendefinition mittels eines Winkels“ durchgeführt. Eine deutliche Verbesserung der Erkennungsleistung gegenüber der Definition des statischen Zeigens zeigte bereits die „Phasendefinition mittels zweier Winkel“, konnte aber auch mit der Erkennungsleistung der „Phasendefinition mittels eines Winkels“ nicht mithalten.

Die Untersuchungen von Fürntratt und Neuschmied [2014] hat ergeben, dass die Genauigkeit bei der Berechnung des Zeigevektors um 24,3 % besser ist, wenn über den Vektor Schulter Richtung Handgelenk im Vergleich zum Vektor Unterarm (Ellenbogen zu Handgelenk) der Zeigevektor ermittelt wird. Fürntratt und Neuschmied [2014] nutzte für die Untersuchung der Zeigebewegung ebenfalls den Kinect® 2 Tiefensensor. Da auch die in dieser Arbeit ausgewerteten

---

<sup>1</sup>Die englischen Begriffe werden aus den Gepflogenheiten der Statistiker heraus verwendet.



**Abbildung 5.3: Analyse des Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ) beim Verlassen bzw. Wiedereintritt in die Ruhephase**

Optimierung der Phasendefinition mittels eines Winkels

Daten ein Unterschied von über 20 % bestätigen, konnte auf eine Optimierung der Parameter der „Phasendefinition mittels zweier Winkel“ verzichtet werden.

In den folgenden Abschnitten werden die Winkelwerte sowie die Parameter des Glättungsfilters untersucht, um die Schwellwerte zu bestimmen, die mit den Testdaten die maximal erreichbare Erkennungsrate ermöglichen. Dabei werden eingegangene Kompromisse angesprochen und offene Fragestellungen für weiterführende Untersuchungen identifiziert.

## Überprüfung des Winkel zwischen den Vektoren $\overline{HS}$ und $\overline{T}$ ( $\Phi_{(\overline{HS}, \overline{T})}$ )

Aus verschiedenen Versuchsreihen, vor allem Praxisversuchen mit der statischen Definition des Zeigens in der Anwendung als Steuerung für die Hebe- und Tragehilfe hat sich der Winkel  $\Phi_{(\overline{HS}, \overline{T})}$  von  $20^\circ$  als gute Einstiegshypothese erwiesen.

Der postulierte Schwellwert von  $20^\circ$  liegt bereits im Bereich des Maximums. Bedingt durch die Zufallsstreuung der Stichprobe wird das lokale Maximum der

Erkennungsrate (siehe Abb. 5.3 auf der vorherigen Seite) mit 85,8 % bei 21 ° erreicht, 0,3 % über der Erkennungsrate bei einem Wert von 20°.

### Bestimmung der optimalen Haltezeit

Zur Ermittlung der optimalen Haltezeit ist die Gesamterkennungsrate mit einer Variationsbreite von 20 frame beginnend bei 0 bestimmt worden. Dabei zeigte sich ein lokales Maximum mit 85,8 % bei einer Haltezeit von 6 frames. Mit einer Haltezeit von 5 Frames kann immer noch eine Gesamterkennungsrate von über 85 % erreicht werden, wohingegen die Erkennungsrate bei 4 bzw. 7 Frames bereits unter 84 % abfällt (siehe Abb. 5.4 auf der nächsten Seite).

Betrachtet man die Erkennungsraten von „Zeigen“ *korrekt erkannt* (True Positive) und „nicht Zeigen“ *korrekt erkannt* (True Negative) im Detail, stellt sich heraus, dass mit steigender Haltezeit die Erkennungsrate des „Zeigens“ von annähernd 86,2 % linear gemäß Formel 5.1 auf 59,0 % abnimmt, während die Erkennungsrate des „nicht Zeigens“ bei 41,1 % beginnt und dann degressiv auf 96,26 % ansteigt. Die durch das Maximum in der Gesamterkennungsrate bestimmte Haltezeit von 6 frames stellt mit dem beschriebenen Versuchsaufbau den besten Kompromiss zwischen Annahme und Ablehnung einer Zeigegeste und hierfür den aktuell besten Schwellwert dar.

$$E_{r_z} = -1,35 * t_H + 88,39 \quad (5.1)$$

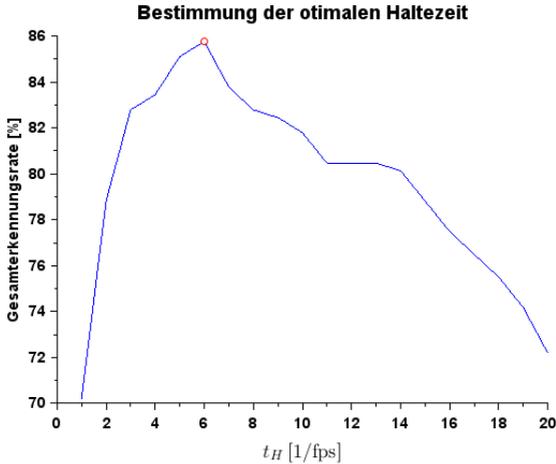
mit  $R^2 = 0.98$

#### Lineare Abhängigkeit, Erkennungsrate „Zeigen“ *korrekt erkannt* ( $E_{r_z}$ ) zu Haltezeit

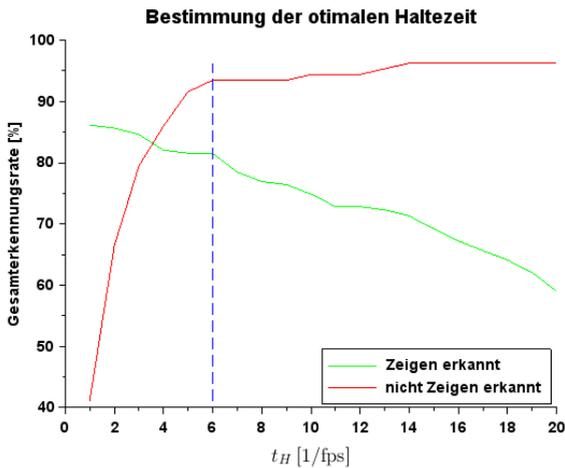
mit Äquivalent der notwendigen Haltezeit ( $t_H$ ) in [1/fps] bei 30 fps

### Auswahl der Filter

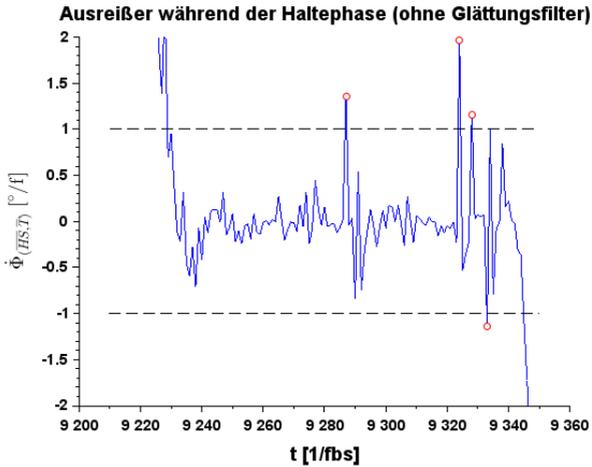
Da bedingt durch das Aufnahmesystem immer wieder Ausreißer (siehe Abb. 5.6 auf Seite 107) Einfluss auf die Identifikation der Haltephase sowie auf die Bewegungsphasen, vor allem die monoton fallende Ausleitung, haben, wurden die Daten testweise mit verschiedenen Filtern geglättet.



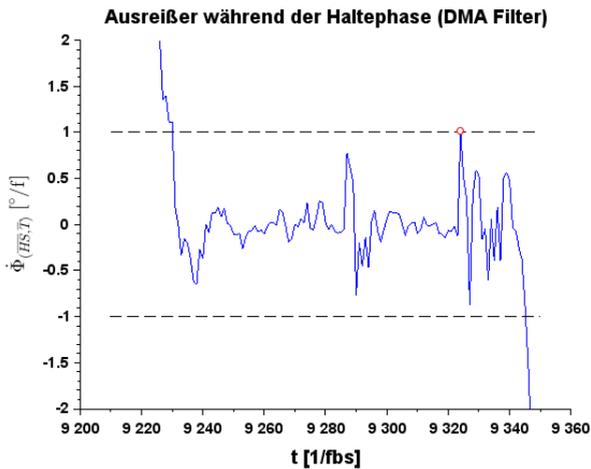
**Abbildung 5.4: Bestimmung der Gesamterkennungsrate in Abhängigkeit des Äquivalent der notwendigen Haltezeit ( $t_H$ )**  
 mit  $t_H$  in  $1/\text{fps}$  bei 30 fps  
 Optimierung der Phasendefinition mittels eines Winkels



**Abbildung 5.5: Bestimmung der Erkennungsrate (Zeigen, nicht Zeigen) in Abhängigkeit des Äquivalent der notwendigen Haltezeit ( $t_H$ )**  
 mit  $t_H$  in  $1/\text{fps}$  bei 30 fps  
 Optimierung der Phasendefinition mittels eines Winkels



**Abbildung 5.6:** Ausreißer in den Werten der Winkelbeschleunigung  $\dot{\Phi}_{(\overline{HS}, \overline{T})}$   
ohne Glättungsfilter  
mit  $t$  in [1/fps] bei 30 fps



**Abbildung 5.7:** Ausreißer in den Werten der Winkelbeschleunigung  $\dot{\Phi}_{(\overline{HS}, \overline{T})}$   
nach Glättung mit dem Double Moving Average Filter  
mit  $t$  in [1/fps] bei 30 fps

Zum Einsatz kamen:

- kein Filter
- gleitender Mittelwert 3. Ordnung
- Median-Filter (RangordnungsfILTER, Filterbreite 3)
- Double Moving Average Filter

Abb. 5.7 auf der vorherigen Seite zeigt die Auswertungsergebnisse des gleichen Datensatz wie Abb. 5.6 auf der vorherigen Seite, nachdem der Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  mit dem Double Moving Averaging Filter geglättet wurde. Von zuvor 4 Ausreißern in der potentiellen Haltephase ist noch einer geblieben, und dieser liegt mit 1,01 %f nur knapp über dem Grenzwert von 1 %f.

Wie aus Tabelle 5.1 ersichtlich, ist die Gesamterkennungsrate für „kein Filter“ und den „RangordnungsfILTER“ vergleichbar, wenn auch ungefähr 2 % niedriger als bei Verwendung des „gleitenden Mittelwert“ und des „Double Moving Averaging Filter“. Ohne Filter ist die Ablehnung der Kontrollgesten als „nicht Zeigen“ mit 95,3 % am höchsten, während die mit dem gleitenden Mittelwert geglätteten Daten mit 82.6 % die beste Erkennung der Zeigegeste ermöglichen.

Auf die Gesamterkennungsrate bezogen bringt der Double Moving Averaging Filter gegenüber des gleitenden Mittelwertes eine Änderung des Gewinns von nur 0,33 %. Im Detail ausschlaggebend für diesen Unterschied ist auch wieder die bessere Erkennungsrate der Kontrollgeste. Für die Gestenerkennung entsprechend der „Phasendefinition mittels eines Winkels“ wurde der Double Moving Averaging Filter ausgewählt, auch wenn bedingt durch die Stichprobengröße nicht sichergestellt ist, dass ein anderer Filter oder die Kombination aus mehreren Filtern bessere Ergebnisse erzielen könnten.

Tabelle 5.1: Erkennungsrate bei Anwendung verschiedener Filter

Filter	Zeigegeste erkannt			Kontrollgeste abgelehnt			Sonderfälle	Erkennungsrate
	ja	nein	Rate	ja	nein	Rate		
kein Filter	149	5	76.41 %	102	43	95.33 %	3	83.11 %
Mittelwertfilter	161	10	82.56 %	97	30	90.65 %	4	85.43 %
RangordnungsfILTER (Medianfilter)	154	10	78.97 %	97	39	90.65 %	2	83.11 %
Double Moving Averaging Filter	159	7	81.54 %	100	34	93.46 %	2	85.76 %

Die Filterauswahl wurde bei einer Fensterbreite von 3 und einer Haltezeit von 6 Frames untersucht.

Der mit 85,4 % Gesamterkennungsrate leistungsfähigste Filter, der Double Moving Averaging Filter ist in der verwendeten Implementierung in der Filterbreite nicht einstellbar, sodass auf eine nähere Analyse der Filterbreite (Ordnung) verzichtet werden kann. Diese wurde der Vollständigkeit halber durchgeführt, deren Ergebnisse finden sich im Anhang A auf Seite 153.

### **Betrachtung der Winkelgeschwindigkeit zwischen den Vektoren $\overline{HS}$ und $\overline{T}$ ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ )**

Die Bewegungsgeschwindigkeit ist in der beschriebenen „Phasendefinition mittels eines Winkels“ das Kriterium zur Definition der Phasenübergänge. Außerdem ist die Idee entscheidend, dass nach der Haltephase eine direkte Ausleitung (monoton fallende Bewegung, siehe Formel 4.14 auf Seite 95 ) folgt, um die Weiterentwicklung der Gestenerkennung auf Basis einer Phasendefinition voranzutreiben. Bei der Zeigegeste tritt die Ausleitung nur nach der Haltephase auf.

Die Optimierung der Schwellwerte der Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  zur Bewegungserkennung, zur Erkennung der Haltephase sowie zur Beschreibung der Ausleitung wird im Anhang A auf Seite 153 vorgestellt.

Die Optimierung dieser Parameter muss kritisch betrachtet werden, da sie die maximale Erkennungsrate mit einer Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  während der Ausleitung von  $< 2\%$  erreicht. Dieser Wert widerspricht den theoretischen Annahmen, sodass für die Beschreibung der Erkennung der Zeigegeste anhand der „Phasendefinition mittels eines Winkels“ auf die zuvor aus der theoretischen Betrachtung hergeleiteten Parameter zurückgegriffen wird und die Schwellwerte der Winkelgeschwindigkeiten  $\dot{\Phi}_{(\overline{HS}, \overline{T})}$  unverändert gelassen werden.

Tabelle 5.2: Auswertung der Probandenversuche

Algorithmus	1 Zeigegeste			
	ausgeführt	korrekt erkannt	Erkennungsrate	
Zeigen Statisch	195	79	40,5 %	
Phasendefinition mittels zweier Winkel	195	93	47,7 %	
Phasendefinition mittels eines Win- kels	definiert	195	159	81,5 %
	optimiert	195	170	87,2 %

Algorithmus	2 Kontrollgeste			
	ausgeführt	korrekt erkannt <sup>1</sup>	Erkennungsrate	
Zeigen Statisch	107	69	64,5 %	
Phasendefinition mittels zweier Winkel	107	100	93,5 %	
Phasendefinition mittels eines Win- kels	definiert	107	100	93,5 %
	optimiert	107	97	90,7 %

Algorithmus	Sonderfälle	3 Gesamt			
		ausgeführt	korrekt erkannt <sup>2</sup>	Erkennungsrate	
Zeigen Statisch	7	302	148	40,0 %	
Phasendefinition mittels zweier Winkel	1	302	193	63,9 %	
Phasendefinition mittels eines Winkels	definiert	2	302	259	85,7 %
	optimiert	2	302	267	88,4 %

1) nicht als Zeigegeste erkannt 2) Zeigegeste erkannt, Kontrollgeste abgelehnt

## 5.4 Kontrollexperiment zum Einfluss einer Drehung um die Körperlängsachse

Im Anschluss an die Überprüfung und Optimierung der Parameter galt es, den Einfluss des Winkels zwischen Probanden und optischer Achse der Kinect® 2 zu betrachten. Wie bereits in der Beschreibung der Versuchsdurchführung erwähnt, wurde der Winkel bei den Teilversuchen 1 und 2 nicht gesondert betrachtet und die Probanden wurden lediglich aufgefordert, annähernd frontal zur Kamera zu stehen und in Richtung der Kamera zu sehen.

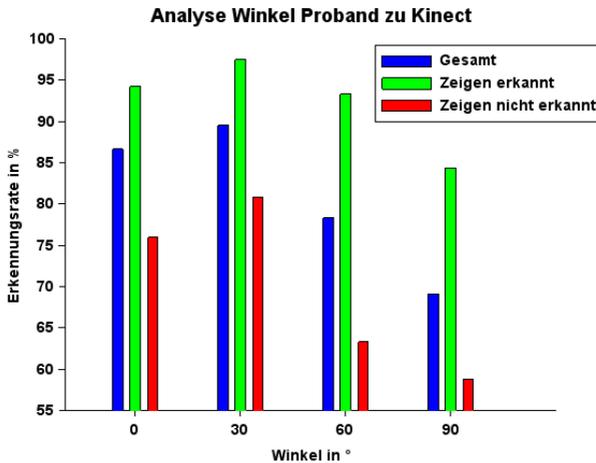
Mit drei Probanden wurden die zwei beschriebenen Teilversuche erneut viermal durchgeführt. Dabei mussten sich die Probanden jeweils um 30° im Uhrzeigersinn um die Körperlängsachse von der Kameraachse wegdrehen. Da im zweiten Versuchsteil die Objekte vom Versuchsleiter so ausgewählt wurden, das jeweils der gesamte Bereich vor und neben der Kamera (ca. 270°) erfasst wurde, wurde auf die Betrachtung der anderen Quadranten verzichtet.

Ziel dieser im folgenden nur noch als „Kontrollexperiment“ bezeichneten Versuche ist, den Einfluss der Orientierung der Probanden zur Kameraachse der Kinect® 2 auf die Erkennungsrate zu untersuchen und die „Phasendefinition mittels eines Winkels“ auch für die veränderten Aufnahmebedingungen zu verifizieren. Dazu wurden mit den Versuchsteilnehmern insgesamt 255 Gestensequenzen aufgenommen, davon 140 Zeigegesten und 115 Kontrollgesten. Die als Sonderfälle aussortierten Gestensequenzen sind alle den Zeigegesten zuzuordnen (siehe Abb. 5.9 auf Seite 113).

Tabelle 5.3: Verteilung der Ausgeführten Gesten im Winkelversuch

Winkel	Zeigegesten	Kontrollgesten	Gesamtanzahl
0°	35	25	60
30°	40 (1)*	26	67
60°	30	30	60
90°	32 (2)*	34	68

\* Zahl in Klammern: Anzahl der Sonderfälle



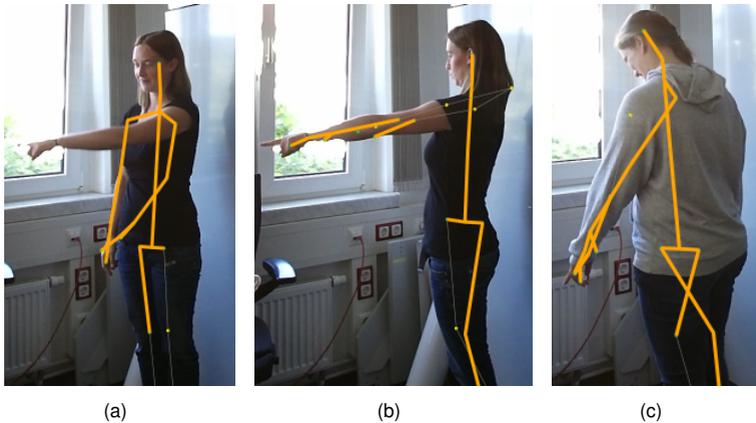
**Abbildung 5.8: Untersuchung des Winkeleinflusses zwischen der optischen Achse der Kinect® 2 und der Frontalebene des Probanden auf die Erkennungsrate**

## Auswertung

Im Kontrollexperiment konnte die Erkennungsrate von über 85 % für eine überwiegend frontale Ausrichtung des Probanden zum Sensor bestätigt werden (siehe Tab. 5.4 auf Seite 115). Bei 30° Abweichung der Probanden von der Kameraachse wurde die Erkennungsleistung mit 89,6 % sogar noch übertroffen. Steigt die Winkelabweichung weiter an, sinkt die Erkennungsrate bis auf 69,1 % bei 90°. Dabei ist auffällig dass die Zeigegeste auch bei 90° noch mit 84,4 % gut erkannt wird, die Abgrenzung der Kontrollgesten aber auf 58,8 % zurückgeht (siehe Abb. 5.8).

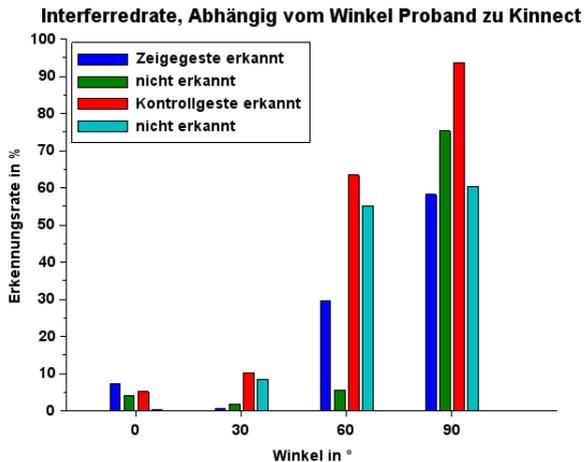
Die Aufzeichnungssoftware erlaubt es, die von der Microsoft® -Bibliothek bereitgestellte Information über die Zuverlässigkeit des Skelettpunktes mit auszuwerten. Dazu wird zu jedem Datenpunkt dessen Status aufgezeichnet. Auch für als gestört (interferred) markierte Datenpunkte versucht die Microsoft® -Bibliothek annähernd korrekte Werte zu liefern. Das dafür verwendete Körpermodell ist allerdings nicht explizit deklariert (siehe Abb. 5.9 auf der nächsten Seite).

Der Algorithmus erkennt bei den drei dargestellten Sonderfällen jeweils, dass beide Arme eine Zeigegeste ausführen.



**Abbildung 5.9: Erkannte Sonderfälle der Winkeluntersuchung, Darstellung der gestörten Bildpunkte**

- (a) 30°, Tracking des zeigenden Arms verloren (b) 90°, Gestörte Punkte am Oberarm  
 (c) 90°, Überwiegend Erkennung des Rückens, fehlerhafte Trennung der Arme



**Abbildung 5.10: Untersuchung des Winkleinflusses zwischen der optischen Achse der Kinect® 2 und der Frontalebene des Probanden auf die Rate der „gestört Skelettunkte“ (Interferedrate) bei der Gestenerkennung**

Für die dargestellten Situation auf Abb. 5.9(a) und Abb. 5.9(c) auf der vorherigen Seite hat die Microsoft®-Bibliothek nicht auf den Verlust des Skeletttrackings hingewiesen, während die Gestensequenz Abb. 5.9(b) nur aus als gestört gemeldeten Daten besteht. Für den menschlichen Beobachter sind allerdings alle Gestensequenzen interpretierbar.

Wie auf Abb. 5.10 auf der vorherigen Seite dargestellt, steigt die Rate der „gestörten“ Skelettpunkte (ausgewertet werden nur die verwendeten Skelettpunkte) ab  $60^\circ$  überproportional an. Auch die Erkenntnis, dass die Erkennungsrate der Kontrollgeste geringer ausfällt als die der Zeigegeste, lässt sich anhand der Störungsrate erklären.

Tabelle 5.4: Ergebnis Winkelversuch

Winkel	1 Zeigegeste		
	ausgeführt	korrekt erkannte	Erkennungsrate
0°	35	33	94.3 %
30°	40	39	97.5 %
60°	30	28	93.3 %
90°	32	27	84.4 %

Winkel	2 Kontrollgeste		
	ausgeführt	korrekt erkannt <sup>1</sup>	Erkennungsrate
0°	25	19	76.0 %
30°	26	21	80.8 %
60°	30	19	63.3 %
90°	34	20	58.8 %

Winkel	Sonderfälle	3 Gesamt		
		ausgeführt	korrekt erkannt <sup>2</sup>	Erkennungsrate
0°	0	60	52	86.7 %
30°	1	67	60	89.6 %
60°	0	60	47	78.3 %
90°	2	68	47	69.1 %

1) nicht als Zeigegeste erkannt 2) Zeigegeste erkannt, Kontrollgeste abgelehnt



# Kapitel 6

## Diskussion

Die vorangegangenen zwei Kapitel stellen mathematische / technische Definitionen der Zeigegeste und ihre Evaluation an Experimenten mit insgesamt 15 Probanden und 302 Gestensequenzen dar. Sowohl bei der theoriebasierten Entwicklung der Gestenbeschreibung wie auch bei der Optimierung der Parameter sind verschiedene Probleme aufgetreten, die im ersten Teil dieses Kapitels näher betrachtet werden sollen.

Darüber hinaus konnte der im Experiment schlecht bewertete Ansatz, die Statische Definition (vergl. Kap. 4.1 auf Seite 78) für ein mobiles Online System zum Heben und Tragen von Lasten bereits erfolgreich umgesetzt werden. Im zweiten Teil des Kapitels werden daher auch die Unterschiede zwischen der offline ausgewerteten Gestenbeschreibung und dem online funktionierenden Ansatz beurteilt und deren Zusammenführung diskutiert.

### 6.1 Phasendefinition

Für den Nutzer können Gesten, abhängig von dem repräsentierten Inhalt dynamisch oder statisch sein. Basierend auf dem Modell von Kendon [2004] sind nicht die Gesten dynamisch oder statisch, sondern nur einzelne Gestenphasen. Kombinieren lassen sich die Nutzerinterpretation einer Geste und die Phase mit dem entsprechenden Informationsgehalt.

Basieren auf der Modellannahme „Phasendefinition mittels eines Winkels“ (siehe Abb. 4.12 auf Seite 96) sind diese Phasen und Phasenübergänge für die Zeigegeste sogar nur mit je einem Parameter zu beschreiben. Dazu muss unterschieden werden in:

- **statische Phasen**, definiert über Schwellwerte des Winkels Hand-Schulter - Torso ( $\Phi_{(\overline{HS}, \overline{T})}$ ),
- **dynamische Phasen**, definiert über das Vorzeichen der Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ ) (Das Vorzeichen bildet die die Bewegungsrichtung der Hand ab.)
- **Phasenübergänge**, die durch einen Schwellwert der Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ ) zu definieren sind,
- *optional eine Zeitdauer*, die den erfolgreichen Abschluss einer jeden Phase bestätigt. Die entsprechende Prüfung kann zum Zeitpunkt des Phasenüberganges erfolgen, siehe dazu Formel 4.13 auf Seite 94

Dieser theoretische Ansatz ist in dem Modell „Phasen der Gestenerkennung“ (siehe Abb. 4.12 auf Seite 96), das heißt auch in der „Phasendefinition mittels eines Winkels“, noch nicht konsequent umgesetzt. Deutlich wird das in der „dynamischen“, gemeinsamen Einleitung und Hauptphase, sowie dem sich anschließenden Phasenübergang in die Haltephase.

Die entsprechenden Definitionen würden neu wie folgt lauten:

$$\dot{\Phi}_{(\overline{HS}, \overline{T})} > 1^\circ / f \quad (6.1)$$

#### Zwangsbedingung der Definition einer gemeinsamen Einleitung und Hauptphase

für eine NEUE Phasendefinition mittels eines Winkels

$$|\dot{\Phi}_{(\overline{HS}, \overline{T})}| < 1^\circ / f \quad (6.2)$$

#### Zwangsbedingung der Definition des Phasenübergangs von der Haupt- in die Haltephase

für eine NEUE Phasendefinition mittels eines Winkels

$$\Phi_{(\overline{HS}, \overline{T})} > 1^\circ / f \quad (6.3)$$

#### Zwangsbedingung der Definition der Haltephase

für eine NEUE Phasendefinition mittels eines Winkels

Eine Implementierung incl. Evaluierung dieses Ansatzes steht aktuell noch aus und ist nicht mehr Bestandteil dieser Abhandlung.

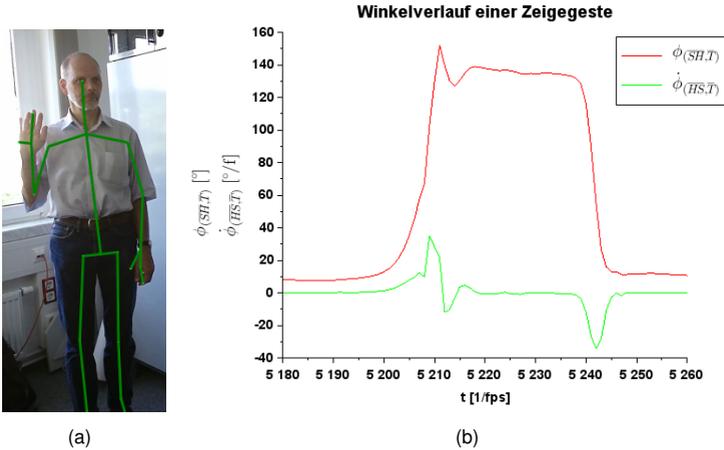
## 6.2 Versuchsdesign

Durch die praktische Anwendung der Zeigegeste zur Verortung von Objekten (z.B. Transportkisten) im Rahmen des parallel verlaufenden Projektes Klara, wurde im Versuchsdesign kein Wert auf exakte und reproduzierbare Bedingungen gelegt. In der alltäglichen Mensch-Mensch-Kommunikation ist es auch nicht notwendig, besondere Bedingungen bei der Ausführung der Gesten einzuhalten. Wie mehrfach beschrieben liegt der primäre Focus auf der Erkennung von natürlichen ausgeführten Gesten. Das Hebe- und Tragesystem als mobil agierendes System findet in den späteren Einsatzbereichen keine definierten Umgebungen mit einheitlichem Hintergrund oder genau ausgerichteten Objekten vor.

Darüber hinaus liegt durch die Verwendung der Microsoft® eigenen Bibliotheken zur Bestimmung der Skelettpunkte die Bildverarbeitung und damit die Genauigkeit der zur Erkennung der Zeigegeste verwendeten Skelettdaten nicht in der Hand des Autoren (siehe Abb. 5.9 auf Seite 113). Das Aufnahmesystem incl. des hinterlegten Skelettmodells ist somit als Black Box anzusehen, deren Ausgabe nur durch die aufgezeichneten Farbbilder überprüft werden kann. Ein alternatives Aufzeichnungssystem wie z.B. Qualisys® könnte in definierten Umgebungen für die Skelettpunkte eine höhere Genauigkeit ermöglichen, wäre aber bei mobilen Anwendungen nicht praktikabel.

Zeitgleich zu den Untersuchungen zur Zeigegeste wurden von Suzaly et al. [2015] und Wenzel et al. [2016] mit Kinect® 1 und Kinect® 2 Untersuchungen zur Körperhaltung beim Heben und Tragen von Lasten durchgeführt. Diese beiden Arbeiten nutzen zur Klassifizierung der Körperhaltung einen Entscheidungsbaum. Die Entscheidungen werden vergleichbar des hier vorgestellten Ansatzes auf Basis berechneter Skelettwinkel getroffen. Leider stellte sich bei diesem Projekt heraus, dass die mit der Kinect® 1 (Open Ni-Bibliothek) bestimmten Schwellwerte für die Weiterentwicklung auf Kinect® 2 (Microsoft® -Bibliothek) nicht einfach übertragbar sind. Sie müssen mit der neuen Soft- / Hardware-Kombination neu ermittelt werden. Somit erscheint es angebracht, die Schwellwerte mit einem möglichst einfachen und durch den Nutzer durchführbaren Kalibrierungstest anpassen zu können.

Vor diesem Hintergrund und bedingt durch die parallel laufende Arbeit an dem mobilen System ist die Entscheidung gefallen, den Aspekt der vollständi-



**Abbildung 6.1: Ungenaue Einweisung des Probanden in die Versuchsaufgabe**

mit Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ), Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ ),  
 $t$  in [1/fps] bei 30 fps

(a) Echtbild (b) Winkerverlauf

gen Reproduzierbarkeit der Versuches einer möglichst robusten Erkennung unterzuordnen. Daher sind auch die für den Versuchsteil 2 notwendigen Zielobjekte frei zwischen den Versuchen verändert, bzw. spontan vom Versuchsleiter ausgewählt worden. Eine getrennte Betrachtung der Verortung der Zielobjekte in Höhe und Raumwinkel war ursprünglich angedacht, wurde aber auf Basis der praktischen Erkenntnisse vorerst verworfen und sollte in weiterführenden Arbeiten genauer untersucht werden.

Während der Versuche hat sich aber auch ein weiterer Fehler im Versuchsdesign bzw. in der Versuchsdurchführung gezeigt. Die Probanden wurden wie bereits beschrieben nur kurz über ihre Aufgabe informiert und sollten ihr persönliches Verständnis der Aufgabe „Zeigen“, vor allem im Teilversuch 1, einbringen. Wie in Abb. 6.1(a) zu erkennen, kann die Aufgabe „Zeige rechts!“ aber missverstanden werden. Der Proband hat eine emblematische Zeigegeste nach Fricke [2007] (siehe Abb. 2.12(a) auf Seite 39) ausgeführt, da er die Aufgabe „Zeige rechts!“ als Zeige die rechte Hand bzw. „Aufzeigen mit Rechts“ interpretiert hat. Der Proband hat diese Interpretation der Aufforderung für

alle vier Zeigegesten im Teilversuch 1 angewendet. Von den vier Zeigegesten wurden durch die „Phasendefinition mittels eines Winkels“ drei richtig erkannt, da der Bewegungsablauf (siehe Abb. 6.1(b) auf der vorherigen Seite) mit dem einer „einfachen Zeigegeste“ vergleichbar ist.

## 6.3 Sonderfälle

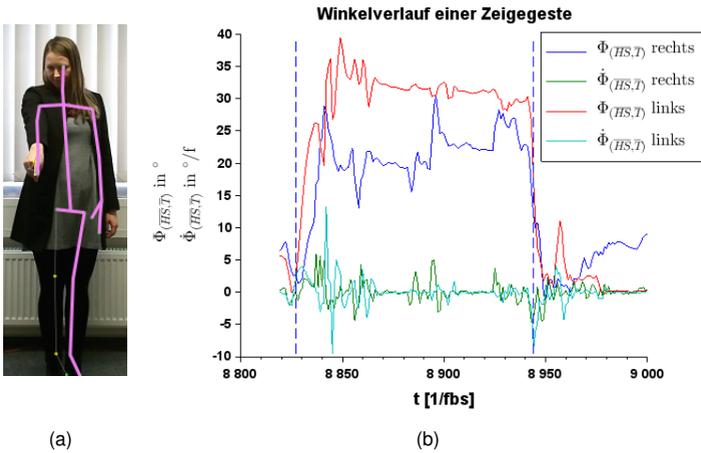
Neben den schon beschriebenen Sonderfällen im Kap. 5.4 auf Seite 112 sind zwei weitere Sonderfälle näher zu betrachten (siehe Tab. 5.2 auf Seite 110). Vorab gilt es aber zu klären, wie der Begriff „Sonderfälle“ in dieser Untersuchung definiert wird. Ein Sonderfall ist ein Fall, der von den vier Ergebniskategorien abweicht. Diese Abweichungen können ihrerseits in drei Kategorien eingeteilt werden:

1. Der Proband hat mit der einen Hand gezeigt, und es wurde die andere erkannt.
2. Es wurde erkannt, dass beide Hände eine Zeigegeste ausführen. Dass der Proband zwei konträre oder auch zwei gleichgesinnte Zeigegesten ausführt, wurde im Rahmen der Auswertung nicht akzeptiert und ist jeweils den Erkennungsfehlern zugeschlagen worden.
3. Es wurden in einer Sequenz keine Geste, weder eine Zeigegeste noch eine Kontrollgeste ausgeführt.

Von den zwei beobachteten Sonderfällen fällt je einer in die Kategorien eins bzw. zwei.

Der Erkennungsfehler der Kategorie 1 (siehe Abb. 6.2 auf der nächsten Seite) ist bei näherer Betrachtung der Winkelverläufe (siehe Abb. 6.2(b) auf der nächsten Seite) durch die extremen Ausreißer geprägt, die bereits aus der als Black Box zu betrachtenden Microsoft® (MS)-Bibliothek übernommen werden. Diese Ausreißer sind auch nicht durch die durchgeführte Filterung zu beseitigen. Erkennungsfehler der Kategorie 1 können im Gegensatz zu Erkennungsfehlern der Kategorie 2 im Onlinesystem später nicht automatisch erkannt werden, da keine Regel für die automatische Auswertung verletzt wird.

Demgegenüber stehen die Sonderfälle der Kategorie 2, „Beide Arme für eine Zeigegeste erkannt“. Der vom System als Sonderfall ausgewiesene Fall ist zwar als fehlerhaft für die Erkennungsrate gewertet worden, hätte aber eigentlich als korrekt (True Positiv, Zeigen richtig erkannt) gewertet werden müssen (siehe Abb. 6.3 auf der nächsten Seite). Dass diese Variante des



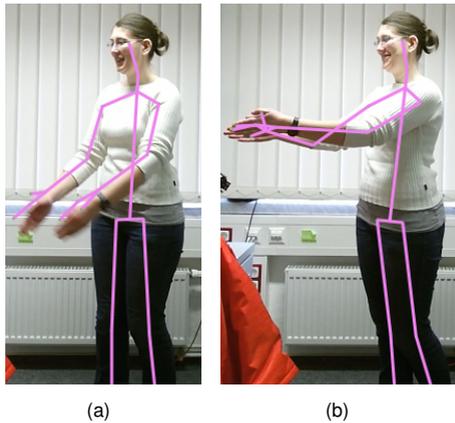
(a)

(b)

**Abbildung 6.2: Sonderfall der Kategorie 1:****Es wurde mit rechts gezeigt und der linke Arm erkannt**

mit Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ), Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ ),

$t$  in [1/fps] bei 30 fps (a) Echtbild (b) Winkelverlauf

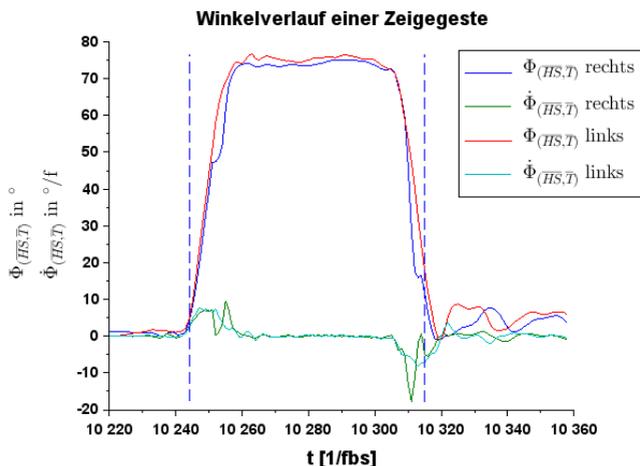


(a)

(b)

**Abbildung 6.3: Sonderfall der Kategorie 2, beide Arme benutzt und richtig erkannt**

(a) Einleitungsphase (b) Haltephase



**Abbildung 6.4: Winkelverlauf für beide Arme,**  
 Sonderfall der Kategorie 2, beide Arme benutzt und richtig erkannt  
 mit Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ), Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$   
 und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ ),  
 $t$  in [1/fps] bei 30 fps

Zeigens mit beiden Armen verbreitet ist konnte der Autor selbst mehrfach beobachten, so z.B. während eines Kongressvortrages.

Betrachtet man die Winkelverläufe für beide Arme, so kann unschwer erkannt werden, dass sich die Bewegung zwischen dem rechten und linken Arm nur unwesentlich unterscheidet (siehe Abb. 6.4). Der Beginn der Einleitung (Beginn Gestenphase) ist für beide Arme einheitlich bei frame 10244 erkannt worden, während der rechte Arm 3 frames (0.1 Sekunden) später die Gestenphase abschließt als der linke.

Für eine Onlineerkennung stellt dieser Sonderfall nur bedingt ein Problem dar. Grundsätzlich muss für die Erkennung einer Geste jedes Bild bzw. jede Sequenz für die rechte und linke Körperhälfte getrennt untersucht werden, solange der Erkennungsalgorithmus nicht auch zweihändige Gesten wie „Klatschen“ identifizieren soll. Somit ist eine Regel des Entscheidungsbaumes vergleichbar der vorgestellten Auswertung zu implementieren, die eine gleichzeitig Erkennung rechts und links als fehlerhaft behandelt und dem Nutzer meldet.

## Gestenausführung mit zwei Händen

Wie bereits bei Baudel und Beaudouin-Lafon [1993] beschrieben, führt zwei-händige Gestenausführung sowie Gesten von mehreren Personen zu Konflikten. Der Fall der Gestenausführung mit zwei Händen tritt nicht selten auf (siehe Abb. 6.3 auf Seite 122 sowie Abb. 6.5 auf der nächsten Seite) und muss daher näher untersucht und weiter aufgeteilt werden.

Dazu sind grundsätzlich drei Teilprobleme zu unterscheiden:

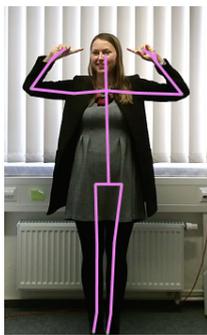
### 1. **Gegensinnig Ausführung von Gesten mit beiden Armen.**

Eine Erkennung der Intention des Gesten ausführenden Menschen ist ausschließlich unter Einbeziehung weiteren kontextabhängigen Wissens möglich. Eine Variante für die Zeigegeste ist, unter Einbeziehung der Zielerkennung das Ziel zu bestätigen. Ein drittes Beispiel aus der Mensch-Mensch-Kommunikation: der Ausführende zeigt zeitversetzt erst mit dem einen Arm auf das Quellobjekt und anschließend mit dem zweiten Arm auf die Zielposition.

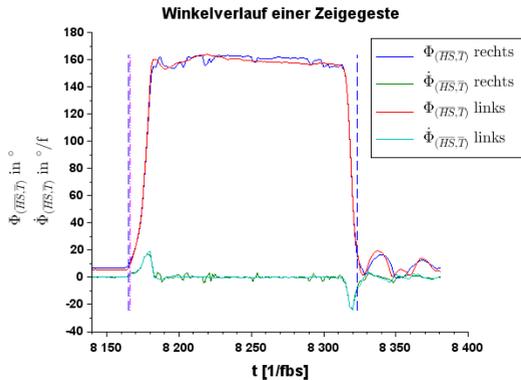
### 2. **Gleichsinnige, gleichrichtige Ausführung von Gesten mit beiden Armen.** Eine Erkennung dieser Gesten ist für das technische System bei entsprechendem Entscheidungsbaum gut möglich, bis zu dem Umstand, dass die doppelte Ausführung die eigentliche Geste verstärkt bzw. die Erkennung durch Redundanz robuster machen kann.

### 3. **Gemeinsame Ausführung von Gesten mit beiden Armen.** Die Kontrollgeste „Klatschen“ stellt eine solche Variante dar. Sie kann nur automatisch erkannt werden, wenn beide Arme gleichzeitig eher gegenläufige Bewegungen ausführen und so die Hände zusammenführen.

Die für die Erkennung von „konträr“ und „gleichsinnig“ ausgeführten Gesten vorgeschlagene Kontext- bzw. Zielanalyse könnte auch für die Erkennung von Sonderfällen der Kategorie 1 verwendet werden. Auch dort könnte das zusätzliche Wissen die Ablehnung der Zeigegeste als Fehlerkennung ermöglichen.



(a)



(b)

### Abbildung 6.5: Kontrollgeste mit zwei Händen

mit Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ), Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ ),  
 $t$  in [1/fps] bei 30 fps

(a) Echtbild (b) Winkerverlauf

## Kontrollgesten

Im Rahmen der vorgestellten Auswertung wurden die Kontrollgesten nicht detailliert betrachtet. Auf eine Erkennung von Sonderfällen bei der Ausführung der in dieser Arbeit ausgewählten Kontrollgesten wurde bewusst verzichtet, da mit der Weiterentwicklung der Gestenerkennung aus den Kontrollgesten regulär zu erkennende Gesten werden. Diese müssen dann regulär evaluiert werden.

Auf eine Besonderheit der auf Abb. 6.5(a) vorgestellten Kontrollgeste soll dennoch eingegangen werden. Die Probanden wurden zu der Gestenausführung mit der Aufgabe „Zeigen auf sich selbst“ bzw. einen „Vogel zeigen“ aufgefordert. Es handelt sich also um eine reguläre Zeigegeste, die durch den charakteristischen Winkerverlauf (siehe Abb. 6.5(b)) korrekt beschrieben wird. Die Geste wurde korrekt als Zeigen (auf sich selbst) erkannt, da es sich aber um eine Kontrollgeste handelt, wurde die Sequenz als „False Negativ“ in der Beurteilung aufgenommen. Eine detaillierte Auswertung dieses Sonderfalles hat ergeben, dass nur 31 % dieser Kontrollgesten als „Zeigen“ erkannt werden,

wohingegen 69 % als „nicht Zeigen“ verworfen werden und damit positiv in die Beurteilung der Kontrollgeste einfließen.

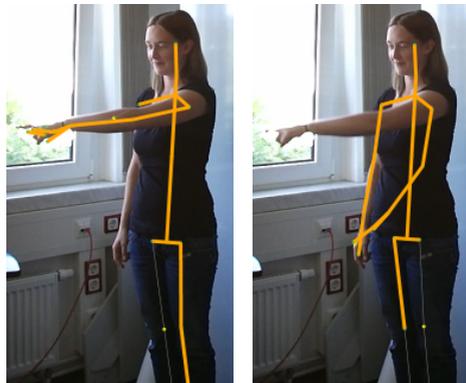
Warum die Geste teilweise als „Zeigen“ erkannt bzw. als „nicht Zeigen“ abgelehnt wird, kann verschiedene Ursachen haben.

- So kann das Zeitverhalten bei der Ausführung die entscheidende Rolle spielen, wenn die minimale Haltezeit nicht erreicht wird.
- Wenn die Ausleitung nicht monoton fallend verläuft, wird die Geste mit „nicht Zeigen“ abgelehnt.
- Eine weitere Variante stellt auch in diesem Fall wieder das korrekte Tracking des Handpunktes dar. Durch den geringen Abstand zum Kopfpunkt können Hand- und Kopfpunkt verschwimmen und damit die Gesamterkennungsleistung negativ beeinflussen.

Im Rahmen einer Weiterentwicklung muss auch diese Kontrollgeste individuell implementiert werden, um eine korrekte Zuordnung zu gewährleisten.

## **Kontrollexperiment zum Einfluss einer Drehung um die Körperlängsachse**

Über die im aktuellen Abschnitt diskutierten Sonderfälle hinaus wurden auch bei dem Kontrollexperiment bereits die dort aufgetretenen Sonderfälle mit angesprochen. Alle drei aufgetretenen Sonderfälle entsprechen der Kategorie 2, „Geste mit beiden Händen erkannt“ (siehe Abb. 5.9 auf Seite 113). Diese Sonderfälle sind korrekt von der „Phasendefinition mittels eines Winkels“ abgelehnt worden. In allen drei Fällen ist das Tracking der relevanten Skelettpunkte nicht korrekt. Die Sonderfälle bei 90° Abweichung zur optischen Achse (Abb. 5.9(b) und Abb. 5.9(c), (S. 113) sind gekennzeichnet durch die „Interferred“ (gestört) Markierung für z.B. den Schultergelenkpunkt. Die gestörten Skelettpunkte werden durch die Microsoft® -Bibliothek parallel zu den sichtbaren Skelettpunkten ausgegeben. Somit erkennt der Algorithmus korrekt je eine Zeigegeste mit rechts und links und muss den Datensatz dem Nutzer als fehlerhaft melden und verwerfen.



(a) Frame 15143

(b) Frame 15144

**Abbildung 6.6: Betrachtung des Sonderfalles „beide Arme“ erkannt**

bei einer 30° Drehung um die Körperlängsachse

Erkennungsfehler mit Trackingverlust, wie er bei 30° aufgetreten ist, (siehe Abb. 6.6(b)) treten mit der Kinect® 1 noch wesentlich gehäufiger auf und lassen sich durch bessere Trackingalgorithmen vermeiden. Ärgerlich ist, dass diese Trackingfehler nicht durch den aktuellen Algorithmus direkt erkannt werden können. Anders sieht es aus, wenn die komplette Gestensequenz und damit auch das vorhergehende Bild betrachtet wird (siehe Abb. 6.6(a)). Auf dem Bild ist sofort erkennbar, dass durch Verdeckung teilweise nicht alle Skelettpunkte durchgängig erkannt werden und somit auch Sprünge der einzelnen Punkte (hier z.B. der Handgelenkpunkt Hg) auftreten können. Diese Sprünge können zur Verbesserung der Erkennungsrate zukünftig mit betrachtet und schwellwertbasiert ausgewertet werden.

## 6.4 Winkelbetrachtung

Unter der Überschrift „Winkelbetrachtung“ werden im Folgenden vier verschiedene Fragestellungen abgehandelt:

1. Das Quadrantenproblem bei der Berechnung eines Winkels  $\Phi$  über den  $\arccos$ .
2. Die Winkelabweichung in der Ruhephase.

3. Der Untersuchungsumfang des Kontrollexperiment zum Einfluss einer Drehung um die Körperlängsachse.
4. Die unterschiedlichen Erkennungsleistungen des Kontrollexperiments für eine Verdrehung um die Körperlängsachse von  $0^\circ$  und  $30^\circ$ .

### arccos-Funktion

Der eingeschlossene Winkel  $\Phi$  zwischen zwei Vektoren  $\bar{a}$  und  $\bar{b}$  berechnet sich nach Formel 6.4 .

$$\cos(\Phi) = \frac{\bar{a} \cdot \bar{b}}{|\bar{a}| \cdot |\bar{b}|} \quad (6.4)$$

#### Eingeschlossener Winkel $\Phi$ im Raum [Zeidler et al., 1996]

Die beiden trigonometrischen Funktionen  $y = \sin x$  und  $y = \cos x$  sind durch unendliche Reihen zu definieren (vgl. z.B. [Zeidler et al., 1996, S. 56]). Sie konvergieren für alle komplexen Zahlen  $x$ .

Die inverse trigonometrische Funktion zu

$$y = \cos x, 0^\circ \leq x \leq 180^\circ$$

besitzt für jede reelle Zahl  $y$  mit  $-1 \leq y \leq 1$  genau eine Lösung mit

$$y = \arccos x, -1 \leq x \leq 1.$$

Durch die Definitionseinschränkung für den Bereich  $0^\circ \leq x \leq 180^\circ$  sind die Ergebnisse der Umkehrfunktion  $\arccos$  nur im ersten und zweiten Quadranten definiert. Sollte es durch die Bewegung des Probanden zu einem überschreiten der Quadrantengrenzen  $0^\circ$  und  $180^\circ$  kommen, wird dies nicht automatisch erkannt.

Bezogen auf die untersuchten Winkelbeziehungen zwischen dem Vektor Hand-Schulter (HS) und dem Vektor Torso ( $\bar{T}$ ) gilt es nun, die Quadrantengrenzen zu untersuchen. Nach Hochschild [2005, S. 112] beträgt das Bewegungsausmaß der Abduktion der Schulter  $180^\circ$ . Praktisches Zeigen hat sich für den Nutzer eher als Peilen über die Hand erwiesen (vergl. Kap. 6.6 auf Seite 132). Daher scheiden Zeigeziele in Verlängerung des Körpers eigentlich aus, denn der Mensch wird zum Zeigen auf entsprechende Objekte seinen Standort verändern. Somit kann diese Quadrantengrenze bei  $180^\circ$  vernachlässigt werden.

Anders sieht es hingegen bei der Adduktion des Schultergelenkes aus. Nach Hochschild [2005, S. 122] beträgt das Bewegungsmaße ca.  $40^\circ$  -  $50^\circ$ . Da in der „Neutral-0-Stellung“ (die auch bei der Gestenausführung angenommen wird) wegen des Rumpfes keine Adduktion möglich ist (Arm würde am Körper anliegen), ist die Adduktion von einer Flexion überlagert. Für die Standardvermessung vor dem Körper beträgt diese je nach Körpergewicht  $45^\circ$  oder gar  $90^\circ$ . Für den in der Arbeit betrachteten Winkel Hand-Schulter - Torso ( $\Phi_{(\overline{HS}, \overline{T})}$ ) kann daher auch die Quadrantengrenze der  $\arccos$ -Funktion bei  $0^\circ$  vernachlässigt werden. Sollte es dennoch in den Randbereichen des am Körper anliegenden Armes zu einem „Quadrantenübergang“ kommen, ist durch die im folgenden Abschnitt geführte Schwellwertdiskussion die Abweichung für die Auswertung irrelevant, da der kritische Winkel laut Formel 4.10 auf Seite 93 nicht überschritten wird.

### **Schwellwert Winkel zwischen den Vektoren $\overline{HS}$ und $\overline{T}$ ( $\Phi_{(\overline{HS}, \overline{T})}$ ) in der Ruhephase**

Direkt aufbauend auf der Fragestellung mit der  $\arccos$ -Funktion für den Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ) stellt sich die Frage nach dem im Zuge der Auswertung ermittelten Schwellwert.

Der theoretische Ansatz geht von einer Parallelität zwischen ruhendem Arm und Oberkörper aus (siehe Abb. 4.9 auf Seite 89). Zwei parallele Geraden schneiden sich theoretisch nicht, der eingeschlossene Winkel  $\Phi_{(\overline{HS}, \overline{T})}$  beträgt  $0^\circ$ . Da wir aber am Menschen nie ideale Bedingungen vorfinden werden, wurde eine Abweichung von der Parallelitätsbedingung und somit ein nicht zu überschreitender Schwellwert für die Ruhephase angenommen.

Bei den parallel durchgeführten Untersuchungen für das Hebe- und Tragesystem wurde für die statische Definition der Winkel von  $20^\circ$  als Schwellwert für das Erreichen des Gestenraumes ermittelt.

Durch die Verbindung des Winkels zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ) mit der Winkelgeschwindigkeit zwischen den Vektoren  $\overline{U}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{U}, \overline{T})}$ ) wurde unter der Annahme, dass die Ruhephase durch die Parallelitätsbedingung und die anwendungsbezogene Definition von Bewegung (Formel 4.7 auf Seite 90) genauer beschrieben wird, ein Winkel von weniger als  $10^\circ$  festgelegt.

Nach einer ersten Optimierung (mit einem eingeschränkten Datensatz) wurde der Winkelwert auf  $11^\circ$  geändert. Auf eine weitere Optimierung der Parameter für die „Phasendefinition mittels zweier Winkel“ wurde abschließend verzichtet, da die „Phasendefinition mittels eines Winkels“ wesentlich bessere Erkennungs-

raten lieferte und diese Erkennungsraten auch durch Optimierung nicht zu erzielen sind.

Für die „Phasendefinition mittels eines Winkels“ wurde der in der Praxis bewährte Winkel zum Erreichen des Gestenraumes als Startwert für die Optimierung ausgewählt und mit der geringen Abweichung von  $1^\circ$  bestätigt. Diese Erkenntnis lässt die Schlussfolgerung zu, dass die Einschränkung durch die  $\arccos$ -Funktion auch an der Quadrantengrenze bei  $0^\circ$  zu vernachlässigen ist.

Wie in Kap. 6.2 auf Seite 119 schon angesprochen, sind die Schwellwerte auch durch die unbekanntenen, in den Microsoft® (MS)-Bibliotheken hinterlegten Modellannahmen nicht auf andere Systeme übertragbar und müssen für jede weitere Generation bzw. jedes vergleichbare System neu bestimmt werden.

## Kontrollexperiment - Winkelbetrachtung

Auf das Kontrollexperiment muss erneut kurz eingegangen werden. Wie bereits bei der Beschreibung des Kontrollperiments erwähnt, wurde die Untersuchung nur im I Quadranten durchgeführt. Da der Mensch „achsensymmetrisch“ zu seiner Längsachse ist, kann auf die Untersuchung des III und IV Quadranten bezogen auf die Hochachse verzichtet werden. Für die betrachtete Fragestellung sind keine anderslautenden Ergebnisse zu erwarten.

Der II Quadrant wurde nicht untersucht, da der Mensch immer mit den Armen von sich weg in Blickrichtung zeigt und diese Richtung dann von der Kamera nicht mehr erfasst wird. Wie auf den Abb. 5.9(c) auf Seite 113 erkennbar ist, sind die Verdeckungen und Überlagerungen bereits bei  $90^\circ$  so groß, dass bei größeren Winkelabweichungen keine auswertbaren Ergebnisse mehr erwartet werden.

## Kontrollexperiment - Erkennungsleistung

Abschließend gilt es noch, die Fragestellung zu betrachten, warum bei dem Kontrollexperiment die Erkennungsrate bei  $30^\circ$  mit 2,9 % besser ausfällt als bei  $0^\circ$ . Da die 2,9 % bei dem Stichprobenumfang von 60 bzw. 67 ausgeführten Gesten gerade einen Unterschied von einer erkannten Geste abbilden, sollte von einem Interpretationsversuch dieser 2,9 % abgesehen werden. Betrachtet man über die Erkennungsrate hinaus den Anteil an Interferred (gestörten) Skelettpunkten, so sollen auch aus den Werten für die Unterschiede zwischen  $0^\circ$  und  $30^\circ$  keine weitergehenden Schlüsse gezogen werden.

## 6.5 Fehlerdiskussion

Auf eine Abschätzung der Fehlerfortpflanzung wird im Rahmen dieser Arbeit verzichtet, da der alle anderen Fehlerquellen dominierende Hauptfehler in der Bewegungsanalyse aus der Weichteilverschiebung in der Größenordnung  $\pm 1$  cm resultiert (vergl. [Sati et al., 1996; Karduna et al., 2001; Cappozzo et al., 2005]). Die aufwändige Kompensation derartiger Fehler, (vergl. z.B. [Vimieiro et al., 2007]) ist insbesondere unter dem Aspekt der „Echtzeitfähigkeit“ für die hier bearbeitete Fragestellung nicht zielführend.

Wie bereits verschiedentlich angesprochen, sind die vom ausgewählten Sensorsystem gelieferten Messdaten bereits stark fehlerbehaftet, insbesondere nicht abbildungstreu (siehe Abb. 6.6 auf Seite 127). Eine interindividuelle Vergleichsmessung, die eine Verifikation bzw. eine Bestimmung der Messunsicherheit des Sensorsystems ermöglicht, wurde im Rahmen der Arbeit nicht durchgeführt. Die von Bonnechere et al. [2014] durchgeführten Vergleichsuntersuchungen zwischen einer Kinect<sup>®</sup> 1 und einem Vicon 8 XT40s-System haben für die in der Arbeit betrachteten Winkel Schulter und Ellenbogen Winkelabweichungen bis zu  $10^\circ$  ergeben (siehe Tab. 6.1). Die zu erwartende Größenordnung stimmt mit den eigenen Erkenntnissen zum gestreckten Arm in der statischen Definition überein (vergl. Formel 4.2 auf Seite 81 ).

Tabelle 6.1: Vergleich „markerbasiertes“ und „markerloses“

### Bewegungserkennungssystem

Kinect<sup>®</sup> 1 mit 30 fps vs. Vicon 8 XT40s mit 60 fps

Bonnechere et al. [2014, S. 595f.]

	Bewegungsbereich (MLS)	Differenz (MLS - MBS)	R <sup>2</sup> (T-test)
<b>Abduktion</b>	111° ( $\pm 17^\circ$ )	0.4 % (2,5°)	0.98
<b>Schulter</b>	109° ( $\pm 18^\circ$ )	0.9 % (9,1°)	0.80
<b>Flexion</b>	127° ( $\pm 11^\circ$ )	8 % (9°)	0.33
<b>Ellbogen</b>	126° ( $\pm 14^\circ$ )	8 % (10°)	0.54

Darüber hinaus führt ein Mensch die zu erkennende Geste nach seinem persönlichen Verständnis durch (vergl. Kap. 2.3 auf Seite 29), so dass für die Probandenversuche eine Reproduzierbarkeit in dieser Hinsicht nicht gewährleistet werden kann.

Maßgeblicher Ansatz der gesamten Arbeit ist die Bestimmung von „Skelett winkeln“ basierend auf den durch die Microsoft<sup>®</sup> -Bibliotheken ermittelten

„Skelettpunkten“. Um die Entscheidung über die Erkennung der Zeigegeste zu treffen, werden diese apriori stark fehlerbehafteten Daten anhand von Schwellwerten beurteilt. Als Einstiegswerte für die durchgeführte Optimierung der einzelnen Parameter wurden theoretisch ermittelte Werte verwendet. Diese wurden wie dargestellt wider Erwarten mit kleinsten Abweichungen bestätigt.

Zusammenfassend wäre eine scheinbar vollständigen Fehlerbetrachtung in mathematischer Fundierung in Anbetracht der vorausgesetzten Schwellwerte für die Einflussgrößen und der beschriebenen sonstigen Voraussetzungen (Restriktionen) für erkannte/nichterkannte Gesten auch irreführend und ist damit obsolet. Es gehört zum Charakter dieser Arbeit, dass nicht die Genauigkeit von Messwerten, sondern die Genauigkeit der Erkennung als Maßstab gesetzt ist.

## 6.6 Zusammenführung von Online- / Offline-Version

Erkenntnisse aus der praktischen Umsetzung der statischen Gestenbeschreibung wurden bereits mehrfach in den vorherigen Abschnitten des aktuellen Kapitels angesprochen. Diese werden nur teilweise im Folgenden wieder aufgegriffen, zusätzliche Elemente, die zur Umsetzung in ein Online-System notwendig sind, sollen hier ebenfalls angesprochen werden.

### Phase Einleitung

Die Frage, wer für das technische System der verantwortliche Nutzer, also der Ausführende ist, lässt sich durch Kontextinformation beantworten, z.B. weil die Person die Stelle des Lageristen besetzt. Alternativ könnte die Anmeldung und Erkennung des verantwortlichen Nutzers über ein zusätzliches Identifikationselement wie z.B. einen Sender erfolgen.

Umgesetzt wurde eine dritte Variante, die ausschließlich über eine in dem Phasenablauf abbildbare und erkennbare Geste bzw. Körperpose definiert wird. Dazu wird die optionale Haltephase im Anschluss an die Einleitung (pre-stroke hold) definiert (Formel 6.5 auf der nächsten Seite). Der Ausführende nimmt mit der ausführenden Hand (Hg) eine Position oberhalb des Schultergelenkes (Sg) ein (siehe Abb. 4.3 auf Seite 84).

$$Hg_{(Y)} \leq Sg_{(Y)} \quad (6.5)$$

**Definition der Anmeldepose (per-stroke-hold)  
im Vorfeld der statischen Gestendefinition**

über dem Y- (Vertikal-)Anteil des Handgelenkpunktes (Hg) und des Schulterpunkte (Sg)  
unter Berücksichtigung des Koordinatensystems der Kinect<sup>®</sup>, x,y Ursprung links oben in der Bildebene

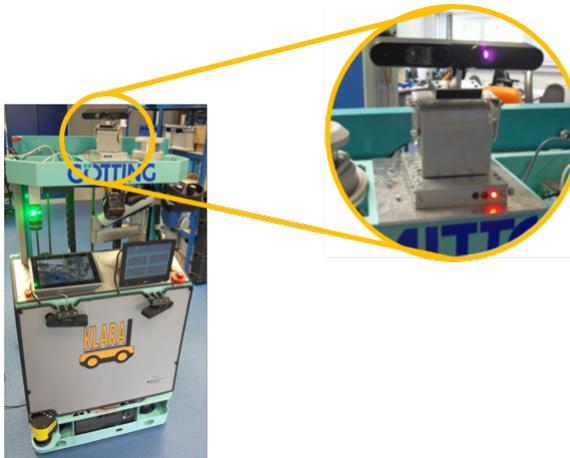
Diese Anmeldepose teilt die umgesetzte Sequenz exakt in die bereits von McNeill [1992] beschriebenen fünf Phasensegmente auf. Im Gegensatz zu den weiterentwickelten Phasendefinitionen sind in dieser Umsetzung die Informationen für das Assistenzsystem in den beiden Haltephase (pre- und post-stroke hold) kodiert. Die dynamischen Phasen (Bewegungen) können in der Implementierung durch den Ausführenden beliebig durchlaufen werden. Durch den definierten Ablauf „Anmeldung“ gefolgt vom „Zeigen“ ergeben sich zwangsläufig charakteristische Winkelverläufe (siehe Abb. 4.4 auf Seite 84). Aus den charakteristischen und wiedererkennbaren Winkelverläufen resultierte die erste dynamische Gestendefinition, die Dynamische Beschreibung mittels linearer Regression (vergl. Kap. 4.2 auf Seite 83).

Für die Umsetzung im Rahmen des Assistenzsystems Klara wurden die Winkelverläufe allerdings nicht weiter betrachtet.

## Systemrückmeldung

Besonderes Augenmerk ist bei Online-Systemen auf eine erkennbare Rückmeldung zu legen. Wie im Kapitel Mensch-Mensch-Kommunikation (vergl. Kap. 2.2 auf Seite 23) beschrieben ist der Kommunikationsprozess ein „implizites Kreislaufmodell“ (siehe Abb. 2.5 auf Seite 28) mit Reaktion und Feedback [Kotler, 2003]. Dieses Kreislaufmodell wird auch in den „Sieben Stadien des Handelns“ von Norman [1988] aufgegriffen. So beschreibt Norman in der Herleitung den Zyklus des Handelns, dass die Ausführung, der Versuch das eigene Ziel in der Welt zu erreichen, durch die Auswertung, den Vergleich zwischen dem was passieren sollte und dem was passiert ist, zu einem geschlossenen Kreislauf wird. Dieser Kreislauf steht auch hinter dem Modell, dass den Geltungsbereich der VDI/VDE 3850-1 und damit eine entsprechend gestaltete Mensch-Technik-Schnittstelle beschreibt (siehe Abb. 2.1 auf Seite 22).

Wie kann nun für eine Gestensteuerung die Systemrückmeldung aussehen? Untersuchungen zur Beantwortung dieser Fragestellung wurde im Umfeld der hier entwickelten und praktisch umgesetzten Gestensteuerung nur am Rande durchgeführt. Das Unterstützungssystem Klara nutzt für Objekterkennung



**Abbildung 6.7: Das Erkennungssystem ist in Bereitschaft**  
 Implementierung im Rahmen des Projektes Klara mit der ASUS® Xtion PRO LIVE  
 (technisch vergleichbar zur Kinect® 1) als Gestensensor

ein zweistufiges System, das der menschlichen Kommunikation nachempfunden wurde. Wenn das Unterstützungssystem positioniert wurde, und eine Zeigegeste des Nutzers erwartet, wird das Erkennungssystem aus der Ruheposition in die Arbeitsposition gehoben; der Nutzer wird, wie in Abb. 6.7 dargestellt, angeschaut.

Für die Nutzererkennung und -identifikation (Einleitung) sowie für die Körperpose „Zeigen“ (gemäß Formeln 4.2 und 4.3 auf Seite 81) ist die Rückmeldung z.Z. über verschiedenfarbige LED's realisiert.

Wenn die Haltezeit abgelaufen ist und das Zielobjekt vorhergesagt werden kann, wird als Rückmeldung für den zweiten Schritt ein Bild des potentiellen Zieles auf einem Touchmonitor präsentiert. Dort kann der Nutzer die Auswahl bestätigen oder verwerfen.

Soll zukünftig die „Phasendefinition mittels eines Winkels“ in den Alltags-einsatz überführt werden, zeichnet sich eine neue Fragestellung für die Nutzerrückmeldung ab. Während die statische Definition bei geeigneter Rückmeldung dem Nutzer ermöglicht, seine Körperhaltung zu korrigieren und darüber hinaus die notwendige Haltezeit zu verharren, sind diese Korrekturmöglichkeiten in der „Phasendefinition mittels eines Winkels“ nicht gegeben. In der „Phasendefinition mittels eines Winkels“ ist neben der Haltezeit und dem

Gestenraum vor allem die Ausleitung das bestimmende Element. Diese muss im Anschluss an eine ausreichend lang bemessene Haltephase eingeleitet und durchgeführt werden. Eine passende Systemrückmeldung für die einzelnen Phase wurde damit schwer bis unmöglich.

Auf der anderen Seite ist durch die mit 6 Frames (0,2 Sekunden) extrem kurze Haltezeit für eine sinnvolle Rückmeldung viel zu kurz. Wenn zur Erkennung einer natürlichen Zeigegeste die „Phasendefinition mittels eines Winkels“ zum Einsatz kommt, ist die Systemrückmeldung nur nach der vollständigen Ausführung der Zeigegeste möglich.

Daher ist zu hinterfragen, ob ein Praxiseinsatz mit der aktuell erreichten Erkennungsrate von über 85 % bereits sinnvoll ist. Beachtet man, dass der Mensch meist flexibler und anpassbarer auf die Technik reagiert als die Technik auf Sonderfälle in der menschlichen Bewegungsausführung, ist nach Meinung des Autoren eine prototypische Implementierung zur weiteren Verbesserung angeraten.

## **Einsatz von Filtern**

Sowohl in der Parametrisierung der Zeigegesten wie auch in der Auswertung der Versuche wurde deutlich, dass die Daten der Microsoft® (MS) Kinect® nur sinnvoll verwendet werden können, wenn Ausreißer in den Winkelverläufen durch den Einsatz von Filtern geglättet werden. Der Einsatz von Mittelwert- oder Rangordnungsfilter in Online-Systemen ist immer mit einem entsprechenden Fehler behaftet, da die Filter nicht den vollständigen Datensatz zur Verfügung haben, sondern nur die bisher aufgezeichneten Daten. Somit kann der Filter nur basierend auf den vergangen Datenpunkten eine Glättung vornehmen.

Berücksichtigt man allerdings, dass anhand der vorgeschlagenen „Phasendefinitionen“ die Entscheidung „Zeigen“ oder „nicht Zeigen“ nur erfolgreich nach Beendigung eins vollständigen Gestenzykluses getroffen werden kann, muss für die Übertragung in ein Online-System die Frage nach den Filtern nicht weiter gestellt werden, da zur Auswertung bereits die gesamte Gestensequenz vorliegt.

Evaluiert werden muss darüber hinaus auch, ob es sinnvoll ist, einzelne Phasen bereits online während der Ausführung zu prüfen. So bietet sich an, neben den Entscheidungen Beginn der Gestenausführung (Formel 4.11 auf Seite 94 ) und Wiedereintritt „Ruhephase“ (Formel 4.10 auf Seite 93 ) auch die Haltephase (Formel 4.12 und Formel 4.13 auf Seite 94 mit  $t > 6f$ ) online zu überwachen. Sollte die notwendige Haltezeit von 6 frames nicht erreicht werden, könnte das System den Nutzer frühzeitig auf eine wahrscheinliche

Fehlerkennung hinweisen oder nach Überschreiten der minimalen Haltezeit einen Hinweis geben, das bereits ausreichend lange gehalten wurde.

Ob diese Hinweise und die Echtzeitüberwachung bestimmter Phasen im Zusammenhang mit einer Erweiterung auf zusätzliche Gesten weiterhin sinnvoll bleiben, kann zum jetzigen Zeitpunkt nicht abgeschätzt werden.

## **Zeigen mit zwei Armen**

Wie im Kap. 6.3 auf Seite 124 bereits angesprochen, ist das Zeigen mit zwei Armen nicht ungewöhnlich. Für die in der Arbeit vorgestellten Phasendefinitionen stellen diese Sonderfälle allerdings noch einige Hindernisse bereit. Wie sind diese in der Übertragung auf ein Online-System zu behandeln bzw. können diese Sonderfälle im Alltagseinsatz sogar an Bedeutung gewinnen? Dazu sollen drei Varianten diskutiert werden:

1. Zeitlichen Versatz bestimmen, wenn zwei potentiell relevante Objekte zu erkennen sind.
2. Vergleich Zielobjekt mit Kontextwissen (Bedingung: Erkennung des Zielobjektes)
3. Vergleich der kleinsten Abweichung für den Schnittpunkt (wie bei KLARA für die Zielbestimmung realisiert).

### **Zeitlicher Versatz**

Wie bereits im vorherigen Abschnitt angesprochen werden von einem Online-System die Zeitpunkte „Beginn Gestenausführung“ und „Wiedereintritt in die Ruhephase“ überwacht. Werden vom Nutzer bei einem System ohne Nutzeridentifikation bzw. Anmeldung die Zeitpunkte „Beginn Gestenausführung“ mit der rechten und linken Hand zeitversetzt erreicht, so ist eine Entscheidung durch das System zu treffen, ob der führende Arm (wahrscheinliche Variante) oder folgende Arm die relevante Information übermittelt. Wenn die Kommunikation so angelegt ist, das zwei Orte adressiert werden, z.B. Behälter und Ablageplatz, dann ist der führende Arm auf die Quelle und der folgende auf die Senke ausgerichtet.

### **Identische Zielrichtung**

Sind beide Arme in die identische Richtung auf ein Ziel ausgerichtet, so ist die seitenrichtige Hand (z.B. rechte Hand beim Zeigen nach rechts) auf Basis der Zielerkennung (Zeigen ist Peilen) die Vorzugsvariante.

### Zielobjekterkennung

Wenn durch alternative Informationsquellen z.B. durch Kommissionierungslisten oder durch das Einsatzspektrum des Gestenerkennungssystems bereits bekannt ist, auf welches Zielobjekt durch den Ausführenden gezeigt wird, sollte diese Information durch das System zur Verifikation der Zeigegeste verwendet werden. Werden zwei konträre Zeigegesten erkannt, so kann unter Verwendung einer Zielerkennung (siehe folgender Abschnitt) und der Kenntnis über das erwarteten Objekt die richtige der beiden Zeigegesten ausgewählt werden.

Eine entsprechende Objekterkennung hätte die Zeigegestenerkennung des parallel laufende Forschungsprojektes Klara trotz der Einschränkungen durch die Kinect<sup>®</sup> 1 und „statischer Gestenbeschreibung“ noch deutlich robuster gemacht. Die Objekterkennung ist nicht Gegenstand der hier vorgestellten Arbeiten und wird bereits ausführlich erforscht (vergl. z.B. [Papazov et al., 2012] ).

### Kleinste Abweichung

Neben dem Einsatz einer Objekterkennung kann auch eine andere Methode zur Unterscheidung von beidseitigen Zeigegesten Verwendung finden, die Minimierung der Abstandskquadrate. Basierend auf den Bilderkennungsbibliotheken für die Kinect<sup>®</sup>-Sensoren ist ein Modus verfügbar, der zu jedem Bildpunkt die entsprechenden Tiefenkoordinaten ausgibt. Durch die Ungenauigkeiten der Kinect<sup>®</sup> durchsticht der aus den bekannten „Skelettpunkten“ Kopf - Hand berechnete Zielvektor nicht genau an der berechneten Koordinate die Tiefenebene sondern nur annähernd. Ähnlich wird es sich auch mit den zu bestimmenden Zielkoordinaten der Zeigegeste beider Arme verhalten. Somit besteht am Ende eine Differenz zwischen dem auf Basis des Zeigevektoren berechneten Zielpunkt und den von der Kinect<sup>®</sup> ermittelten Zielpunkten für die rechte und linke Hand. Die Entscheidung über die zeigende Hand kann anhand der kleinsten Abweichung getroffen werden.

Mathematisch heißt das , auf Basis der Gerade (g) Kopf - Hand kann für jeden Punkt der Gerade (P) ein Koordinatentripel  $P_x, P_y$  und  $P_z$  bestimmt werden. Wird nun über die Kinect<sup>®</sup> ein Tiefenbild (B) aufgenommen, kann in diesem für jeden Punkt Q mit den Koordinaten  $Q_x = P_x$  und  $Q_y = P_y$  die Tiefenkoordinate  $Q_z$  bestimmt werden. Die kleinste Quadratische Differenz  $d = (P_z - Q_z)^2$  wird bei Klara für die Zielerkennung eingesetzt. Wird die Quadratische Differenz für die rechts  $d_r$  und links  $d_l$  ausgeführte Zeigegeste getrennt bestimmt, ist die Entscheidung über die zeigende Hand anhand der geringeren Abweichung zu treffen.

Für die zwei zuletzt vorgestellten Entscheidungsvarianten „Zielobjekterkennung“ und „kleinste Abweichung“ ist die im Folgenden kurz angesprochene

Zielerkennung notwendig. Diese ist aber nicht Bestandteil der vorgestellten Arbeiten.

## Zielerkennung

Zeigen ist für den Menschen die Möglichkeit, ein Objekt im Raum zu verorten. Dazu muss neben dem beschriebenen und technisch anspruchsvollen Erkennen der Zeigegeste auch das Zielobjekt erkannt werden. Dazu werden im folgenden grundsätzliche Betrachtungen angesprochen und anschließend zwei vollkommen unterschiedliche technische Ansätze diskutiert.

Allgemein muss die Frage gestellt werden: „Was ist denn Zeigen?“

Eine erste Antwort findet sich bei Fricke [2007, S. 59]. „Die [...] Zeigegeste [...] lokalisiert eine bestimmte Entität in Bezug auf einen Raumpunkt, der sich aus der Verlängerung der durch die Origo und den Arm bzw. die Hand markierten Geraden ergibt.“ Im Projekt Klara wurde diese Frage für die durch das Assistenzsystem erkannte Zeigegeste betrachtet [Nowack et al., 2015a]. Es wurden 40 Zeigegesten analysiert. Dazu wurden mit der Kinect® 1 die „Skelettpunkte“ Kopf, Schulter und Hand näher betrachtet. Wie in Abbildung 6.8 auf der nächsten Seite dargestellt ist mit 65 % zu 40 % der Vektor Kopf-Hand bei der Zielerkennung genauer. Daher lässt sich sagen, Zeigen ist Peilen über die Hand. Nur so lässt sich erklären warum der Mensch nicht nur mit gestrecktem sondern auch sehr häufig mit gebeugtem Arm zeigt.

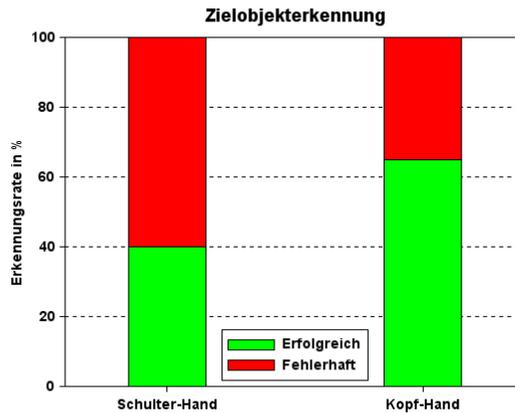
Für die technische Zielerkennung gilt es darüber hinaus die Frage zu diskutieren, „Wie wird denn das durch den Zeigevektor definierte Ziel erkannt?“ Dazu sind zwei komplett unterschiedliche Situation zu betrachten:

- Bekannte Ziele bei ortsfester Installation.
- Unbekannte Ziele bei mobiler Installation.

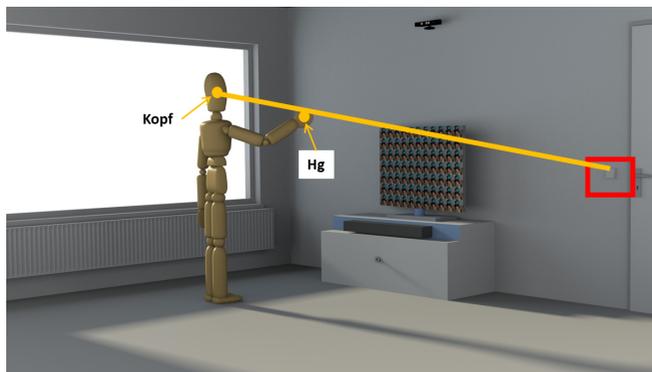
**Bekannte Ziele bei ortsfester Installation** bedeutet für die Zielerkennung bzw. Zielidentifikation:

- Der Gestensensor ist ortsfest, z.B. an einer Wand montiert.
- Die Zielobjekte befinden sich ebenfalls an definierten Positionen. Sie sind z.B. in Vorrichtungen eingespannt. Diese Variante wurde bereits 2013 von Fraunhofer IOSB und BMW vorgestellt [Schick und Sauer, 2013].

Bei dieser Variante kann das Ziel bzw. die Zielposition ohne weitere Information aus der erkannten Zeigegeste abgeleitet werden. Da die „Skelettpunkte“



**Abbildung 6.8:** Vergleich Zielerkennung, Zeigevektor Schulter-Hand / Kopf-Hand [Nowack et al., 2015a]



**Abbildung 6.9:** Zielerkennung bei ortsfester Installation mit Punkten Handgelenkpunkt (Hg), Kopf und eine Zielfläche in rot



**Abbildung 6.10: Zielerkennung bei mobiler Installation**

Realisierung des Projektes Klara

Kopf und Hand bekannt sind, kann auf deren Basis ein Zielvektor bestimmt werden. Dessen Durchstoßpunkt durch das Zielobjekt, lässt sich über Vektorrechnung bestimmen, solange für die bekannten Zielobjekte nicht nur ein Zielpunkt sondern eine Zielfläche zu treffen ist (siehe Abb. 6.9 auf der vorherigen Seite).

### **Unbekanntes Ziel bei mobiler Installation**

Bei der Verwendung für ein mobiles System (z.B. die Hebe- und Tragehilfe Klara) sind keine ortsfesten Koordinaten zu erwarten. Alle Koordinaten beziehen sich auf die aktuelle Sensorposition. Für die Zielerkennung muss basierend auf der aktuellen Sensorposition die Umgebung untersucht und das Zielobjekt bestimmt werden. Im Projekt Klara kann dazu der Gestensensor geschwenkt und geneigt werden (siehe Abb. 6.10). So kann der durch Kopf und Handpunkt bestimmte Zeigevektor segmentweise abgetastet werden. In jedem Segment wird entlang des Vektors, also zu jeder  $x,y$  Koordinate, die berechnete  $Z$  Koordinate mit der vom Sensor gemessenen  $Z$  Koordinate verglichen. Würde es keine Messfehler geben, sollten beide  $Z$  Koordinaten am Zielobjekt identisch sein. Die Praxis hat aber gezeigt, dass die Koordinaten stark voneinander abweichen. Mit der Methode der kleinsten Abstandsquadrate konnte eine sehr gute Näherung zur Zielerkennung gefunden werden.

Die beschriebene Variante ist Bestandteil des Forschungsprojektes Klara und soll nicht der hier aufgeführten Arbeit zugeordnet werden. Sie soll aber eine mögliche Variante zur Zielerkennung verdeutlichen, wobei auch bei solchen Systemen über weitere Sensoren oder durch zusätzliches Kontextwissen (z.B. Karten der Fahrnavigation) die Zielobjekte berechnet werden können, wie es bereits für die ortsfesten System vorgestellt wurde.



## Kapitel 7

# Zusammenfassung

Ziel der vorliegenden Arbeit ist die Identifikation von Parametern, um mit einem einfachen und preisgünstigen Skeletterkennungssystem, wie es mit der Kinect® auf dem Markt verfügbar ist, einen Algorithmus zu entwickeln, um Freiraumgesten für die Mensch-Technik-Interaktion nutzbar zu machen. Am Beispiel der Zeigegeste konnte gezeigt werden, dass auf Basis des Phasenmodells für die sprachbegleitende Geste und der aktuell verfügbaren einfachen technischen Sensoren verschiedene „Parameter-Konfigurationen“ erstellt werden konnten, die diesen Anforderungen unterschiedlich gut gerecht werden können.

Die Akzeptanz eines solchen Systems im zukünftigen Praxiseinsatz wird hauptsächlich vom Nutzerverständnis abhängen. Im Rahmen einer Akzeptanzuntersuchung zeigte sich, dass „Zeigen“ nicht unbedingt als dynamische Bewegung, sondern eher als statische Körperpose identifiziert wird.

Zusätzliche Kontextinformation, wie sie der Mensch dank seiner kognitiven Fähigkeiten nutzen kann, finden in den hier vorgestellten Ansätzen keine Berücksichtigung, da diese teilweise nicht durch den ausgewählten Sensor erkannt werden können (z.B. die Ausrichtung der Handfläche oder eine Nutzung des Zeigefingers). Anderes Kontextwissen, z. B. welche Objekte der Nutzer auswählen wird, könnte von dem vorgestellten System genutzt werden, wurde aber in dieser Arbeit noch nicht umgesetzt. Dafür muss nicht nur das Zeigen sondern auch das Zeigeziel ausgewertet werden. Darüber hinaus müssten Objekte durch Bilderkennungsalgorithmen zusätzlich identifiziert werden.

Auf der Grundlage umfassender theoretischer Betrachtungen zur Zeigegeste und den Ergebnissen der Akzeptanzuntersuchung konnten drei Ansätze zur Erkennung ausgearbeitet werden. Die Parametrisierung der statischen Körperpose wurde im ersten funktionierenden Ansatz zur Erkennung des „Zei-

gens“ vorgenommen. Da das Erkennungssystem vor allem die großflächigen Bewegungen und kaum die Bewegung der kleinen Gliedmaßen wie Finger erkennt, muss der Algorithmus anhand der Armstellung die Geste erkennen können. Dazu wird der Ellbogen- sowie der Schulterwinkel einbezogen. Mit der statischen Definition konnte aus den Daten der Probandenversuche „Zeigen“ erkannt oder eine Kontrollgeste als „nicht Zeigen“ abgelehnt werden. Die Gesamterkennungsrate beträgt bei der „statischen Definition“ 51,3 %.

Die Überlegungen zur statischen Definition bilden darüber hinaus die Basis für alle weiterführenden Definitionen. Diese gehen vom Phasenmodell der Geste aus, und werten die vollständigen Gestensequenz aus, auch wenn die Information der Zeigegeste überwiegend in der Haltephase übertragen wird.

Die „Phasendefinition mittels zweier Winkel“ konnte unter Verwendung der Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ) und Winkel zwischen den Vektoren  $\overline{U}$  und  $\overline{T}$  ( $\Phi_{(\overline{U}, \overline{T})}$ ) eine Gesamterkennungsrate von 63,9 % realisieren. Parallel wurde auch die „Phasendefinition mittels eines Winkels“ entwickelt, die nur den Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ) auswertet und die Gesamterkennungsrate auf 85,7 % steigern konnte.

Die Entwicklungsarbeit zur Definition (Parameterkonfiguration) der „Zeigegeste“ führten zur Erkenntnis, dass eine akzeptable Umsetzung nur mit Einschränkungen gelingt:

- **Zeigegeste:** Die vorliegende Arbeit wurde generell auf die Erkennung der „Zeigegeste“ eingegrenzt. Die Erkennung von Gesten mit einem System, das bereits ein „Skelett“ und nutzbare „Skelettpunkte“ liefert, erscheint einfach, da der Teil Bilderkennung nicht mehr behandelt werden muss. Aber die Vielzahl an möglichen Gesten für das gleiche Kommando und die mögliche Variantenvielfalt bei der Gestenausführung machen die Eingrenzung auf die „Zeigegeste“ notwendig.
- **Nutzbare Skelettpunkte:** Weil die Gestenerkennung einfach und kostengünstig in die Praxis zu überführen sein soll, wurde die Kinect® als Gestensensor ausgewählt. Der Sensor hat neben dem Preis den großen Vorteil, dass die Nutzer keine besondere Kleidung oder sonstige Bedienelemente mitführen müssen. Dieses optische, markerlose System bildet dagegen auch nur eine eingeschränkte Anzahl von nutzbaren „Skelettpunkten“, vor allem im Bereich der Hand ab. Die Erkennung der Finger und Nutzung als Kontextinformation für die Zeigegeste ist nicht möglich. Systeme, die diesen Bereich mit abdecken wie die „Leap Motion“, sind für den Anwendungskontext nicht weiter nutzbar, da der Erkennungsbereich für Ganzkörperbewegungen zu klein ist.

- **Freie Sicht:** Ein optisches System erfordert freie Sicht zwischen dem zu beobachtenden Nutzer und dem optischen Sensor, bei der Kinect® der RGB-Kamera und der IR-Kamera.
- **Orientierung des Ausführenden** Die „Skelettpunkte“ sind nur sinnvoll nutzbar, wenn der Ausführende sich frontal bis max. 30° um die Körperlängsachse verdreht vor dem Sensor befindet.
- **Nutzbarer Gestenraum:** Basierend auf den Erkenntnissen von McNeill [1992] werden „deiktische Gesten“ im mittleren und oberen, peripheren bzw. extrem peripheren Raum ausgeführt. Dieser Sachverhalt konnte für die Arbeit bestätigt werden und führte konsequent zu der Festlegung, dass die Hand die Ruhephase verlässt, wenn der Winkel zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\Phi_{(\overline{HS}, \overline{T})}$ ) größer als 21° wird.

Für die Probandenversuche wurden noch weitere Einschränkungen festgelegt:

- **Vollständige Sichtbarkeit:** Der Proband soll während der Versuche vollständig von der Skeletterkennung erfasst werden. Nur so ist sichergestellt, dass fehlerbehaftete Erkennung von Segmenten der unteren Extremitäten die Beurteilung der Gesten nicht beeinflussen.
- **Nur der Proband:** Da die, für die Probandenversuche verwendete, Kinect® 2 bis zu sechs Personen im Bild erkennen und einigermaßen zuverlässig verfolgen kann, diese Fähigkeit aber für die Auswertung hinderlich ist, durfte nur der Proband im Bild sein, der auch der Ausführende ist.
- **Vereinzelung der Gesten:** Die Gesten sind von dem Probanden als Folge nacheinander auszuführen. Da die vorgestellten Algorithmen auf der Annahme basieren, dass die Gestensequenz vollständig durchlaufen wird, muss die Gestensequenz in der Ruhelage der Arme beginnen und dort auch wieder enden.

Mit Hilfe der Probandenexperimente wurden alle Schwellwerte optimiert. Als Ausgangswert für die Optimierung wurden die Schwellwerte mit Werten vorgelegt, die aus theoretischen Betrachtungen abgeleitet wurden. Sonderfälle, die während der Versuche bzw. bei der Auswertung beobachtet wurden konnten, wurde abschließend diskutiert und führten zu vielen weiteren Fragestellungen, die im Ausblick aufgeführt werden.

Nach den in der vorliegenden Arbeit ermittelten Ergebnisse kann zusammenfassend empfohlen werden, „Zeigen“ anhand der „Phasendefinition mittels eines Winkels“ zu detektieren.

Für die Hebe- und Tragehilfe war es hingegen notwendig, die statische Definition zu implementieren. Diese hat für die Praxisüberführung den Vorteil, dass nicht erst die vollständige Gestensequenz durchlaufen werden muss, bevor das System eine Rückmeldung an den Nutzer geben kann. Durch die zeitnahe Rückmeldung und die konsequente Umsetzung einer Mensch-Technik-Kommunikation mit Rückkoppelschleife kann die „statische Definition“ zur Erkennung der Zeigegeste mittels eines Technischen Systems erfolgreich angewendet werden.

## Kapitel 8

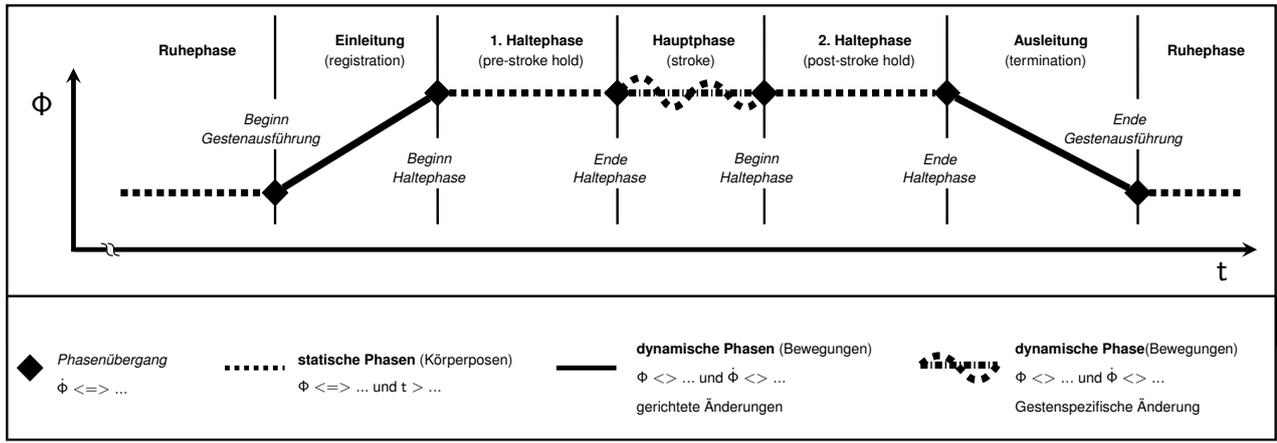
# Ausblick

Die vorliegende Arbeit lässt den Blick in die Zukunft der „Mensch-Technik-Interaktion mittels Freiraumgesten“ in zweierlei Richtung zu. Da diese Arbeit auf die Zeigegeste eingeschränkt wurde, gilt es auf der einen Seite, andere Gesten algorithmisch zu beschreiben und zu evaluieren. Auf der anderen Seite lebt die Mensch-Technik-Interaktion von der Übertragung der vorgestellten Ansätze in die Praxis. Wenn potentielle Nutzer die Vorteile dieser Interaktionsform erkennen, wird die Akzeptanz steigen und die Gestenbeschreibung wird erweitert werden.

### 8.1 Weiterentwicklung der Gestenbeschreibung

Bereits in den vorangegangenen Kapiteln wurde mehrfach darauf eingegangen, dass in der Arbeit nur „Zeigen“ erkannt oder die „Kontrollgesten“ als „nicht Zeigen“ abgelehnt werden. Um die Erkennung der Kontrollgesten zu verbessern, ist der nächste Schritt, weitere Gesten zu definieren, die mit vergleichbaren Algorithmen erkannt werden können. Dazu bietet sich das „Modell einer vollständigen Phasendefinition“ an (siehe Abb. 8.1 auf der nächsten Seite). Dieser Modellansatz soll in weiterführenden Arbeiten z.B. mit der Geste „Heranwinken“ auf seine allgemeingültige Tauglichkeit für die Gestenerkennung, aber im Besonderen die Erkennung von dynamischen Gesten, evaluiert werden.

Die Geste „Heranwinken“ wurde in der Akzeptanzuntersuchung von den Probanden als vorwiegend dynamisch erkannt. Im Phasenmodell sollte sich der Charakter der Geste in der Hauptphase abbilden lassen.



**Abbildung 8.1: Modell einer vollständigen Phasendefinition**

Muss in zukünftigen Arbeiten weiter untersucht werden.

Neben der Evaluation des „Modells einer vollständigen Phasendefinition“ an einer weiteren Geste sollte das Modell auch an der Zeigegeste evaluiert werden. Dazu müssen die bereits beschriebenen Zwangsbedingungen (Formel 6.1 bis Formel 6.3 auf Seite 118) in z.B. Scilab implementiert und mit den bestehenden Experimentaldaten evaluiert werden (vergl. Kap. 6.1 auf Seite 117).

In der Diskussion der Ergebnisse der Algorithmenentwicklung und vor allem der Versuchsauswertung wurde mehrfach angesprochen, dass durch die modellbasierte Vervollständigung der „Skelettpunkte“ diese nicht akkurat mit dem entsprechenden Körperpunkten im Echtbild übereinstimmen (siehe Abb. 6.6 auf Seite 127). Ohne die „Skeletterkennung“ vollständig durch eine Eigenentwicklung zu ersetzen, sollte die Position der relevanten „Skelettpunkte“ kontinuierlich überwacht werden. Auch wenn generell von der Verwendung der Absolutwerte der „Skelettpunkte“ abgeraten werden muss, bietet sich an, die Positionsdifferenz zwischen zwei aufgezeichneten Frames zu betrachten. Die menschliche Bewegung lässt nur eine bestimmte Verschiebung innerhalb einer 1/30 Sekunde zu. Wird dieser Wert übersprungen oder in zwei aufeinanderfolgenden Frames eine annähernd identische Strecke vor und wieder zurück gesprungen, so kann für diesen „Skelettpunkt“ unabhängig der von Microsoft® bereitgestellten Fehlererkennung zusätzlich ein Fehler erkannt werden. Ob diese Erkennung eine automatische Korrektur dieser Fehler und damit eine Verbesserung der Gesamterkennungsrate zulässt, sollte untersucht werden.

In der Informatik werden für die Gestenerkennung „selbst lernende Systeme“ eingesetzt. Dieser konzeptionelle Ansatz sollte auch die hier vorgestellten Schwellwerte verbessern, da diese dann nicht nur anhand eines Datensatzes optimiert und verbessert werden. Sie optimieren sich eigenständig während der Laufzeit des Systems. Die für die Optimierung notwendige Entscheidung, ob die Erkennung korrekt war, ist dann Aufgabe des Menschen, der die Gestenerkennung einsetzt. Eine entsprechende Möglichkeit zur Rückmeldung muss geschaffen werden. Für die Praxisanwendung taugliche zusätzliche Entscheidungshilfen können bei der Bewertung der erkannten Geste Unterstützung geben.

## 8.2 Übertragung in die Praxis

Auch oder gerade für die Übertragung in die Praxis gilt es zukünftig noch viel zu tun. So wird die statische Definition (vergl. Kap. 4.1 auf Seite 78) bereits erfolgreich für eine Hebe- und Tragehilfe eingesetzt. Um auch die in der Arbeit am besten bewertete „Phasendefinition mittels eines Winkels“ in die

Praxis übertragen zu können, müssen die entsprechenden Randbedingungen evaluiert werden.

Während die statische Definition durch die lange Haltephase eine direkte Rückmeldung an den Ausführenden zulässt, muss für die „Phasendefinition mittels eines Winkels“ der vollständige Gestenzyklus durchlaufen werden, bevor das Zeigen und damit ein Zeigeziel erkannt werden kann.

Im Kontext einer gestengesteuerten Hausautomation bietet sich allerdings die „Phasendefinition mittels eines Winkels“ an, da der Ausführende dort die Geste flüssig und intuitiv ausführen wird. Wenn die Zeigeziele, wie bereits angesprochen eine ausreichende Größe aufweisen, sollte die Übertragung der „Phasendefinition mittels eines Winkels“ für diese Anwendungen kurzfristig realisierbar sein (vergl. Kap. 6.6 auf Seite 138).

Unter Einbeziehung der Zielerkennung, sei es durch Berechnung des Zeigeziels aus dem Zeigevektor oder durch optische Verfolgung des Zeigevektors kann mit hoher Wahrscheinlichkeit eine Steigerung der Gesamterkennungsrate erreicht werden. Vor allem die Entscheidung, eine Geste als „nicht Zeigen“ abzulehnen, ist durch die zusätzliche Information zu unterstützen. So können Zeigeziele, die nicht den erwarteten Objekten entsprechen, oder nicht im festgelegten Zielfeld enden, im Vorfeld abgelehnt werden. Auch könnte, wenn ein zu erwartendes Objekt nicht direkt am Ende des Zielvektors aber im Ergebnisbild erkannt wird, der Zielpunkt entsprechend korrigiert werden.

Ein wichtiger Punkt in der Mensch-Technik-Interaktion ist die Systemrückmeldung an den Nutzer. Während diese Systemrückmeldung für die statische Definition mit der langen Haltephase einfach zu implementieren ist, ist für die „Phasendefinition mittels eines Winkels“ die Frage nach einer verständlichen Systemrückmeldung offen. Die Akzeptanz einer gestenbasierten Interaktion wird aber maßgeblich von dieser Systemrückmeldung abhängen. Für die „Phasendefinition mittels eines Winkels“, in der die Entscheidung „Zeigen“ oder „nicht Zeigen“ während der Ausleitung zu treffen ist, kann die Rückmeldung erst gegeben werden, wenn der Ausführende die Geste bereits vollständig abgeschlossen hat. Bei einer Gesamterkennungsrate von 85% bedeutet dies, dass jede achte bis neunte Geste nicht erkannt wird und vom Ausführenden wiederholt werden muss. Eine geeignete Vorhersage der Zeigegeste und des Zeigeziels ist für die Rückmeldung wahrscheinlich unverzichtbar.

Abschließendes Fazit der vorliegenden Arbeit: Die „Mensch-Technik-Interaktion mittels Freiraumgesten“ wird für interaktive Systeme in der Zukunft an Bedeutung gewinnen. Wenn der Mensch die Maschine im Arbeitsumfeld nicht mehr als den Gegenspieler ansieht, der Aufgaben wahrnimmt die er ihr abtritt, weil er sie nicht realisieren kann oder will, sondern das technische System als

Mitspieler ansieht, mit dem er sich Aufgaben teilen darf, dann werden alternative Interaktionsformen weiterentwickelt werden müssen.

Ob hierfür die Performance von Sensoren aus der Spielwelt ausreicht oder qualitativ hochwertigere Sensoren für Messsysteme zu nutzen sind, wird eher zu vernachlässigen sein. Die Fragestellungen nach technischen Beschreibungen weiterer vom Menschen intuitiv ausgeführter Gesten gilt es für zukünftige Anwendungen der automatischen Erkennung, in erster Linie zu beantworten. Ein weiteres großes Forschungsfeld ist die Übertragung der menschlichen Fähigkeiten zur komplexen Wahrnehmung, z.B. gleichzeitige Auswertung des wahrgenommenen und Identifikation eines bekannten gesuchten Zustandes in ein technisches System. Dazu könne z.B. Systeme mit unscharfer Logik (Fuzzylogik) verwendet und weiter entwickelt werden. Wann eine Erkennung mit einer der menschlichen Gestenwahrnehmung vergleichbaren Erkennungsrate und einer vergleichbaren Gestenvielfalt bereitsteht, bleibt weiterhin abzuwarten.



## Anhang A

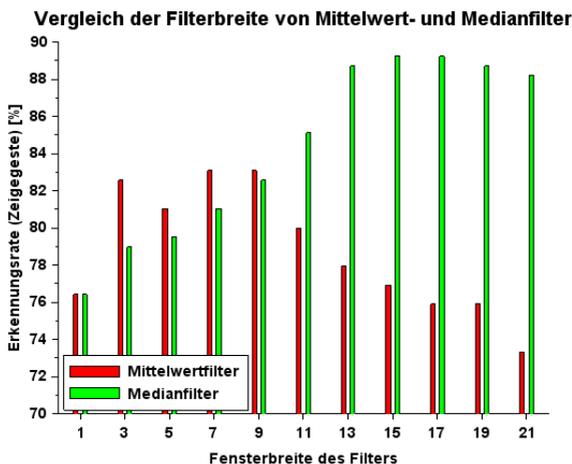
# Vollständige Optimierung

### Auswahl der Fenstergröße

Der mit 85,8 % Erkennungsrate leistungsfähigste Filter, der Double Moving Averaging Filter ist in der verwendeten Implementierung in der Filterbreite nicht einstellbar, sodass auf eine nähere Analyse der Filterbreite (Ordnung) verzichtet werden kann.

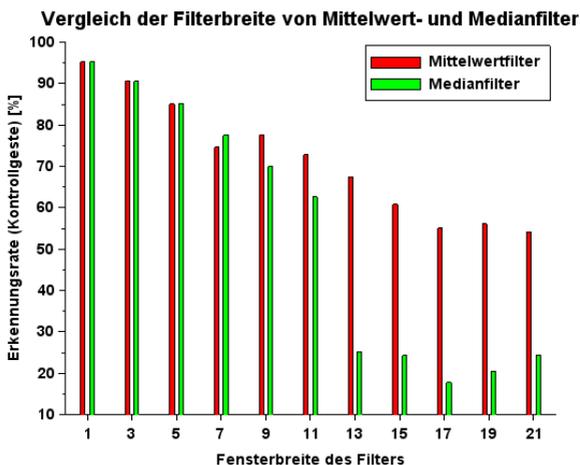
Aber da sich aber bereits im Filtervergleich eine Unterschied zwischen der Erkennungsrate der Zeigegeste und der Kontrollgeste gezeigt hat, wurden die detaillierten Werte des gleitenden Mittelwertes und des Rangordnungsfilters in Abhängigkeit von der Fensterbreite betrachtet.

Die Fensterbreite von drei für den Mittelwertfilter ergibt insgesamt den besten Kompromiss mit 85,4 % Erkennungsrate. Betrachtet man die Erkennungsrate für die Zeige- (siehe Abb. A.1 auf der nächsten Seite) und die Kontrollgeste (siehe Abb. A.2 auf der nächsten Seite) getrennt, bietet der Rangordnungsfilter bessere bzw. gleiche Ergebnisse wie der gleitende Mittelwert. Dazu müsste die Zeigegeste bei einer Filterbreite von 15 oder 17 (Erkennungsrate 89,2 %) identifiziert werden, wohingegen die Kontrollgeste als „nicht Zeigen“ mit einer Filterbreite von 1, dass heißt ohne Filtereinfluss bei einer Erkennungsrate von 95,3 % abgelehnt werden. Die Erkennungsleistung der Kontrollgesten nimmt dann linear mit der Fensterbreite bei dem gleitenden Mittelwert ab (vergl. Formel A.1 auf Seite 155). Der Rangordnungsfilter hingegen nimmt nur bis zu einer Filterbreite von 11 linear ab (vergl. Formel A.2 auf Seite 155) und macht dann einen Sprung um 37,4 % auf unter 26 %.



**Abbildung A.1: Vergleich der Fensterbreite von Mittelwert- und Medianfilter bei Ausführung der Zeigegeste**

Optimierung der Phasendefinition mittels eines Winkels



**Abbildung A.2: Vergleich der Fensterbreite von Mittelwert- und Medianfilter bei Ausführung der Kontrollgeste**

Optimierung der Phasendefinition mittels eines Winkels

$$Er_{nz} = 2,15 * win + 95,39$$

mit  $R^2 = 0,95$  (A.1)

### Lineare Abhängigkeit, Erkennungsrate Kontrollgeste als „nicht Zeigen“ korrekt abgelehnt zu Fensterbreite (gleitender Mittelwert)

mit Erkennungsrate „nicht Zeigen“ *korrekt abgelehnt* ( $Er_{nz}$ ), Fensterbreite ( $win$ )

$$Er_{nz} = -3,32 * win + 95,39$$

$win \in \{1, 3, \dots, 9, 11\}$  (A.2)

mit  $R^2 = 0,99$

### Lineare Abhängigkeit, Erkennungsrate Kontrollgeste als „nicht Zeigen“ korrekt abgelehnt zu Fensterbreite (Rangordnungsfiler)

mit Erkennungsrate „nicht Zeigen“ *korrekt abgelehnt* ( $Er_{nz}$ ), Fensterbreite ( $win$ )

## Optimierung der Winkelgeschwindigkeit zwischen den Vektoren $\overline{HS}$ und $\overline{T}$ ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ )

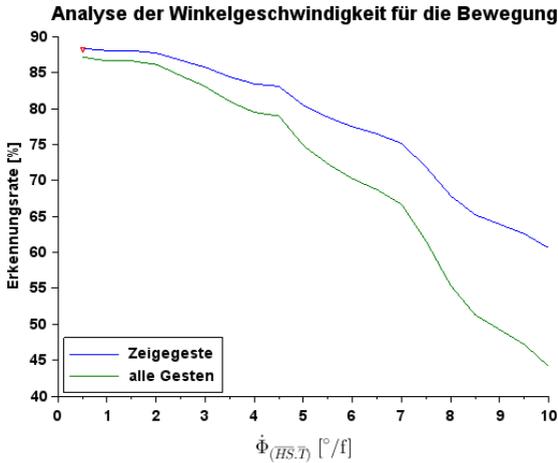
Im Rahmen der Versuchsauswertung wurden alle Parameter genauer untersucht und so optimiert, dass die Erkennungsrate maximiert wurde. Dabei wurden auch die verschiedenen Winkelgeschwindigkeiten  $\dot{\Phi}_{(\overline{HS}, \overline{T})}$  einer genaueren Betrachtung unterzogen.

Dazu wurden von  $\dot{\Phi}_{(\overline{HS}, \overline{T})}$  drei Schwellwerte unterschieden und getrennt untersucht:

- Der Beginn der Bewegung laut Formel 4.11 auf Seite 94 .
- Der Haltetremor in der Haltephase nach Formel 4.12 und 4.13 auf Seite 94.
- Die Ausleitung, welche nach Formel 4.14 auf Seite 95 streng monoton fallend sein sollte.

Für den Beginn der Bewegung und die Haltephase wurde der Wertebereich von 0,5%/f bis 10%/f variiert, für die Releasephase der Bereich von „<5%/f“ bis „<-1%/f“.

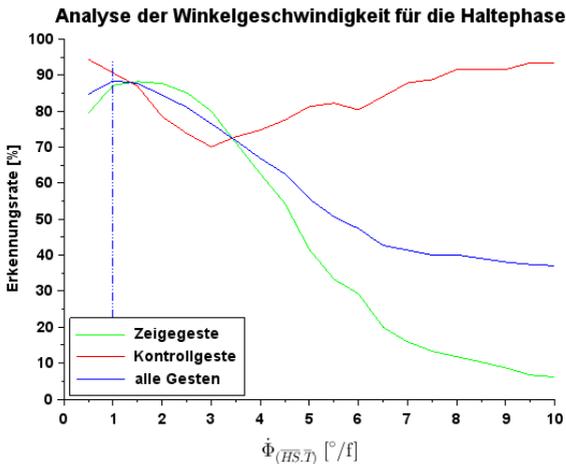
Der Beginn der Bewegung wird am besten mit einer Winkelgeschwindigkeit von 0,5%/f erkannt. Die Erkennungsrate nimmt dabei linear (Regressionsgerade siehe Formel A.3 auf Seite 157 ) ab. Die Gesamterkennungsrate ist direkt



**Abbildung A.3: Überprüfen der Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ ) zum Verlassen der Ruhephase**

mit  $t$  in [1/fps] bei 30 fps

Optimierung der Phasendefinition mittels eines Winkels



**Abbildung A.4: Überprüfen der Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ ) während der Haltephase**

mit  $t$  in [1/fps] bei 30 fps

Optimierung der Phasendefinition mittels eines Winkels

abhängig von der Erkennungsrate der Zeigegeste (siehe Abb. A.3 auf der vorherigen Seite) und unabhängig von der Erkennungsrate der Kontrollgeste, die über den Optimierungsbereich mit 90,7 % konstant bleibt.

$$Er = -3,11 * \dot{\phi} + 94,1 \quad (\text{A.3})$$

mit  $R^2 = 0.95$

### Linearer Zusammenhang zwischen Winkelgeschwindigkeit zur Detektion des Beginn der Bewegung und der Erkennungsrate

mit Erkennungsrate (Er)

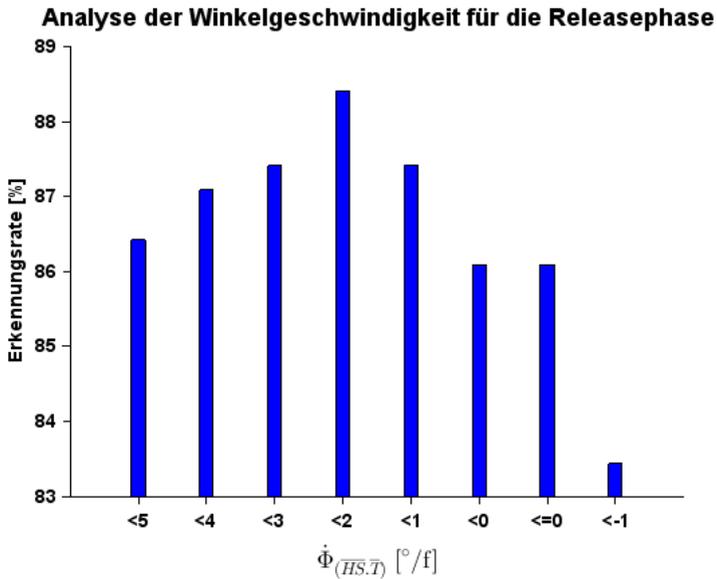
Für die Haltephase erwies sich der in der Definition postulierte Wert von „1 %f“ als bester Kompromiss für die Erkennung von Zeige- und Kontrollgeste (siehe Abb. A.4 auf der vorherigen Seite).

Die Optimierung der Winkelgeschwindigkeit während der Ausleitung hat ergeben, dass ein Schwellwert von <2 %f die beste Gesamterkennungsleistung bei den vorliegenden Daten ergibt. Dieser Schwellwert lässt sich aber durch die Theorie nicht begründen und widerspricht der zur Definition diese Parameters getroffenen Annahme.

Alle anderen Parameter (siehe Tab. A.1) bleiben trotz der möglichen Optimierung bei der Bewegungsgeschwindigkeit zur Bewegungserkennung und in der Ausleitung gleich. Die Gesamterkennungsrate nach der Optimierung liegt mit 88.4% nur 2,6% höher als mit der in Kapitel 4.4 auf Seite 93 beschriebenen Phasendefinition mittels eines Winkels.

Tabelle A.1: Parameter der Gestenerechnung nach der Optimierung

Äquivalent der notwendigen Haltezeit ( $t_H$ )	6 [1/fps] mit 30 fps
Filter	Double Moving Averages Filter
$\Phi_{(HS, \bar{T})}$	21 °
$\dot{\Phi}_{(HS, \bar{T})}$ (Bewegung)	0,5 %f
$\dot{\Phi}_{(HS, \bar{T})}$ (Haltephase)	1 %f
$\dot{\Phi}_{(HS, \bar{T})}$ (Ausleitung)	< 2 %f



**Abbildung A.5: Überprüfen der Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\dot{\Phi}_{(\overline{HS}, \overline{T})}$ ) während die Releasephase**  
mit  $t$  in [1/fps] bei 30 fps

Optimierung der Phasendefinition mittels eines Winkels

# Begriffsdefinitionen

**Anmeldepose** Die „Anmeldepose“ ist eine Körperhaltung, die vom Ausführenden während des pre-stroke hold eingenommen wird, um sich am System anzumelden und u.U. den zu erkennenden Arm festzulegen.

**Ausführender** Der „Ausführende“ ist jener Nutzer, welcher das Kommando über das durch Gesten zu steuernde System innehat. Er führt die durch das technische System zu erkennende Geste aus.

**Einleitung** Während der „Einleitung“ bewegt der Ausführende die Hand in den Gestenraum. Für die Zeigegeste ist diese Phase nicht genauer zu definieren.

**Gestenausführung** Die „Gestenausführung“ beinhaltet jene fünf Phasen, die ursprünglich von Kendon [2004] definiert wurden.

**Gestenraum** Als „Gestenraum“ wird der Bereich oberhalb der Hüfte bezeichnet, in dem der Ausführende die Geste intuitiv ausführt.

**Gestenzklus** Das vollständige Durchlaufen aller für die Gestenausführung bzw. -erkennung notwendigen Phasen.

**Haltephase** Die „Haltephase“ der Zeigegeste ist der nach dem erweiterten Modell von Kendon als post-stroke hold bezeichnete Zeitabschnitt, in dem der Ausführende die als Zeigen interpretierte Körperhaltung eingenommen hat.

**Haltezeit** Die „Haltezeit“ ist die Zeitdauer ausgedrückt in frames, während der Ausführende die Körperhaltung in der Haltephase beibehalten muss.

**Hauptphase** Die „Hauptphase“ ist nach Kendon [2004, S. 112] jene Phase der Bewegung, in welcher, der Ausdruck der Geste (was auch immer diese sein mag) erreicht wird. Für die Zeigegeste ist diese Phase nicht genauer zu definieren.

**Kontrollexperiment** Das „Kontrollexperiment“ ist ein mit drei Probanden durchgeführtes Experiment, um den Einfluss der Orientierung des Probanden zur Kameraachse der Kinect® 2 auf die Erkennungsrate zu untersuchen. Dazu drehen sich die Probanden jeweils um 30° im Uhrzeigersinn um die Körperlängsachse von der Kameraachse weg.

**Kontrollgeste** Als „Kontrollgeste“ werden in der vorliegenden Arbeit alle Gesten bezeichnet, die nicht der Definition einer Zeigegeste entsprechen.

**Marker** Der „Marker“ ist ein zusätzlicher Datensatz, der von dem Versuchsleiter manuell der Datenaufzeichnung hinzugefügt werden kann. Der Marker besteht aus einer Zeile Nullen, die bei der Auswertung zur Trennung der Gestensequenzen eingesetzt wird.

**nicht Zeigen** Mit „nicht Zeigen“ werden Gesten von den Algorithmen abgelehnt, die nicht als Zeigen erkannt werden.

**Online-System** Das „Online-System“ ist ein System, mit dem eine Geste zeitgleich oder nach Abschluss der Ausführung ausgewertet wird, um eine Systemreaktion anzuschließen.

**Ruhephase** Die „Ruhephase“ bezeichnet einen Zeitabschnitt vor der Einleitung und nach der Ausleitung, in dem die Hände des Ausführenden den Gestenraum verlassen haben.

**Skelett** Der Begriff „Skelett“ wird in dieser Arbeit für das Stickfigur des Gestenaufnahmesystems verwendet und nicht anatomisch korrekt für das menschliche Skelett.

**Sonderfälle** Unter „Sonderfällen“ werden alle Fälle verstanden, die nicht in die vier Ergebniskategorien fallen.

**statisches Halten** Das „statische Halten“ bezeichnet die Einhaltung der mathematischen Definition der Zeigepose in der statischen Beschreibung einer Zeigegeste.

**Technisches System** Unter dem „technischen System“ wird entsprechend der VDI/VDE 3850-1 [2000] ein in sich geschlossenes System aus „Bediensystem“, „Steuerung“ und „Maschine / Prozess“ verstanden.

**Zeigegeste** Die „Zeigegeste“ ist eine spezielle Geste, welche zur Verortung von Objekten dient (vergl. Kap. 3.3 auf Seite 66).

**Zeigen** Mit „Zeigen“ werden die von den Algorithmen erkannten Gesten angenommen.

**Zeigepose** Körperhaltung der Zeigegeste beim statischen Halten.

**Zeigevektor** Die Verbindung vom Kopf über die Hand zum Zielobjekt. Das Peilen über die Hand beschreibt Zielerkennung.

**Zielmarkierung** Die „Zielmarkierungen“ sind im Versuchsraum angebrachte Markierungen, auf welche die Probanden im Versuchsteil 2 zeigen. Ihre Position befindet sich im Raum vor der Frontalebene des Probanden. Für den Probanden befinden sich die Zeigeeziele im Bereich von ca.  $\pm 90^\circ$  um seine Längsachse und  $-60 / + 90^\circ$  um seine Querachse.

**Zielobjekt** Das „Zielobjekt“ ist jenes Zeigeeziel, auf welches der Ausführende mit seiner Zeigegeste verweist.



# Abbildungsverzeichnis

2.1	Geltungsbereich der Richtlinie VDI/VDE 3850 . . . . .	22
2.2	Vereinfachtes Modell der Informationsaufnahme . . . . .	24
2.3	Kommunikationsmodell von Shannon . . . . .	25
2.4	Kommunikationsmodell nach Shannon und Weaver (Interpretation von Herzeg) . . . . .	26
2.5	Kommunikationsmodell (Interpretation von Kotler) . . . . .	28
2.6	Sieben Stadien des Handelns . . . . .	30
2.7	Prozessmodell nach Rasmussen . . . . .	31
2.8	Stufen des Wahrnehmungsprozess . . . . .	32
2.9	Optische Täuschung . . . . .	33
2.10	Phasen einer Geste nach Kendon . . . . .	36
2.11	Ausrichtung der Handfläche beim Zeigen bzw. Heranwinken . . . . .	38
2.12	emblematische Zeigegesten . . . . .	39
2.13	Bewegungserkennungssysteme . . . . .	42
2.14	Stickfigure und Skelettpunkte von Kinect® 1 und 2 im Vergleich . . . . .	44
2.15	Anwendungsrahmen für Gebrauchstauglichkeit . . . . .	46
2.16	zeitliche Phasen eines Interaktionsschritts . . . . .	47
2.17	Zeitliche Abhängigkeit von Interaktionsfehlern . . . . .	49
2.18	Intuitive Sitzverstellung . . . . .	51
3.1	Gebärden für Ortsangaben und Zeigebewegungen . . . . .	59
3.2	Zeichen aus der Tauchersprache . . . . .	60
3.3	Klassifikation natürlicher Gebärden nach Wundt . . . . .	62
3.4	Klassifikation redegleitender Gesten nach Ekman/Friesen . . . . .	63
3.5	Klassifikation redegleitender Gesten nach McNeill . . . . .	63

3.6	Typischer Gestenraum eines erwachsenen Sprechers . . . . .	68
3.7	Nutzung des Gestenraumes für deiktische Gesten . . . . .	68
3.8	Tätigkeitsabhängige Arbeitshöhe . . . . .	70
3.9	Drei Phasen sportlicher Bewegungsakte und ihre Relationen . .	71
3.10	Phasenmodell nach [McNeill, 1992] . . . . .	72
3.11	Modell der Gestenkommunikation . . . . .	74
3.12	Einflussfaktoren zur Entwicklung intuitiver Freiraumgesten . . .	75
4.1	Phasen einer Geste, erweitertes Modell . . . . .	80
4.2	Statische Definition der Zeigegeste . . . . .	80
4.3	Anmelderpose (pre-stroke hold) . . . . .	84
4.4	Charakteristische Winkeländerung, Dynamische Beschreibung mittels linearer Regression . . . . .	84
4.5	Handgelenkpunkt (Hg), Dynamische Beschreibung mittels line- arer Regression, Erwartung . . . . .	85
4.6	Handgelenkpunkt (Hg), Dynamische Beschreibung mittels line- arer Regression, Realer Verlauf . . . . .	85
4.7	Zeigen ohne den Arm zu beugen . . . . .	86
4.8	Zeigen mit angewinkeltem Ellenbogen . . . . .	87
4.9	Winkelverlauf einer typischen Zeigegeste, incl. Ruhephase, (Pha- sendefinition mittels zweier Winkel) . . . . .	89
4.10	Körperhaltung in den verschiedenen Gestenphasen, Phasende- finition mittels zweier Winkel . . . . .	89
4.11	Ablaufdiagramm der Phasendefinition mittels zweier Winkel . .	92
4.12	Phasen der Gestenerkennung, Phasendefinition mittels eines Winkels . . . . .	96
4.13	Ablaufdiagramm der Phasendefinition mittels eines Winkels . .	97
5.1	Aufzeichnungssoftware für die Probandenversuche . . . . .	100
5.2	Erkennungsprobleme in Abhängigkeit von der Zeigegeste . . .	102
5.3	Analyse des Schulterwinkels für die Ruhephase, Phasendefiniti- on mittels eines Winkels . . . . .	104
5.4	Bestimmung der Gesamterkennungsrate in Abhängigkeit des Äquivalent der notwendigen Haltezeit ( $t_H$ ), Phasendefinition mit- tels eines Winkels . . . . .	106

5.5	Bestimmung der Erkennungsrate (Zeigen, nicht Zeigen) in Abhängigkeit des Äquivalent der notwendigen Haltezeit ( $t_H$ ), Phasendefinition mittels eines Winkels . . . . .	106
5.6	Ausreißer ohne Glättungsfilter . . . . .	107
5.7	Ausreißer geglättet mit dem Double Moving Average Filter . . . . .	107
5.8	Untersuchung des Winkleinflusses zwischen der optischen Achse der Kinect® 2 und der Frontalebene des Probanden auf die Erkennungsrate . . . . .	112
5.9	Erkannte Sonderfälle der Winkeluntersuchung, Darstellung der gestörten Bildpunkte . . . . .	113
5.10	Winkleinfluss Frontalebene Proband - Optische Achse Kinect® 2 auf die Interferedrate . . . . .	113
6.1	Ungenauere Einweisung des Probanden in die Versuchsaufgabe . . . . .	120
6.2	Sonderfall der Kategorie 1: Es wurde mit rechts gezeit und der linke Arm erkannt . . . . .	122
6.3	Sonderfall der Kategorie 2, beide Arme benutzt und richtig erkannt	122
6.4	Sonderfall Kategorie 2, Winkelverlauf für beide Arme . . . . .	123
6.5	Kontrollgeste mit zwei Händen . . . . .	125
6.6	Sonderfall - „beide Arme“ erkannt - bei 30° Drehung um die Körperlängsachse . . . . .	127
6.7	Erkennungssystem in Bereitschaft . . . . .	134
6.8	Vergleich Zielerkennung, Zeigevektor Schulter-Hand / Kopf-Hand	139
6.9	Zielerkennung bei ortsfester Installation . . . . .	139
6.10	Zielerkennung bei mobiler Installation . . . . .	140
8.1	Modell einer vollständigen Phasendefinition . . . . .	148
A.1	Vergleich der Fensterbreite von Mittelwert- und Medianfilter bei Ausführung der Zeigegeste . . . . .	154
A.2	Vergleich der Fensterbreite von Mittelwert- und Medianfilter bei Ausführung der Kontrollgeste . . . . .	154
A.3	Überprüfen der Winkelgeschwindigkeit zwischen den Vektoren $\overline{HS}$ und $\overline{T}$ ( $\dot{\phi}_{(\overline{HS}, \overline{T})}$ ) zum Verlassen der Ruhephase . . . . .	156
A.4	Überprüfen der Winkelgeschwindigkeit zwischen den Vektoren $\overline{HS}$ und $\overline{T}$ ( $\dot{\phi}_{(\overline{HS}, \overline{T})}$ ) während der Haltephase . . . . .	156

- A.5 Überprüfen der Winkelgeschwindigkeit zwischen den Vektoren  $\overline{HS}$  und  $\overline{T}$  ( $\dot{\phi}_{(\overline{HS}, \overline{T})}$ ) während die Releasephase . . . . . 158

# Tabellenverzeichnis

3.1	Auswertung der Untersuchung zum Intuitiven Gestenverständnis	76
5.1	Erkennungsrate bei Anwendung verschiedener Filter . . . . .	108
5.2	Auswertung der Probandenversuche . . . . .	110
5.3	Verteilung der Ausgeführten Gesten im Winkelversuch . . . . .	111
5.4	Ergebnis Winkelversuch . . . . .	115
6.1	Vergleich „markerbasiertes“ (MBS) und „markerloses“ (MLS) Bewegungserkennungssystem . . . . .	131
A.1	Parameter der Gestenerkennung nach der Optimierung . . . . .	157



# Formelverzeichnis

3.1	minimaler Schulterwinkel des Gestenraumes (männlich) . . . . .	69
3.2	minimaler Schulterwinkel des Gestenraumes (weiblich) . . . . .	69
4.1	Eingeschlossener Winkel . . . . .	78
4.2	Zwangsbedingung für den Ellbogenwinkel (Zeigen statisch) . . .	81
4.3	Zwangsbedingung für den Schulterwinkel (Zeigen statisch) . . .	81
4.4	Zwangsbedingung zur Definition des statischen Haltens . . . . .	82
4.5	Zwangsbedingung der Definition Geste (dynamisch, lineare Regression) . . . . .	83
4.6	Zwangsbedingungen zur Definition der Ruhephase (Parallelitätsbedingung), (Phasendefinition mittels zweier Winkel) . . . . .	88
4.7	Zwangsbedingung zur Definition Beginn Gestenausführung, (Phasendefinition mittels zweier Winkel) . . . . .	90
4.8	Zwangsbedingung zur Definition Zeigen in der Haltephase, (Phasendefinition mittels zweier Winkel) . . . . .	90
4.9	Zwangsbedingung der Definition Ausleitung, (Phasendefinition mittels zweier Winkel) . . . . .	91
4.10	Zwangsbedingung der Definition Ruhephase, (Phasendefinition, ein Winkel) . . . . .	93
4.11	Zwangsbedingungen der Definition Beginn Gestenphase (Phasendefinition, ein Winkel) . . . . .	94
4.12	Zwangsbedingung der Definition Beginn Haltephase (Phasendefinition, ein Winkel) . . . . .	94
4.13	Zwangsbedingung der Definition Ende Haltephase (Phasendefinition, ein Winkel) . . . . .	94
4.14	Zwangsbedingung zur Definition Ausleitung (Phasendefinition, ein Winkel) . . . . .	95

5.1	Abhängigkeit Erkennungsrate Zeigen / Haltezeit . . . . .	105
6.1	Zwangsbedingung der Definition der gemeinsame Einleitung und Hauptphase . . . . .	118
6.2	Zwangsbedingung der Definition des Phasenübergang Haupt- Haltephase . . . . .	118
6.3	Zwangsbedingung der Definition der Haltephase . . . . .	118
6.4	Eingeschlossener Winkel . . . . .	128
6.5	Definition Anmeldepose (statisch) . . . . .	133
A.1	Abhängigkeit Erkennungsrate Kontrollgeste / Fensterbreit (glei- tender Mittelwert) . . . . .	155
A.2	Abhängigkeit Erkennungsrate Kontrollgeste / Fensterbreit (Rang- ordnungsfiler) . . . . .	155
A.3	Linearer Zusammenhang zwischen Winkelgeschwindigkeit zur Detektion des Beginn der Bewegung und der Erkennungsrate .	157

# Literaturverzeichnis

- Argyle, M. und Trower, P. (1981). Signale von Mensch zu Mensch: Die Wege der Verständigung. Beltz Verlag, Weinheim.
- Azimi, M. (2016). Skeletal Joint Smoothing White Paper: Kinect for Windows 1.5, 1.6, 1.7, 1.8. Online verfügbar unter: <https://msdn.microsoft.com/en-us/library/jj131429.aspx#ID4ELNAC>; Zuletzt geprüft: 16.09.2016.
- Baudel, T. und Beaudouin-Lafon, M. (1993). Charade: Remote Control of Objects Using Free-hand Gestures. *Commun. ACM*, 36(7): 28–35.
- Beecken, A., Keller, J., Prillwitz, S., und Zienert, H. (1999). Grundkurs Deutsche Gebärdensprache, Band 3 der Gebärdensprachlehre. Signum-Verlag, Hamburg.
- Beelich, K. H. und Schwede, H.-H. (1991). Denken, Planen, Handeln: Grundtechniken für zweckmäßiges Lernen und Arbeiten mit vielen Erläuterungen und Anwendungsbeispielen. Vogel-Fachbuch. Vogel, Würzburg, 3., überarbeitete Auflage.
- Bernsau, K. M. (2003). Das Mißverständnis der Informationsgesellschaft. Semiotische Kritik eines leichtfertigen Begriffs.
- Birbaumer, N. (1993). Neuro- und Sinnesphysiologie: Mit 11 Tabellen. Springer-Lehrbuch. Springer, Berlin [u.a.].
- Bonnechere, B., Jansen, B., Salvia, P., Bouzahouene, H., Omelina, L., Moiseev, F., Sholukha, V., Cornelis, J., Rooze, M., und van Sint Jan, S. (2014). Validity and reliability of the Kinect within functional assessment activities:

- comparison with standard stereophotogrammetry. *Gait & Posture*, 39(1): 593–598.
- Brockhaus (1988). *Brockhaus Enzyklopädie*: In vierundzwanzig Bänden. F.A. Brockhaus, Mannheim, 19. Auflage.
- Bundesregierung und dem Bundesministerium für Wirtschaft und Arbeit mit Zustimmung des Bundesrates (12.08.2004). *Verordnung über Arbeitsstätten (Arbeitsstättenverordnung)*: ArbStättV.
- Cappozzo, A., Della Croce, U., Leardini, A., und Chiari, L. (2005). Human movement analysis using stereophotogrammetry Part 1: Theoretical background. *Gait & Posture*, 21(2): 186–196.
- Dahm, M. (2006). *Grundlagen der Mensch-Computer-Interaktion. Informatik : Software-Ergonomie*. Pearson Studium, München [u.a.].
- Deuschl, G., Bain, P., und Brin, M. (1998). Consensus Statement of the Movement Disorder Society on Tremor. *Movement Disorders*, 13(S3): 2–23.
- Dey, A. K. (2001). Understanding and Using Context. *Personal Ubiquitous Comput*, 5(1): 4–7.
- Ekman, P. und Friesen, W. (1969). The repertoire of nonverbal Behavior: Categories, origins, usage and coding. *Semiotica*, 1(1): 49–98.
- Freeman, W. T. und Roth, M. (1994). Orientation Histograms for Hand Gesture Recognition. In *International Workshop on Automatic Face and Gesture Recognition*, 296–301.
- Fricke, E. (2007). *Origo, Geste und Raum: Lokaldeixis im Deutschen*, Band 24 der *Linguistik, Impulse & Tendenzen*. De Gruyter, Berlin and New York.
- Fürntratt, H. und Neuschmied, H. (2014). Evaluating Pointing Accuracy on Kinect V2 Sensor. In Arya, A. und Brom, C., Hrsg., *Proceedings of the 2nd International Conference on Human-Computer Interaction*, Paper No. 124.
- Hassenzahl, M., Burmester, M., und Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In Szwillus, G. und Ziegler, J., Hrsg., *Mensch & Computer 2003: Interaktion in Bewegung*, S. 187–196. Vieweg+Teubner Verlag, Wiesbaden.

- Herczeg, M. (2009). Software-Ergonomie: Theorien, Modelle und Kriterien für gebrauchstaugliche interaktive Computersysteme. Lehrbuchreihe interaktive Medien. Oldenbourg, München, 3., vollständig überarbeitete und erweiterte Auflage.
- Hochschild, J. (2005). Grundlagen zur Wirbelsäule, HWS und Schädel, BWS und Brustkorb, obere Extremität, volume 1 of Strukturen und Funktionen begreifen: funktionelle Anatomie - therapierelevante Details / Jutta Hochschild. Thieme, Stuttgart [u.a.], 3., unveränderte Auflage.
- Hurtienne, J., Mohs, C., Meyer, H. A., Kindsmüller, M. C., und Israel, J. H. (2006). Intuitive Use of User Interfaces - Definition und Herausforderungen. In : i-com : Zeitschrift für interaktive und kooperative Medien, 5(3): 38–41.
- Johannsen, G. (1993). Mensch-Maschine-Systeme. Springer-Lehrbuch. Springer, Berlin Heidelberg, 1 Auflage.
- Karduna, A. R., McClure, P. W., Michener, L. A., und Sennett, B. (2001). Dynamic Measurements of Three-Dimensional Scapular Kinematics: A Validation Study. Journal of Biomechanical Engineering, 123(2):184.
- Kendon, A. (2004). Gesture: Visible action as utterance. Cambridge University Press, Cambridge, New York.
- Kotler, P. (2003). Grundlagen des Marketing. Wirtschaft : Marketing. Pearson Studium, München and Boston [u.a.], 3., überarbeitete Auflage.
- Le Hong, S. und Biesterfeldt, J. (2010). Weltweit berührt: Studie zur Untersuchung kultureller Unterschiede und Gemeinsamkeiten bei der gestenbasierten Bedienung von Multitouch-Oberflächen. User Interface Design GmbH.
- Livingston, M. A., Sebastian, J., Ai, Z., und Decker, J. W. (2012). Performance measurements for the Microsoft Kinect skeleton. In 2012 IEEE Virtual Reality (VR), S. 119–120.
- McNeill, D. (1995, ©1992). Hand and mind: What gestures reveal about thought. University of Chicago Press, Chicago.
- Meinel, K. und Schnabel, G. (2015). Bewegungslehre - Sportmotorik: Abriss einer Theorie der sportlichen Motorik unter pädagogischem Aspekt. Meyer und Meyer, Aachen, 12.

- Mohs, C., Hurtienne, J., Kindsmüller, M. C., Israel, J. H., Meyer, H. A., und Group, I. R. (2006). IUUI - Intuitive Use of User Interfaces: Auf dem Weg zu einer wissenschaftlichen Basis für das Schlagwort "Intuitivität". MMI-Interaktiv, S. 75–84.
- Norman, D. A. (1983). Some Observations on Mental Models. In Gentner, D. und Stevens, A. L., Hrsg., *Mental models, Cognitive science*, S. 7–14. L. Erlbaum Associates, Hillsdale, N.J.
- Norman, D. A. (1988). *The design of everyday things*. Basic Books, New York.
- Nöth, W. (2000). *Handbuch der Semiotik*. Metzler, Stuttgart, 2., vollständig neu bearbeitete und erweiterte Auflage.
- Nowack, T., Jehring, S., und Kurtz, P. (2015a). Zeigen – Objektauswahl für ein industrielles Assistenzsystem zum Heben und Tragen von Kleinladungsträgern. In Wienrich, C., Zander, T. O., und Gramann, K., Hrsg., *Trends in Neuroergonomics*, S. 152–158, Berlin. Universitätsverlag der TU Berlin.
- Nowack, T., Suzaly, N., Lutherdt, S., Schürger, K., Jehring, S., Witte, H., und Kurtz, P. (2015b). Phases of technical gesture recognition. In Kurosu, M., Hrsg., *Human-Computer Interaction*, volume 9169 of LNCS sublibrary. SL 3, Information systems and applications, incl. Internet/Web, and HCI, S. 130–139. Springer, Cham.
- Nowack, T., Wenzel, S., und Kurtz, P. (2016). Technische Beschreibung einer dynamischen und einer statischen Variante der Zeigegeste. In *Arbeit in komplexen Systemen - digital, vernetzt, human?!*, A.4.12. GfA-Press, Dortmund.
- Papazov, C., Haddadin, S., Parusel, S., Krieger, K., und Burschka, D. (2012). Rigid 3D geometry matching for grasping of known objects in cluttered scenes. *The International Journal of Robotics Research*, 31(4): 538–553.
- Pavlovic, V. I., Sharma, R., und Huang, T. S. (1997). Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7): 677–695.
- Preim, B. (1999). *Entwicklung interaktiver Systeme: Grundlagen, Fallbeispiele und innovative Anwendungsfelder ; mit 20 Tabellen*. Springer-Lehrbuch. Springer, Berlin [u.a.].

- Sati, M., de Guise, J. A., Larouche, S., und Drouin, G. (1996). Quantitative assessment of skin-bone movement at the knee. *The Knee*, 3(3): 121–138.
- Schick, A. und Sauer, O. (2013). Gestenbasierte Qualitätskontrolle: Intuitive Mensch-Maschine-Interaktion in der Industrie. Online verfügbar unter <http://www.iosb.fraunhofer.de/servlet/is/44590/Artikel%20Gesteninteraktion.pdf?command=downloadContent&filename=Artikel%20Gesteninteraktion.pdf>
- Schmauder, M. und Spanner-Ulmer, B. (2014). *Ergonomie - Grundlagen zur Interaktion von Mensch, Technik und Organisation*. REFA-Fachbuchreihe Arbeitsgestaltung. Carl Hanser, München.
- Schmidtke, H. und Bernotat, R. (1993). *Ergonomie*. Hanser, München, 3., neubearbeitete und erweiterte Auflage.
- Shannon, C. E. und Weaver, W. (1964 [©1949]). *The mathematical theory of communication*. University of Illinois Press, Urbana.
- Shneiderman, B. und Plaisant, C. (2010). *Designing the user interface: Strategies for effective human-computer interaction*. Addison-Wesley, Boston, 5th ed. edition.
- Skinner, A., Vice, J., und Baraniecki, L. (2015). Development of Tactile and Gestural Displays for Navigation, Communication, and Robotic Control. In Yamamoto, S., Hrsg., *Human Interface and the Management of Information. Information and Knowledge Design*, volume 9172 of *Lecture Notes in Computer Science*, S. 445–457. Springer International Publishing, Cham.
- Southwood, P. (2011). Dive hand signal. Online verfügbar unter [de.wikipedia.org](http://de.wikipedia.org).
- Steinhorst, P., Schäfer, M., Reinhold, S., Schomakers, V., und Teistler, M. (2016). Selbstbeschreibungsfähigkeit von Touch-Gesten mobiler Geräte. DOI: 10.18420/muc2016-mci-0258
- Suzaly, N., Sprenger, S., Nowack, T., Witte, H., und Kurtz, P. (2015). Beurteilung der Körperhaltung beim Heben und Tragen von Lasten mit Hilfe Kinect® Kamera. In Gesellschaft für Arbeitswissenschaft e.V., Hrsg., *VerANTWORTung für die Arbeit der Zukunft*, Beitrag B.1.23, Dortmund. GfA-Press.

- Vimieiro, C., Andrada, E., Lipfert, S., Witte, H., und Pinotti, M. (2007). Calculation of the cardan angles at the hip joint using the ISB-recommendation to define the joint coordinate system. In COBEM 2007.
- Walter, R., Bailly, G., und Müller, J. (2013). StrikeAPose: Revealing Mid-air Gestures on Public Displays. In Mackay, W. E., Hrsg., CHI 2013, S. 841–850, New York, NY. ACM.
- Webster, D. und Celik, O. (2014). Experimental evaluation of Microsoft Kinect's accuracy and capture rate for stroke rehabilitation applications. In 2014 IEEE Haptics Symposium (HAPTICS), S. 455–460.
- Wenzel, S., Nowack, T., und Kurtz, P. (2016). Unterstützung der Körperhaltungsbewertung laut Leitmerkmalmethode „Ziehen und Schieben“ mit Hilfe einer Tiefenkamera. In Arbeit in komplexen Systemen - digital, vernetzt, human?!, S. B.4.9. GfA-Press, Dortmund.
- Weyerstall, F. und Saul, M. (2013). Vergleich der Multitouch-Steuerung mit berührungslosen Gesten. FHWS SCIENCE JOURNAL, 1(2). Online verfügbar unter [https://opus4.kobv.de/opus4-fhws/frontdoor/deliver/index/docId/49/file/Vergleich\\_der\\_Multitouch-Steuerung.mit\\_beruehrungslosen.Gesten.pdf](https://opus4.kobv.de/opus4-fhws/frontdoor/deliver/index/docId/49/file/Vergleich_der_Multitouch-Steuerung.mit_beruehrungslosen.Gesten.pdf)
- Wilson, A. D., Bobick, A. F., und Cassell, J. (1996). Recovering the temporal structure of natural gesture. In Second International Conference on Automatic Face and Gesture Recognition, S. 66–71.
- Wundt, W. (1900). Völkerpsychologie: eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos, und Sitte. W. Engelmann, Leipzig.
- Zeidler, E., Bronštejn, I. N., und Grosche, G. (1996). Teubner-Taschenbuch der Mathematik. Teubner, Stuttgart [u.a.], vollständige Neufassung.
- Zimbardo, P. G., Gerrig, R. J., und Hoppe-Graff, S. (1999). Psychologie: Mit 70 Tabellen. Springer-Lehrbuch. Springer, Berlin [u.a.], 7. Auflage.
- Zühlke, D. (2004). Useware-Engineering für technische Systeme: Mit 39 Tabellen. VDI. Springer, Berlin [u.a.].

# Verzeichnis der verwendeten Normen

[DIN 33402-2 2005] Ergonomie - Körpermaße des Menschen - Teil 2: Werte.  
ICS 13.180. Berlin, 2005-12

[DIN EN ISO 9241-11 1999] Norm DIN EN ISO 9241-11 Januar 1999. Ergo-  
nomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten Teil 11:  
Anforderungen an die Gebrauchstauglichkeit – Leitsätze

[VDI/VDE 3850-1 2000] Norm VDI/VDE 3850-1 Mai 2000. Nutzergerechte  
Gestaltung von Bediensystemen für Maschinen

