Universität der Bundeswehr München
Institut für
Flugsysteme

Visuelles Kommunikationssystem zur nonverbalen Mensch-UAV-Interaktion

Alexander Schelle

Vollständiger Abdruck der von der Fakultät für Luft- und Raumfahrttechnik der Universität der Bundeswehr München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Gutachter: Univ.-Prof. Dr.-Ing. Peter Stütz
Univ.-Prof. Dr. rer. nat. Andreas Wendemuth

Die Dissertation wurde am 21.06.2022 bei der Universität der Bundeswehr München eingereicht und durch die Fakultät für Luft- und Raumfahrttechnik am 23.11.2022 angenommen. Die mündliche Prüfung fand am 20.12.2022 statt.

Kurzfassung

Die nonverbale Verständigung per Handzeichen stellt in der bemannten Luftfahrt ein bewährtes Alternativverfahren bei technisch- oder umgebungsbedingten Kommunikationsproblemen dar. Trotz modernster abbildender Sensorik und zunehmender Automatisierung fehlt bei unbemannten Fluggeräten (UAVs) bislang eine derartige visuelle Schnittstelle, sodass der Informationsaustausch auch innerhalb der Sichtweite primär über eine Funkverbindung stattfindet, welche nur über ein kompatibles, technisches Kommunikationsgerät hergestellt werden kann. Mit dem Verlust dieser Verbindung geht im Allgemeinen auch ein Verlust der Kontrolle über das unbemannte System einher. Zudem kann die aktive Ausstrahlung einer elektromagnetischen Signatur in bestimmten Anwendungsfällen unerwünscht oder sogar gefährlich für den Systemnutzer sein.

Im Bereich der *collocated Human-Drone Interaction* (cHDI) wurden daher verschiedene Anstrengungen unternommen, um eine natürliche und direkte Mensch-UAV-Interaktion ohne den Einsatz zusätzlicher Kommunikationsgeräte zu ermöglichen. Die nonverbale Interaktion über eine visuelle Schnittstelle bietet dabei im Vergleich zu anderen Modalitäten die meisten Vorteile, wobei das Potential dieser Schnittstelle gegenwärtig nicht vollständig ausgeschöpft wird, da primär der *Interaktionsaspekt* und nicht der *Kommunikationsaspekt* behandelt wird. Ein bidirektionaler und zielgerichteter Informationsaustausch stellt jedoch eine Grundvoraussetzung für eine natürliche Mensch-UAV-Schnittstelle dar, wofür auch eine *Intentionserkennung* benötigt wird. Das Ziel dieser Arbeit ist es somit, einen Beitrag zum Ausbau der kommunikativen Fähigkeiten von UAVs zu leisten und einen alternativen Systemzugriff zu ermöglichen.

Hierfür wurde zunächst ein interdisziplinäres Konzept für ein visuelles Mensch-UAV-Kommunikationssystem entwickelt, welches relevante nonverbale Körpersignale optisch erfassen und mittels eines aufgabenbasierten Dialogsystems kontextsensitiv in Nutzerintentionen dekodieren kann. Auf dieser methodischen Grundlage wurde ein echtzeitfähiges Experimentalsystem in einen Flugversuchsträger integriert und dessen Klassifikationsleistung bewertet. In anschließenden Mensch-Maschine-Experimenten wurde das Systemkonzept evaluiert und zusätzlich untersucht, welche nonverbalen Subsignale, insbesondere von ungeschulten Nutzern, zur Kommunikation eingesetzt werden, welche Führungsverfahren sich für eine kollokierte UAV-Führung über eine visuelle Schnittstelle eignen und welche Faktoren Verständigungsprobleme begünstigen.

Die Eignung des Systemkonzepts konnte in unterschiedlichen Nutzungskontexten nachgewiesen und eine signifikante Effektivitätserhöhung durch eine kontextsensitive Signaldekodierung bestätigt werden. Dialogunterstützungsfunktionen können zudem dabei helfen, nutzerinduzierte (z.B. Darstellungsfehler) als auch systeminduzierte Verständigungsprobleme (z.B. Klassifizierungsfehler) während eines nonverbalen Mensch-UAV-Dialogs zu reduzieren und das Kommunikationsziel effizienter zu erreichen. Für die Übermittlung höherwertiger Kommunikationsinhalte (z.B. Missionsdaten) werden jedoch höhere Anforderungen an die nonverbalen Fähigkeiten des Nutzers gestellt, da ein größerer Satz an Signalen dargestellt werden muss. Ein mehrdeutiger Wortschatz mit gestischen Metaphern und Synonymen kann die Intuitivität steigern und dem Nutzer Handlungsoptionen bei nutzerspezifischen Darstellungsdefiziten oder systembedingten Erkennungsproblemen bieten. Weiteres Potential liegt in der dreidimensionalen Repräsentation des Nutzers, der Einbeziehung zusätzlicher visuell wahrnehmbarer Körpersignale, wie der Mimik, als auch in einem insgesamt intelligenteren UAV-Verhalten auf dem Weg zu einer natürlichen Mensch-UAV-Kommunikation.

Abstract

In manned aviation, non-verbal communication via hand signals is a well-established alternative method in the event of technical or environmental communication problems. Despite state-of-the-art imaging sensor technology and increasing automation, unmanned aerial vehicles (UAVs) have so far lacked such a visual interface, so that the exchange of information even within visual range primarily takes place via a radio link, which can only be established via a compatible, technical communication device. Loss of this link is generally accompanied by loss of control of the unmanned system. In addition, the active emission of an electromagnetic signature can be undesirable or even dangerous for the system user in certain applications.

In the field of *collocated Human-Drone Interaction* (cHDI), various efforts have therefore been made to enable natural and direct human-UAV interaction without the use of additional communication devices. In this context, nonverbal interaction via a visual interface offers the most advantages compared to other modalities, although the potential of this interface is currently not fully exploited because primarily the *interaction aspect* is addressed rather than the *communication aspect*. However, bidirectional and goal-directed information exchange is a basic requirement for a natural human-UAV interface, for which *intention recognition* is also needed. Thus, the goal of this work is to contribute to the expansion of the communicative capabilities of UAVs and to enable an alternative system access.

For this purpose, an interdisciplinary concept for a visual human-UAV communication system was first developed, which can optically detect relevant nonverbal body signals and decode them into user intentions in a context-sensitive manner by means of a task-based dialog system. Based on this methodological foundation, a real-time capable experimental system was integrated into an UAV and its classification performance was evaluated. In subsequent human-machine experiments, the system concept was evaluated and additionally investigated which nonverbal subsignals are used for communication, especially by untrained users, which guidance procedures are suitable for collocated UAV guidance via a visual interface, and which factors promote comprehension problems.

The suitability of the system concept could be proven in different usage contexts and a significant increase in effectiveness could be confirmed by context-sensitive signal decoding. Dialog support functions can also help to reduce user-induced (e.g., representation errors) as well as system-induced comprehension problems (e.g., classification errors) during a nonverbal human-UAV dialog and to achieve the communication goal more efficiently. However, for the transmission of more sophisticated communication content (e.g., mission data), higher demands are placed on the user's nonverbal capabilities, as a larger set of signals must be reproduced. An ambiguous vocabulary with gestural metaphors and synonyms can increase intuitiveness and provide the user with options for action in case of user-specific representation deficits or system-related recognition problems. Further potential lies in the three-dimensional representation of the user, the inclusion of additional visually perceptible body signals such as facial expressions, as well as in an overall more intelligent UAV behavior on the way to a natural human-UAV communication.

„ (…) only humans can lose one modality (e.g., hearing) and make up for this deficit by communicating with complete competence in a different modality (e.g., signing). “

- Hauser, Chomsky und Fitch, 2002 [1, S. 1575]

Inhaltsverzeichnis

Kurzfassung i
Abstract iii
Abkürzungsverzeichnis xi
Abbildungsverzeichnis xiii
Tabellenverzeichnis xvii
1 Einführung 1
1.1 Motivation 2
1.2 Aufbau der Arbeit 3
2 Grundlagen 5
2.1 Unbemannte Flugsysteme 5
2.1.1 Einordnung 5
2.1.2 Militärische Nutzung von UAS 8
2.1.3 Kommerzielle Nutzung von UAS 10
2.1.4 Remote Access 12
2.1.5 Grenzen funkbasierter Kommunikation 12
2.2 Nonverbale interpersonelle Kommunikation 14
2.2.1 Klassische Ansätze zur Klassifizierung gestischer Äußerungen 16
2.2.2 Technische Handzeichensysteme in der Luft- und Raumfahrt 18
2.2.3 Visuelle Boden-Luft Kommunikation 21
2.2.4 Visuelle Luft-Luft und Luft-Boden Kommunikation 24
2.2.5 Gebärdensprachen 24
2.2.6 Phonologie visueller Handzeichensysteme am Beispiel der Deutschen Gebärdensprache 25
2.3 Zusammenfassung 29
3 Stand der Forschung 31
3.1 Collocated Human-Drone Interaction 32
3.1.1 Neuartige Anwendungen 33
3.1.2 Interaktionsmodalitäten 34
3.1.3 Kommunikationsinhalte 36
3.1.4 Proxemik 37
3.1.5 Evaluierungsmethoden 38
3.1.6 Zusammenfassung und Literaturübersicht 39
3.2 Bildbasierte Gestenerkennungssysteme in der cHDI 41
3.2.1 Sensortechnologien zur Erfassung von Bewegungen 42
3.2.2 Körpermodelle 44
3.2.3 Methoden zur luftgestützten Erfassung von Körpermerkmalen 45
3.2.4 Methoden zur Klassifizierung von Körpersignalen 50
3.2.5 Zusammenfassung und Literaturübersicht 54
3.3 Design und Verfügbarkeit von gestischen Vokabularen in der cHDI 55
3.3.1 Motive 55
3.3.2 Gestendatensätze 57
3.4 Übermittlung von Systemantworten über einen visuellen Rückkanal 58
3.4.1 Flugdynamische Kommunikationsverfahren 58
3.4.2 Lichtbasierte Kommunikationsverfahren 59
3.5 Research Gaps und Forschungsfragen 62
3.5.1 Fähigkeitslücke 1: Visuelles Kommunikationssystem zur Intentionserkennung 63
3.5.2 Fähigkeitslücke 2: Erkennung generischer Nutzereingaben 63
3.5.3 Wissenslücke 1: Nonverbale Kommunikationssignale von Nutzern ohne Systemerfahrung im Kontext einer simulierten Notsituation 64

3.5.4 Wissenslücke 2: Kollokierte UAV-Führung über eine visuelle Schnittstelle ... 64
3.5.5 Wissenslücke 3: Verständigungsprobleme bei nonverbalen Dialogen mit technischen Systemen ... 64
3.5.6 Fähigkeitslücke 3: Dialogunterstützungsfunktionen zur Verbesserung der nonverbalen Mensch-UAV-Kommunikation ... 65

4 Konzept für ein visuelles Mensch-UAV-Kommunikationssystem ... 67
4.1 Annahmen und Umgebungsbedingungen ... 67
4.2 Definitionen ... 68
4.3 Anforderungen an ein visuelles Mensch-UAV-Kommunikationssystem ... 70
4.3.1 Variabler Interaktionsbereich ... 70
4.3.2 Anpassungsfähigkeit an unterschiedliche Nutzergruppen ... 70
4.3.3 Inhärente Erkennung intentionaler Körpersignale ... 72
4.3.4 Zugangsregelung ... 72
4.3.5 Echtzeitfähigkeit des Systems ... 74
4.3.6 Visuelle Übermittlung der Systemantworten ... 75
4.3.7 Ergonomisches Vokabular ... 75
4.3.8 Sicherheitsrelevante Aspekte ... 77
4.3.9 Zusätzliche Anforderungen an ein Experimentalsystem ... 78

4.4 Das UAV als „Dienstleister“ ... 79
4.4.1 Kommunikationsmodell ... 79
4.4.2 VisCom-Schichtenmodell ... 80
4.4.3 Kommunikationsprotokoll ... 82
4.4.4 Interaktionsmodell des Nutzers ... 84
4.4.5 Modell zur Beschreibung des Systemverhaltens ... 85
4.4.6 Dialog- und UAV-bezogene Dienste ... 85

4.5 Visuelle Nutzer- und Subsignalmodellierung ... 86
4.5.1 Anatomisches Modell ... 86
4.5.2 Nonverbale Kommunikationsmittel ... 87

4.6 Kontextsensitive Signaldekodierung mittels Motion-to-Symbol-Ansatz ... 88
4.7 NLU-gestützte Diensterkennung ... 90
4.8 Modellierung einer symbolbasierten Mensch-UAV-Verkehrssprache ... 91

5 Umsetzung des Konzepts in ein Experimentalsystem ... 95
5.1 Nutzererfassungssystem ... 96
5.1.1 Sensorsteuerung ... 96
5.1.2 Posenbestimmung ... 96
5.1.3 Nutzerauthentifizierung ... 98
5.1.4 Gimbalsteuerung ... 100

5.2 Klassifizierungssystem ... 101
5.2.1 Visueller Nutzerzustand ... 102
5.2.2 Orientierung des Kopfes und Torsos ... 103
5.2.3 Stellung der Arme ... 104
5.2.4 Bewegungsrichtung und Oszillation des Kopfes und der Hände ... 106
5.2.5 Gestikulationsstärke ... 108
5.2.6 Mundbild ... 108
5.2.7 Position der Hände ... 110
5.2.8 Bewegungspfad der Hände ... 111
5.2.9 Berührung ... 112
5.2.10 Handform ... 112
5.2.11 Signaldekodierung ... 115

5.3 Dialogsystem ... 116
5.3.1 Dialogmanager ... 117
5.3.2 Natural Language Processing ... 117
5.3.3 Displaysteuerung für den visuellen Rückkanal ... 118

5.4 Flugsteuerungssystem ... 119
5.4.1 Flugmanager ... 119
5.4.2 UAV-Schnittstelle ... 119

5.5 Dienste und Systemkonfigurationen ... 120
5.6 Realisierung und Integration des Experimentalsystems in eine fliegende Plattform ... 122

6 Experimentelle Evaluierung ... 125
6.1 Bewertung der Klassifikationsleistung des Experimentalsystems ... 125
6.2 Experiment A: Nonverbale Körpersignale von Nutzern ohne Systemerfahrung ... 128
6.2.1 Untersuchungsinhalte ... 128
6.2.2 Untersuchungsdesign ... 129
6.2.3 Stichprobenbeschreibung ... 130
6.2.4 Datenerhebung ... 131
6.2.5 Versuchsdurchführung ... 131
6.2.6 Operationalisierung und Messmethoden ... 131
6.2.7 Ergebnisse ... 132
6.2.8 Diskussion ... 136

6.3 Experiment B: kollokierte UAV-Führung im Realflug ... 138
6.3.1 Untersuchungsinhalte ... 138
6.3.2 Untersuchungsdesign ... 141
6.3.3 Stichprobenbeschreibung ... 142
6.3.4 Datenerhebung ... 142
6.3.5 Versuchsdurchführung ... 143
6.3.6 Operationalisierung und Messmethoden ... 144
6.3.7 Ergebnisse ... 145
6.3.8 Diskussion ... 150

6.4 Experiment C: nonverbaler Dialog ... 152
6.4.1 Untersuchungsinhalte ... 152
6.4.2 Untersuchungsdesign ... 154
6.4.3 Stichprobenbeschreibung ... 156
6.4.4 Datenerhebung ... 156
6.4.5 Versuchsdurchführung ... 156
6.4.6 Operationalisierung und Messmethoden ... 158
6.4.7 Ergebnisse ... 159
6.4.8 Diskussion ... 164

7 Zusammenfassung und Ausblick ... 171
7.1 Beiträge dieser Arbeit zum Stand der Forschung ... 171
7.2 Bewertung der entwickelten Verfahren zur Adressierung der Fähigkeitslücken und Überprüfung der aufgestellten Hypothesen ... 172
7.2.1 Fähigkeitslücke 1: Visuelles Kommunikationssystem zur Intentionserkennung ... 172
7.2.2 Fähigkeitslücke 2: Erkennung generischer Nutzereingaben ... 173
7.2.3 Fähigkeitslücke 3: Dialogunterstützungsfunktionen zur Verbesserung der nonverbalen Mensch-UAV-Kommunikation ... 173

7.3 Beantwortung der identifizierten Wissenslücken ... 174
7.3.1 Wissenslücke 1: Nonverbale Kommunikationssignale von Nutzern ohne Systemerfahrung im Kontext einer simulierten Notsituation ... 174
7.3.2 Wissenslücke 2: Kollokierte UAV-Führung über eine visuelle Schnittstelle ... 175
7.3.3 Wissenslücke 3: Verständigungsprobleme bei nonverbalen Dialogen mit technischen Systemen ... 176

7.4 Ausblick und Anknüpfungspunkte für weitere Forschung ... 177

Danksagung ... 179
Literaturverzeichnis ... 181
Anhang ... 197

A1 Phoneme der Deutschen Gebärdensprache 197
A2 Klassifikationsleistung des Merkmalbeschreibers für statische Armstellungen 198
A3 Klassifikationsleistung des Merkmalbeschreibers für gezeichnete Symbole 199
A4 Auszug aus dem Handformendatensatz 200
A5 Vokabular und Syntax des Experimentalsystems (Interview Mode für Experiment A) 201
A6 Vokabular und Syntax des Experimentalsystems (Control Mode für Experiment B) 202
A7 Vokabular und Syntax des Experimentalsystems (Briefing Mode für Experiment C) 203
A8 Klassifikationsleistung des Experimentalsystems 204
A9 NASA Raw TLX Fragebogen 206
A10 Ermittelte Daten aus Experiment A 207
A11 Ermittelte Daten aus Experiment B 208
A12 Ermittelte Daten aus Experiment C 209
A13 Veröffentlichungen des Autors 210

Abkürzungsverzeichnis

A

API	*Application Programming Interface*
ASL	*American Sign Language*

B

BKS	*Bodenkontrollstation*
BOS	*Behörden und Organisationen mit Sicherheitsaufgaben*
BVLOS	*Beyond Visual Line of Sight*

C

cHDI	*Collocated Human-Drone-Interaction*
CNPC	*Control and Non-Payload Communication Link*
CPM	*Convolutional Pose Machine*

D

DGS	*Deutsche Gebärdensprache*

E

EAR	*Eye Aspect Ratio*
EASA	*European Union Aviation Safety Agency*
EVA	*Extra Vehicular Activity*

F

FARP	*Forward Arming and Refueling Point*
FCS	*Flight Control System*

H

HDI	*Human-Drone Interaction*
HMI	*Human-Machine Interaction*
HPE	*Human Pose Estimation*
HRI	*Human-Robot Interaction*
HWK	*Hardware-Komponente*

I

ICAO	*International Civil Aviation Organization*
IPP	*Image Processing Pipeline*
ISM	*Industrial, Scientific, Medical*
ISO	*International Organization for Standardization*

K

KS	*Kontextsensitivität*

L

LWIR	*Longwave Infrared*

M

MAR	*Mouth Aspect Ratio*

N

NLU	*Natural Language Understanding*
NUI	*Natural User Interface*

P

PAF	*Part Affinity Field*
PCL	*Payload Communication Link*
PLOC	*Prolonged Loss of Communications*

R

ROI	*Region of Interest*
ROS	*Robot Operating System*

S

SoC	*State of Charge*
STANAG	*Standardization Agreement*
SWK	*Software-Komponente*

T

TCP/IP	*Transmission Control Protocol/Internet Protocol*
ToF	*Time-of-Flight*

U

UA	*Unmanned Aircraft*
UAS	*Unmanned Aircraft System*
UAV	*Unammend Aerial Vehicle*
UCAV	*Unmanned Combat Aerial Vehicle*
UDP	*User Datagram Protocol*
UI	*User Interface*

V

V2I	*Vehicle-To-Infrastructure*
V2V	*Vehicle-To-Vehicle*
VLOS	*Visual Line of Sight*
VTOL	*Vertical Take-Off and Landing*

W

WoZ	*Wizard of Oz*

Abbildungsverzeichnis

Abb. 1 Nonverbale Landeeinweisung einer NH-90 Besatzung auf der niederländischen Fregatte Zr. Ms. Evertsen 2
Abb. 2 Inhaltlicher Aufbau der Arbeit 3
Abb. 3 Bandbreite an unbemannten Fluggeräten, bezogen auf die maximale Abflugmasse und operative Flughöhe 8
Abb. 4 Stationäre Bodenkontrollstation für ein UAS aus der HALE-Kategorie (links) im Vergleich zum portablen Steuergerät eines microUAS (rechts) 9
Abb. 5 Start eines UCAV vom Typ X-47B von einem amerikanischen Flugzeugträger (rechts). 9
Abb. 6 Funkbasierte Koordination der Rollbewegungen eines unbemannten Tankflugzeugs vom Typ MQ-25 mittels DIABLO 10
Abb. 7 Prognose der Nachfrageentwicklung auf dem Markt für privat und kommerziell genutzte UAS in Deutschland bis 2030 11
Abb. 8 Interpersonelle Kommunikationskanäle 15
Abb. 9 Typische Gestenräume eines Erwachsenen nach McNeill 16
Abb. 10 Taxonomie von Morris [114] zur Einteilung von Gesten nach ihrer kommunikativen Funktion und Grad der Formalisierung 17
Abb. 11 Auszug aus den ICAO Marshalling Signals für Rollbewegungen 18
Abb. 12 Auszug aus den ICAO Marshalling Signals für schwebefähige Luftfahrzeuge 19
Abb. 13 Handsignale für Kraftstoffmengen der U.S. Navy 19
Abb. 14 Handsignale der Apollo-Rettungstaucher (links) und Handzeichensystem von Höhlentauchern (rechts) 20
Abb. 15 Auszug aus der Studie der NASA zur extraterrestrischen Kommunikation mittels Handsignalen 21
Abb. 16 Visuelles Alternativverfahren zur Übermittlung von Flugfreigaben aus dem Kontrollturm an ein Luftfahrzeug mittels Lichtsignalen 22
Abb. 17 Symbolische Boden-Luft-Signale der U.S. Army 22
Abb. 18 Internationale Notsignale 23
Abb. 19 Visuelle Notsignale der U.S. Army 23
Abb. 20 Hauptphoneme der Deutschen Gebärdensprache 26
Abb. 21 Ausführungsstellen im Raum 27
Abb. 22 Die sechs Grundhandformen 27
Abb. 23 Gestisches Vokabular zur Übermittlung von Aufträgen an ein UAV mittels einer formalen Sprache 37
Abb. 24 Einordnung der untersuchten Gestenerkennungssysteme im Kontext der HDI 41
Abb. 25 Wesentliche Verarbeitungsschritte eines generischen Gestenerkennungssystems 42
Abb. 26 In der HPE verwendete Körpermodelle 44
Abb. 27 Lokalisierung der Fingerkuppen (gelbe Kreise) mittels Defekten in der konvexen Hüllkurve (blaue Kontur) der farbsegmentierten Hand (weiße Kontur) 46
Abb. 28 Lokalisierung der Extremitäten in stark verrauschten Tiefendaten eines stereoskopischen Sensors aus einem schwebenden UAV 46
Abb. 29 Posenbestimmung mittels Tiefendaten und geodätischen Distanzen 47
Abb. 30 DNN basierte Pipeline von OpenPose zur Posenschätzung mittels Part Confidence Maps und Part Affinity Fields... 48
Abb. 31 Inferenzzeiten verschiedener DNN-Ansätze zur Multi-Person HPE im Vergleich 49
Abb. 32 a) Gestische „home zones“ (gestrichelt) um einen Nutzer, b) das verwendete Gestenvokabular 51
Abb. 33 Gestenklassifizierung anhand von dreidimensionalen Gestenräumen 53
Abb. 34 Farbkodierte Betriebsindikatoren (rote Markierungen) eines kleinen Hexakopters (links) sowie eines UAVs der HALE-Kategorie (rechts) 60
Abb. 35 Visualisierung von vier Anzeigemetaphern zur Richtungsindikation mittels eines umlaufenden LED-Rings 60
Abb. 36 Darstellung von Text und Grafiken mittels rotierender LED-Propeller 61
Abb. 37 Auf den Boden projizierte Schaltflächen zur gestischen Interaktion mit einem schwebenden UAV 61
Abb. 38 Die identifizierten Forschungslücken im Überblick 63
Abb. 39 Die thematische Positionierung dieser Arbeit im Kontext der *collocated Human-Drone Interaction* 66
Abb. 40 Relevante Flugphasen für eine visuelle Interaktion (blaue Markierung) 68
Abb. 41 Warnschutzweste (links) und ArUco-Marker (rechts) als physikalische Authentifizierungsmittel 74
Abb. 42 Handzeichen sortiert nach Komfort 76
Abb. 43 Signal-Darstellung-Zuordnung als multikriterielles Optimierungsproblem 77
Abb. 44 Vereinfachte Darstellung der visuellen Mensch-UAV-Interaktion als Client/Server-Modell mithilfe eines zwischengeschalteten Kommunikationssystems 79
Abb. 45 Das VisCom-Kommunikationsmodell 80
Abb. 46 Unterschiedliche Verarbeitungsebenen des Kommunikationssystems 83
Abb. 47 Modellierung des Nutzerinteraktionsverhaltens als endlicher Automat 84
Abb. 48 Das Dienstmodell des VisCom-Konzepts 86
Abb. 49 Das BODY_25 Körpermodell von OpenPose inklusive der VisCom-Erweiterungen 87
Abb. 50 Phoneme des VisCom-Konzepts 88
Abb. 51 Kodierungsebenen des VisCom-Konzepts 89
Abb. 52 Kontextabhängige Dekodierung von Beobachtungen in Symbole 90

Abb. 53 NLU-gestützte Erkennung von UAV- und Dialogdiensten ... 91
Abb. 54 NLP-basiertes Verfahren zur grammatikalischen Komplexitätsreduktion von verbal formulierten UAV-Auftragsbeschreibungen ... 92
Abb. 55 Übersicht der funktionalen Hard- und Softwarekomponenten des Experimentalsystems ... 95
Abb. 56 Experimentell ermittelte Detektionsleistung von Openpose in Abhängigkeit von der Abbildungsgröße einer Person und der gewählten Auflösung des Detektornetzwerkes ... 97
Abb. 57 Ablaufdiagramm zur Nutzerauthentifizierung ... 98
Abb. 58 Visualisierung der implementierten Authentifizierungsverfahren ... 99
Abb. 59 Zielvorgaben für die Sensornachführung mit dem Nacken als Referenzpunkt für die X/Y-Positionierung und der Torsolänge als Referenzwert für die Abbildungsgröße ... 100
Abb. 60 Verlauf der Zoomregelung zur Optimierung der Abbildungsgröße und Aufrechterhaltung der Sichtverbindung ... 101
Abb. 61 Klassifikationsmodule des Klassifizierungssystems ... 102
Abb. 62 Verwendete Merkmale zur groben Vorklassifizierung des Nutzerzustandes ... 103
Abb. 63 Bestimmung der Orientierung des Kopfes mittels Konfidenzmethode ... 103
Abb. 64 Bestimmung der Orientierung des Torsos mittels Konfidenzmethode (oben) und Distanzmethode (unten) ... 104
Abb. 65 Zweistufiges, geometrisches Verfahren zur Klassifikation von Kopf- und Handbewegungen anhand von 2D-Positionsdaten ... 107
Abb. 66 Erkennung von Blinkvorgängen anhand von Längenverhältnissen am Auge ... 109
Abb. 67 Algorithmus zur Klassifikation des Mundbildes ... 109
Abb. 68 Verwendete Gesichtsmerkmale für die Klassifizierung des Mundbildes anhand des MAR ... 109
Abb. 69 Visuelle Erkennung von lautem Rufen ... 110
Abb. 70 Klassifizierung der Handpositionen anhand von zehn zweidimensionalen Gestenräumen ... 111
Abb. 71 Axiale und radiale Berührungsbereiche, grün hervorgehoben ist der von der linken Hand aktivierte Körperteil ... 112
Abb. 72 Verfahren zur automatisierten Annotation von Bilddaten mit Handformen ... 113
Abb. 73 Veranschaulichung der *Area of Overlap* und *Area of Union* zur Berechnung der *Intersection over Union* ... 114
Abb. 74 Beispielhafte Beschreibung der „Fernglas-Geste“ ... 116
Abb. 75 Im Dialogsystem implementierte Dialogstrategie ... 117
Abb. 76 Auszug aus dem Datensatz zur Beschreibung der Intention *LiftObjectDirection* mittels YAML-Notation, ... 118
Abb. 77 Auszug aus dem Datensatz zur Beschreibung der Entitäten *object* und *hdirection* mittels YAML-Notation ... 118
Abb. 78 Implementiertes Anzeigekonzept für den visuellen Rückkanal ... 119
Abb. 79 Das realisierte Experimentalsystem in der Frontal- und Seitenansicht ... 122
Abb. 80 Elektrische Verbindungsübersicht aller Hardware-Komponenten des Experimentalsystems ... 123
Abb. 81 Das vollintegrierte Experimentalsystem im Flug ... 123
Abb. 82 Konfusionsmatrix mit der Klassifikationsleistung des Experimentalsystems im semantischen Kontext *Action* ... 126
Abb. 83 Konfusionsmatrix mit der Klassifikationsleistung des Experimentalsystems im semantischen Kontext *Dialog* ... 127
Abb. 84 Laboraufbau im Experiment A ... 130
Abb. 85 Systemverhalten im Experiment A ... 130
Abb. 86 Verteilung der Darstellungsformen (links) sowie der Beteilung der Arme (rechts oben) und des Mundes (rechts unten) zur Signalisierung eines Interaktionswunsches ... 133
Abb. 87 Die verwendeten Darstellungsformen zur visuellen Rückmeldung ... 133
Abb. 88 Die Verteilung der Subsignale während der Rückmeldung ... 134
Abb. 89 Verteilung der Darstellungsformen zur Beschreibung eines körperlichen Problems (links), der Handbeteiligung (rechts oben) und der genutzten Handformen (rechts unten) ... 134
Abb. 90 Erkennungsdauer von Notsignalen mittels der Gestikulationsstärke mit Markierung bei 20 s ... 135
Abb. 91 Kontinuierliches und diskretes Führungsverfahren im Zeitverlauf ... 139
Abb. 92 Räumliche Flugführungspräferenzen unter Beibehaltung einer Sichtverbindung ... 140
Abb. 93 Versuchsaufbau im Experiment B ... 141
Abb. 94 Das für Experiment B verwendete Testgelände mit hervorgehobenen Abflug-/Landezonen A und B sowie dem fiktiven Hindernis dazwischen ... 143
Abb. 95 Das Experimentalsystem aus der Sicht eines Probanden kurz vor der Landung in Zone A ... 144
Abb. 96 Trajektorien der Flüge mit kontinuierlichem Führungsverfahren (KF-Verfahren) ... 146
Abb. 97 Trajektorien der Flüge mit diskretem Führungsverfahren (DF-Verfahren) ... 146
Abb. 98 Subjektive Beanspruchung in Abhängigkeit vom verwendeten Führungsverfahren ... 147
Abb. 99 Subjektive Beanspruchung in Abhängigkeit vom Führungsverfahren gruppiert nach Erfahrungsniveau ... 147
Abb. 100 Anzahl der empfangenen Kommandos pro Flug in Abhängigkeit vom verwendeten Führungsverfahren ... 148
Abb. 101 Landegenauigkeit in Abhängigkeit vom verwendeten Führungsverfahren ... 148
Abb. 102 Flugdauer in Abhängigkeit vom verwendeten Führungsverfahren ... 148
Abb. 103 Flugführungspräferenzen in Abhängigkeit von der Flugphase ... 149
Abb. 104 Entfernung zum UAV nach der Landung in Abhängigkeit vom Erfahrungsniveau ... 150
Abb. 105 Laboraufbau im Experiment C ... 154
Abb. 106 Ablaufschema des Experiments C ... 157
Abb. 107 Einfluss des Systemverhaltens auf die Effektivität der Interaktion ... 160
Abb. 108 Einfluss des Systemverhaltens auf die Anzahl der Nutzereingaben (links) und Gesamtdauer der Interaktionen in Sekunden (rechts) während der erfolgreichen Übermittlung aller Dienstbeschreibungen der Klasse A ... 161
Abb. 109 Verteilung der Ursachen von Kommunikationsproblemen ... 162

Abb. 110 Verteilung der Reaktionen der Probanden nach Auftreten eines Kommunikationsproblems 163
Abb. 111 Verteilung der Nutzereingaben pro Dienstbeschreibung bis zum Dialogabbruch 164
Abb. 112 Verteilung der Interaktionsdauer pro Dienstbeschreibung in Sekunden bis zum Dialogabbruch 164
Abb. 113 Beobachtete Problemlösungsstrategien bei nonverbalen Verständigungsproblemen 169

Tabellenverzeichnis

Tab. 1 Typische Datenlinkkonfigurationen für kleine UAS unter BVLOS-Bedingungen ... 13
Tab. 2 Hauptkomponenten, Parameter und Ausprägungen des manuellen Kanals nach [99] ... 28
Tab. 3 Literaturübersicht der betrachteten Arbeiten in der cHDI ... 40
Tab. 4 Literaturübersicht der betrachteten Verfahren zur visuellen Erfassung von Körpermerkmalen in der cHDI ... 50
Tab. 5 Literaturübersicht der betrachteten Gestenerkennungsverfahren in der cHDI ... 55
Tab. 6 Literaturübersicht geeigneter Datensätze zur nonverbalen cHDI ... 58
Tab. 7 Literaturübersicht visueller Rückmeldeverfahren für dein Einsatz auf UAVs ... 62
Tab. 8 Einteilung der Nutzergruppen nach nonverbalem Ausdrucksvermögen, Funktionsbedarf und Wissensstand über das Kommunikationssystem ... 72
Tab. 9 Sicherheitsbezogene Anforderungen an ein visuelles Kommunikationssystem ... 78
Tab. 10 Schichtenmodelle von OSI, TCP/IP und VisCom im Vergleich ... 81
Tab. 11 In einer Voruntersuchung ermittelte Dienstbeschreibungen nach der Komplexitätsreduktion ... 93
Tab. 12 Der erzeugte Datensatz für statische Armstellungen ... 105
Tab. 13 Klassifikationsleistung des Merkmals Armstellung (rechter Arm) ... 105
Tab. 14 Implementierte, generische Bewegungsarten ... 106
Tab. 15 Der erzeugte Datensatz für Erkennung von beidarmigen Zeichengesten ... 111
Tab. 16 Klassifikationsleistung für das Merkmal „Bewegungspfad“ ... 112
Tab. 17 Übersicht des erzeugten Datensatzes für Handformen und die Detektionsleistung des implementierten Objektdetektors, im Anhang A4 befinden sich zusätzliche Abbildungen zu den jeweiligen Handformen ... 114
Tab. 18 Übersicht der im Experimentalsystem verfügbaren semantischen Kontexte ... 116
Tab. 19 Übersicht der implementierten Dienste und Betriebsmodi im Experimentalsystem, Elemente in Klammern stellen Platzhalter für die entsprechende semantische Gruppe von Symbolen dar ... 121
Tab. 20 Im Probandenversuch ermittelte Klassifikationsleistung des Experimentalsystems (Briefing Mode) ... 126
Tab. 21 Übersicht der Messgrößen zur Operationalisierung im Experiment A ... 132
Tab. 22 Erkennungsdauer in Sekunden für nonverbale Notsignale mittels der Gestikulationsstärke, Standardabweichung in Klammern ... 135
Tab. 23 Vorgaben für Probanden in Experiment B ... 142
Tab. 24 Übersicht der Messgrößen zur Operationalisierung im Experiment B ... 145
Tab. 25 Vergleich der mittleren Fluggeschwindigkeiten (SD in Klammern) ... 149
Tab. 26 Übersicht der zu übermittelnden Dienstbeschreibungen inklusive Anmerkungen ... 155
Tab. 27 Signalpaare mit identischen Darstellungen ... 155
Tab. 28 Signalpaare mit minimal unterschiedlichen Darstellungen ... 155
Tab. 29 Verwendetes Kodierschema A zur Annotation der Ursachen von Kommunikationsproblemen ... 158
Tab. 30 Verwendetes Kodierschema B zur Annotation der Reaktionen nach Kommunikationsproblemen ... 158
Tab. 31 Übersicht der Untersuchungsinhalte und Messgrößen zur Operationalisierung im Experiment C ... 159
Tab. 32 Ermittelte Daten der Interaktionen, bei denen alle Dienstbeschreibungen der Klasse A erfolgreich übermittelt werden konnten ... 160

1 Einführung

Während zu Beginn des Jahrtausends unbemannte Flugsysteme (*Unmanned Aircraft Systems*, UAS) vorrangig im militärischen Nutzungskontext anzutreffen waren, finden miniaturisierte Varianten vermehrt Einzug in den zivilen Alltag als fliegende Kameras, Paketzusteller und Rettungshelfer in Not. Der stetige Ausbau intelligenter Unterstützungssysteme für eine sichere Flugführung, Energieversorgung und Kollisionsvermeidung führt zu Veränderungen im Umgang mit dieser Art von Fluggeräten. Neben speziell geschulten Experten, nutzen vermehrt auch Laienanwender, Rettungskräfte und Behörden miniaturisierte UAS [2, S. 79] als fliegende Hilfsmittel, sodass auch die Bedürfnisse dieser Anwendergruppen zunehmend in den Fokus der Forschung rücken.

Die steigende Nachfrage nach UAS-basierten Anwendungen trägt bereits jetzt maßgeblich zur Erhöhung der Flugbewegungen bei und erfordert neue Regularien und Verfahren für die sichere Koexistenz von bemannter und unbemannter Luftfahrt. Aktuelle Hochrechnungen [3, 4] gehen davon aus, dass bereits in den 2030er-Jahren UAS den europäischen Luftraum dominieren werden. Für das Jahr 2050 geht man von rund sieben Millionen Freizeitdrohnen und etwa 400.000 für behördliche und kommerzielle Zwecke genutzte UAS in Europa aus [4, S. 4]. Neben dem militärischen Markt gilt der zivile Sektor als maßgeblicher Treiber für die rasante Verbreitung dieser Technologie. So werden unbemannte Fluggeräte (*Unmanned Aerial Vehicles*, UAVs) nicht mehr nur als Sensorplattformen verwendet, sondern helfen bereits dabei, Feuer an schwer zugänglichen Orten zu löschen [5], Schwimmhilfen zu Menschen in Seenot zu transportieren [6] oder etwa zeitkritische Blutproben in Labore zu fliegen [7]. Die bereits stattfindende Transformation von der passiven Beobachterrolle hin zu einem mit der Umgebung aktiv interagierenden, robotischen Flugsystem (*Aerobot*[1]) mit manipulatorischen Fähigkeiten stellt den nächsten Schritt in der Evolution von UAVs dar [10]. Umso wichtiger erscheint es, dass zukünftige UAVs auch über Fähigkeiten verfügen müssen, mit ihrer lokalen Umgebung interagieren zu können. Im Fokus bisheriger Forschung stand die Kommunikation mit anderen fliegenden (*Vehicle-to-Vehicle,* V2V) oder am Boden befindlichen (*Vehicle-to-Infrastructure,* V2I) technischen Systemen [11–16].

Ein bisher noch wenig untersuchter, aber zunehmend wichtiger Bereich wird in der englischsprachigen Literatur als *collocated Human-Drone Interaction* (cHDI[2]) bezeichnet [17, 18]. Bei kollokierten[3] Anwendungen teilen sich UAV und Nutzer denselben physikalischen Raum und beeinflussen sich dadurch gegenseitig [20, S. 2029]. Sei es bei der direkten Interaktion mit Hilfesuchenden in entlegenen Gebieten, beim angeleiteten Aufnehmen und Absetzen von Transportgut oder bei der kollaborativen, kleinräumigen Zusammenarbeit von Mensch und UAV zur Erfüllung gemeinsamer Missionsziele. In diesen Situationen kann nicht immer davon ausgegangen werden, dass beim (möglicherweise ungeschulten) Nutzer überhaupt ein zum UAV kompatibles Kommunikationsgerät zur Verfügung steht. Daher werden alternative und nutzerzentrierte Schnittstellen für die direkte Mensch-UAV-Interaktion von der Forschergemeinde gefordert [2, S. 79].

---

[1] Akronym für **Aer**ial R**obot**. Ursprünglich eine Bezeichnung aus der Raumfahrt für unbemannte, flugfähige Erkundungsvehikel für interplanetare Missionen. Der Begriff wird aber zunehmend auch für terrestrische Multirotor-UAVs verwendet, die mit Aktuatoren und Manipulatoren ausgestattet sind [8, 9].

[2] Um den Bezug zur englischsprachigen Literatur zu erhalten, werden in dieser Arbeit die englischen Akronyme verwendet.

[3] In Anlehnung an das Konzept der „Kollokation“ aus den Sprachwissenschaften [19, S. 105], d.h. dem häufigen gemeinsamen Auftreten bestimmter Wortverbindungen (z.B. „dunkle Nacht“ oder „dickes Buch“), wird in dieser Arbeit zur Verdeutlichung der räumlichen Nähe von Mensch und UAV das Adverb „kollokiert“ verwendet.

1.1 Motivation

Einen alternativen und zugleich natürlichen Weg der Kommunikation bietet die Verständigung mittels visueller Signale in Form von graphischen Symbolen, Licht- oder Handzeichen, welche unter dem Oberbegriff der *visuellen Kommunikation* zusammengefasst werden. Standardisierte Hand- und Lichtzeichen stellen insbesondere in der Luftfahrt eine etablierte Alternative zu funkbasierten Kommunikationssystemen dar und bilden einen wichtigen Pfeiler einer resilienten Kommunikation sowohl zwischen Flugzeugbesatzungen in der Luft als auch zwischen Luftfahrzeugführern und Bodenpersonal (Abb. 1).

Abb. 1 Nonverbale Landeeinweisung einer NH-90 Besatzung auf der niederländischen Fregatte Zr. Ms. Evertsen, übernommen aus [21]

Bildgebende Sensoren gehören bereits jetzt zur Grundausstattung der allermeisten unbemannten Fluggeräte, werden jedoch primär zu Mess- oder Aufzeichnungszwecken verwendet. Die Nutzung dieser Sensoren als Kommunikationsschnittstelle ist insofern attraktiv, als dass dadurch die Umweltwahrnehmung von UAVs um eine menschliche Komponente erweitert und gleichzeitig eine alternative Zugriffsmöglichkeit auf das System bei Datenlinkausfällen geschaffen wird. Darüber hinaus erleichtert eine visuelle Benutzerschnittstelle auch eine nahtlose Integration von unbemannten Transportsystemen in bestehende menschenzentrierte Roll- und Landeverfahren an Land (Flugplätze) und zu Wasser (Flugzeugträger). Der visuelle Austausch von Informationen und Intentionen in einem nonverbalen Dialog eröffnet zudem gänzlich neue Interaktionsanwendungen in der kleinräumigen Mensch-UAV-Interaktion. Ähnliche Konzepte für natürliche Mensch-Maschine-Schnittstellen werden bereits in der Robotik erfolgreich eingesetzt, jedoch können diese Ansätze nicht ohne Weiteres auf die Interaktion mit fliegenden Systemen übertragen werden. Neben den methodischen Grundlagen zur technischen Umsetzung einer geeigneten, visuellen Mensch-UAV-Schnittstelle, sind insbesondere die Vorgänge während einer solchen berührungs- und lautlosen Mensch-UAV-Interaktion in der Literatur bislang wenig erforscht.

1.2 Aufbau der Arbeit

Der thematische Fokus der vorliegenden Arbeit liegt somit auf dem bewussten Einsatz nonverbaler Körpersignale in kommunikativer Absicht zur direkten Interaktion mit unbemannten Fluggeräten, zu denen keine Funkverbindung aufgebaut werden kann, weil eine entsprechende Infrastruktur fehlt oder nicht verfügbar ist. Neben dem aktuellen Einführungskapitel gliedert sich diese Arbeit in sechs weitere Abschnitte (Abb. 2).

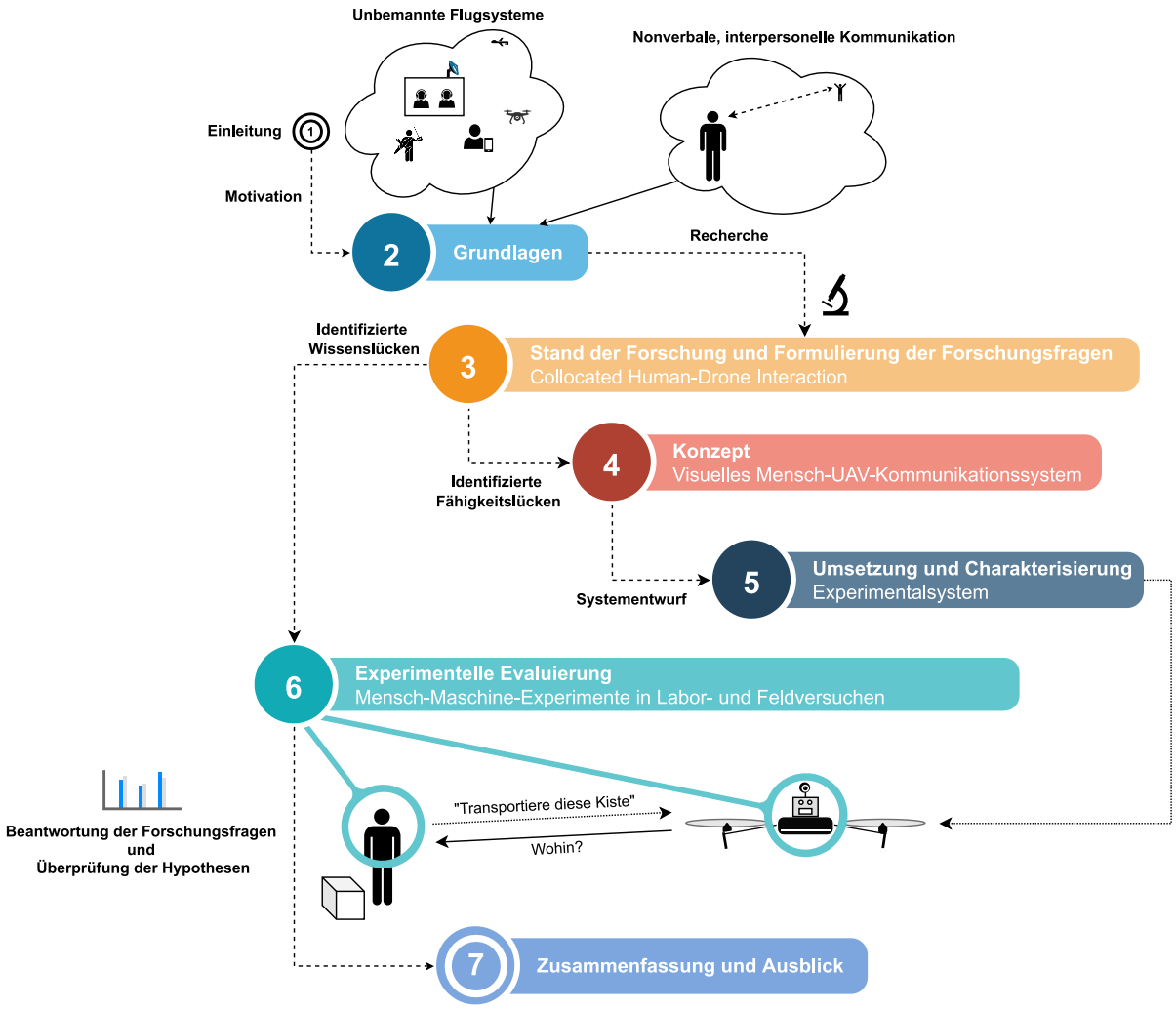


Abb. 2 Inhaltlicher Aufbau der Arbeit

Kapitel 2 bietet einen Einstieg in die zwei Hauptthemengebiete dieser Arbeit. Im ersten Teil wird eine Einordnung der verschiedenen UAS-Varianten vorgenommen und deren aktuelles Nutzungsspektrum im militärischen und zivilen Kontext aufgezeigt. Darüber hinaus wird hier auch auf die generelle Problematik funkbasierter Schnittstellen für die Kommunikation mit unbemannten Fluggeräten eingegangen. Im zweiten Teil wird beschrieben, welche klassischen Ansätze zur Charakterisierung gestischer Äußerungen existieren und unter welchen Bedingungen Hand- und Lichtsignale zur nonverbalen, interpersonellen Kommunikation, insbesondere im Bereich der Luft- und Raumfahrt, genutzt werden. Basierend auf der Taxonomie der Deutschen Gebärdensprache wird anschließend eine systematische Einordnung derjenigen visuell erfassbaren Körpermerkmale vorgenommen, die im zwischenmenschlichen Bereich in kommunikativer Absicht verwendet werden.

Kapitel 3 zeigt den aktuellen Stand der Forschung in der cHDI auf und legt dabei einen besonderen Fokus auf den visuellen Informationsaustausch sowie den Vorteilen des optischen Kanals für die Interaktion mit fliegenden Systemen gegenüber anderen Kommunikationskanälen. Neben aktuellen bildbasierten Methoden zur markerfreien Erfassung menschlicher Körpermerkmale und deren computergestützten Klassifizierung, behandelt dieser Abschnitt auch die Frage, mit welchen Zielvorstellungen Gestenvokabulare üblicherweise entwickelt werden und welche Funktionen in klassischen Gestenerkennungssystemen fehlen, die jedoch für eine visuelle Mensch-UAV-Benutzerschnittstelle unter realistischen Einsatzbedingungen benötigt werden. Dieses Kapitel enthält zudem den gegenwärtigen Erkenntnisstand über das Raumnutzungsverhalten von Menschen in der unmittelbaren Nähe von schwebefähigen kleinen UAVs und stellt geeignete visuelle Rückmeldeverfahren vor, die für die Übermittlung von Systemantworten genutzt werden können. Auf der Grundlage der identifizierten Wissenslücken innerhalb der cHDI werden in diesem Kapitel anschließend die Forschungsfragen dieser Arbeit formuliert.

Kapitel 4 stellt das interdisziplinäre Konzept für ein visuelles Kommunikationssystem vor, welches die identifizierten technischen Fähigkeitslücken adressiert und damit eine möglichst realitätsnahe Evaluierung der nonverbalen Mensch-UAV-Interaktion ermöglicht. Dazu werden zunächst operationelle Annahmen für den Systementwurf aufgestellt, systemische als auch nutzerzentrierte Anforderungen an eine visuelle Mensch-UAV-Schnittstelle formuliert und geeignete Modelle zur Beschreibung der Kommunikationsform, des Interaktionsverhaltens und der Informationsübermittlung mittels visuell erfassbarer körperlicher Subsignale vorgestellt. Im letzten Teil des Kapitels wird eine symbolbasierte Mensch-UAV-Verkehrssprache vorgeschlagen, die eine nonverbale Beschreibung von Nutzerintentionen in einem dialogähnlichen Vorgang erlaubt.

Kapitel 5 enthält die technische Umsetzung des Konzepts in ein echtzeitfähiges Experimentalsystem und dessen Integration in eine schwebefähige Flugplattform. Neben der Beschreibung der systemischen und algorithmischen Realisierung der Systemkomponenten wird in diesem Abschnitt auch die Klassifikationsleistung der implementierten statistischen Verfahren zur Mustererkennung bewertet.

Kapitel 6 widmet sich der Evaluierung der Gesamtleistung des Experimentalsystems bezüglich der Erkennung nonverbaler Nutzereingaben und der Beantwortung der formulierten Forschungsfragen mit Hilfe von Mensch-Maschine-Experimenten unter Labor- sowie Realflugbedingungen. Hierfür wurden drei geeignete Anwendungsfälle als Experimente nachgebildet, um verschiedene Aspekte der nonverbalen Mensch-UAV-Interaktion zu beleuchten.

Kapitel 7 spannt einen Bogen über die gesamte Arbeit, stellt deren Beiträge zum Stand der Forschung gebündelt dar und fasst die gewonnenen Erkenntnisse zusammen. Im Ausblick werden auch potentielle Anknüpfungspunkte für weitergehende Untersuchungen vorgeschlagen.

2 Grundlagen

Dieses Kapitel bietet einen Einstieg in die zwei Hauptthemengebiete dieser Arbeit. Der erste Abschnitt zeigt auf, welche Typen von UAVs existieren, wofür sie klassischerweise im militärischen und kommerziellen Kontext genutzt werden und welche Probleme bei der Nutzung funkbasierter Technologien zur Kommunikation mit unbemannten Fluggeräten entstehen können. Der zweite Teil dieses Kapitels widmet sich der interpersonellen, visuellen Kommunikation mit einem thematischen Schwerpunkt auf der gestischen und lichtbasierten Verständigung im Bereich der Luft- und Raumfahrt als alternatives Kommunikationsverfahren. Eine tiefergehende Betrachtung etablierter Handzeichensysteme kann erste Hinweise darauf geben, welche systemischen Komponenten für die technische Umsetzung einer intuitiv bedienbaren, visuellen Mensch-UAV-Schnittstelle benötigt werden und welche Besonderheiten bei dieser Interaktionsform beachtet werden müssen. Im Anschluss daran folgt eine systematische Beschreibung der nonverbalen Kommunikationsmittel visueller Handzeichensysteme am Beispiel der Deutschen Gebärdensprache.

2.1 Unbemannte Flugsysteme

Vor den ersten bemannten Erprobungsflügen der Gebrüder Montgolfier 1783 (Heißluftballon) oder der Gebrüder Wright 1901 (Doppeldecker-Gleitapparat) gingen zahlreiche unbemannte Tests voraus [22]. Im Gegensatz zu heutigen Systemen verfügten diese ersten unbemannten Prototypen jedoch über keine Möglichkeit der Fernsteuerung. UAS bestehen typischerweise aus mindestens drei Hauptkomponenten: einer stationären oder mobilen Bodenkontrollstation (BKS), dem eigentlichen Fluggerät (*Unmanned Aerial Vehicle*, UAV oder *Unmanned Aircraft*, UA), sowie einer funkbasierten Kommunikationsschnittstelle, über die beide Teilsysteme untereinander Daten austauschen. Im Gegensatz zu ferngesteuerten Modellfluggeräten, bei denen das manuelle Fliegen innerhalb des Sichtbereichs des Piloten (*Visual Line of Sight*, VLOS) im Vordergrund steht, transportieren UAVs Nutzlasten in Form von Sensoren und Messequipment, welche für die Durchführung einer Aufgabe meist außerhalb der Sichtweite des Fernpiloten (*Beyond Visual Line of Sight*, BVLOS) benötigt werden. Zahlreiche weitere Sensoren zur Lage- und Positionsbestimmung sowie ein Flugkontrollsystem (*Flight Control System*, FCS) erlauben dabei eine automatisierte Flugdurchführung, bei der eine kontinuierliche Steuerung des Fluggerätes durch einen Menschen nicht erforderlich ist.

2.1.1 Einordnung

Aufgrund der vielfältigen Einsatzmöglichkeiten von UAS im militärischen und kommerziellen Kontext und den resultierenden Anforderungen an das Fluggerät bezüglich Flugleistung, Nutzlastkapazität und operationellen Fähigkeiten, wurden im Laufe der Zeit eine Vielzahl von UAV-Varianten entwickelt und diverse Ansätze zur Kategorisierung dieser Fluggeräte in verschiedenen Abstufungen vorgestellt [23–27]. In dieser Arbeit erfolgt die Unterscheidung anhand der Art der Auftriebserzeugung als auch dem Grad der Führungsautomation und -autonomie. Zur weiteren Klassifizierung dient der Ansatz von Bento [28, S. 55], der unbemannte Fluggeräte im Hinblick auf deren Abflugmasse und operative Flughöhe in drei Hauptkategorien[4] einteilt: *kleine UAVs (nano, micro, mini)*, *taktische UAVs*

[4] In der beschriebenen Taxonomie existiert eine vierte Kategorie von *„UAVs mit Spezialaufgaben“*, welche jedoch in dieser Arbeit nicht näher betrachtet wird und daher nicht aufgeführt ist.

und *strategische UAVs*. Für eine vertiefende Lektüre zu unbemannten Flugsystemen werden dem interessierten Leser die Arbeiten von [29, 30] empfohlen.

Arten der Auftriebserzeugung

Fluggeräte, welche aerodynamischen Auftrieb durch Vorwärtsbewegung eines feststehenden Flügels erzeugen, gehören zur Kategorie der *Starrflügler* und entsprechen dem allgemeinen Verständnis eines Flugzeugs. In dieser Konfiguration ist ein energieeffizienter Reiseflug über große Strecken möglich, für Start und Landung wird jedoch eine ausreichend lange und präparierte Piste vorausgesetzt. *Drehflügler* hingegen erzeugen Auftrieb nach dem Prinzip eines rotierenden Flügels und besitzen die Fähigkeit, ohne Vorwärtsbewegung abzuheben und landen zu können (*Vertical Take-Off and Landing*, VTOL). Hierzu zählen klassische Helikopter wie auch Multikopter in verschieden Rotorkonfigurationen (z.B. Quad-, Hexa- und Oktokopter). Eine Kombination aus Starr- und Drehflüglern stellen Hybridsysteme dar, die auch als *Transitionsflieger* bezeichnet werden [31]. Diese verfügen über schwenkbare Flügel (*Tilt-Wing*), Rotoren (*Tilt-Rotor*) oder einen schwenkbaren Rumpf (*Tilt-Body*) und beherrschen somit sowohl den Schwebeflug als auch einen energetisch effizienteren Horizontalflug. *Schwingflügler* (Ornithopter) bilden eine weitere Klasse an Fluggeräten, deren Auftriebserzeugungsprinzip sich am Flügelschlag von Vögeln orientiert [32]. Der Vollständigkeit halber seien an dieser Stelle auch diejenigen UAV-Konfigurationen erwähnt, welche aerostatischen Auftrieb nach dem „Leichter-als-Luft"-Prinzip erzeugen (*Aerostaten*), wie beispielsweise unbemannte Luftschiffe.

Automatisierung und Autonomie

Automatisierung und Autonomie werden manchmal stellvertretend verwendet, dabei beziehen sich beide Begriffe auf unterschiedliche Eigenschaften eines technischen Systems. Die DIN 19233 definiert *Automatisierung* als „das Ausrüsten einer Einrichtung, sodass sie ganz oder teilweise ohne Mitwirkung des Menschen bestimmungsgemäß arbeitet" [33]. Derlei „Einrichtungen" finden sich in fast allen UAVs als Teilfunktionen eines Autopilotensystems wieder, dessen Aufgabe es ist, die Fluglage des Fluggeräts in unterschiedlichen Flugphasen zu stabilisieren oder vorprogrammierte Wegpunkte mit möglichst minimalen Abweichungen abzufliegen. Der Automatisierungsgrad beschreibt somit den Anteil dieser automatisierten Funktionen an den Gesamtfunktionen des Flugsystems. Darüber hinaus existiert der Begriff der *Autonomie*, der sich auf die Fähigkeit eines Systems bezieht, eigene Entscheidungen auf der Basis von Umweltwahrnehmung und vorhandenem oder erschlossenem Wissen treffen zu können, um vorgegebene Ziele zu erreichen. Die Literatur enthält mehrere formale Definitionen zur Autonomie von fliegenden Systemen [34–36], sodass eine erschöpfende Definition des Begriffs im begrenzten Rahmen dieser Arbeit nicht möglich und auch nicht beabsichtigt ist. Für eine tiefergehende Lektüre sei deshalb der interessierte Leser auf die genannten Arbeiten verwiesen. Um Missverständnissen im weiteren Verlauf dieser Arbeit vorzubeugen, werden diejenigen automatisierten Systemfunktionen, die aufgrund des notwendigen minimalen menschlichen Einwirkens den Eindruck von Autonomie erwecken, als „hochautomatisiert" oder „intelligent" bezeichnet.

Nano, Micro und Mini UAVs

Fluggeräte dieser Kategorie gehören zu den kleinsten Systemen ihrer Art und operieren im bodennahen Luftraum bis zu einer Höhe von etwa 300 Metern. Nano UAVs, wie der RoboFly [32], sind oftmals nicht größer als Insekten und wiegen nur Bruchteile eines Gramms. Ausgestattet mit miniaturisierten Lochkameras [37] oder spezialisierten Sensoren, beispielsweise zur Messung von Gaskonzentrationen [38], können diese Systeme in engen und schwer zugänglichen Orten Daten sammeln. Systeme bis 100 Gramm bilden die nächstgrößere Klasse von Micro UAVs mit höheren Maximalgeschwindigkeiten und längeren Flugzeiten und entsprechen der Größenordnung von kleinen Vögeln [39]. Der Großteil

der Fluggeräte dieser Kategorie gehört jedoch zur Gruppe der Mini UAVs, die manchmal auch als kleine UAVs bezeichnet werden. Angetrieben von Verbrennungs- oder Elektromotoren und einer Höchstabflugmasse von bis zu 25 kg[5], verfügen diese Systeme über ausreichend Nutzlastkapazität für das Mitführen von Missionsrechnern, Sensoren, Manipulatoren (vgl. robotisches System *Droria* in Abb. 3) oder den Transport von leichten Lasten. Aufgrund der einfachen Anpassbarkeit und den verhältnismäßig geringen rechtlichen Anforderungen für den Flugbetrieb eignen sich die Systeme dieser Kategorie als fliegende Trägerplattformen für vielfältige Mensch-Maschine-Experimente.

Taktische UAVs

Fluggeräte dieser Kategorie werden vorrangig im militärischen Kontext unter BVLOS-Bedingungen betrieben und können bis zu 40 Stunden in mittleren Flughöhen bis 10 Kilometer operieren. Mit maximalen Flügelspannweiten, vergleichbar mit denen von Kleinflugzeugen sowie Abflugmassen von bis zu 1,5 Tonnen benötigen diese Systeme zusätzliche Infrastrukturen für Start und Landung. Dies können Katapultsysteme (z.B. System *Kleinfluggerät Zielortung*, KZO), Fangnetze (z.B. System *LUNA*) oder präparierte Pisten sein, die beispielsweise für das System *Heron 1* benötigt werden. Weitere Unterteilungen beziehen sich auf die Einsatzentfernung von der Bodenkontrollstation (z.B. *close*, *short*, *medium*, *long range*) oder auf die maximale Flughöhe und Einsatzdauer, beispielsweise als „*Medium Altitude Long Endurance*“ (MALE).

Strategische UAVs

In dieser Kategorie finden sich primär Starrflügler mit Spannweiten von bis zu 40 Metern und Abflugmassen von mehreren Tonnen, die in großen Flughöhen bis 20 Kilometer operieren und mehrere Tage bis mehrere Wochen ohne Unterbrechung betrieben werden können (*High Altitude Long Endurance*, HALE). Um diese Flugleistungen zu erreichen, werden entsprechende Mengen an Treibstoff mitgeführt oder die Energieversorgung über Solarzellen bei leichtgewichtigen, hochfliegenden HAPS-Systemen (*High Altitude Pseudo Satellites*) realisiert. Zu den Hauptanwendungen zählt die großräumige Überwachung und Aufklärung sowie die Bereitstellung von Telekommunikationsinfrastrukturen.

[5] National sowie international existieren unterschiedliche Auffassungen darüber, bis wann ein UAS noch als „klein“ bezeichnet werden kann. Zur besseren Unterscheidung von taktischen UAS und noch größeren Systemen wird in dieser Arbeit die Auslegung der Europäischen Union verwendet, die für kommerziell genutzte UAS eine Grenze bei 25 kg zieht [40].

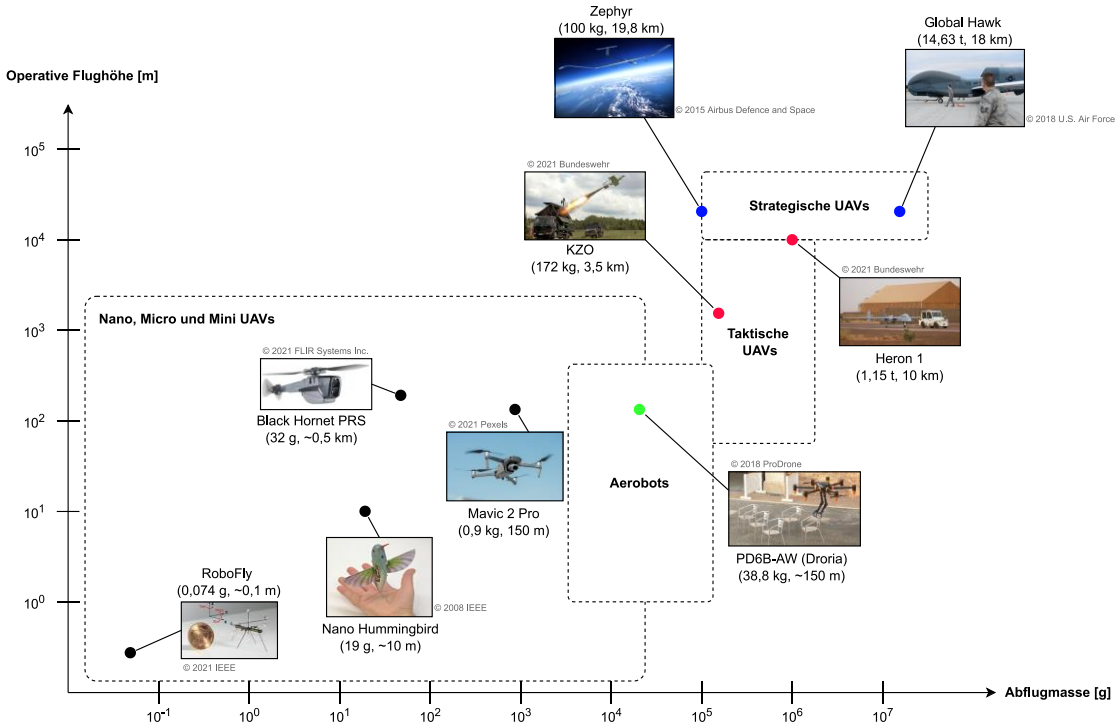


Abb. 3 Bandbreite an unbemannten Fluggeräten[6] bezogen auf die maximale Abflugmasse und operative Flughöhe

2.1.2 Militärische Nutzung von UAS

UAS eignen sich für sogenannte D3-Einsätze (*Dull, Dirty and Dangerous*) [48, S. 8], d.h. für diejenigen Missionen, die für die Flugbesatzung potentiell langwierig, schmutzig (im Sinne einer atomaren, biologischen oder chemischen Kontamination) oder überaus gefährlich sein könnten. Im militärischen Kontext zählt hierzu neben der Zieldarstellung auch der Einsatz von UAS in Krisengebieten zur Überwachung, Zielerfassung sowie zur elektronischen und abbildenden Aufklärung[7]. Das Spektrum der verwendeten Fluggeräte reicht von wenige Gramm leichten Minihelikoptern für die Nahaufklärung bis hin zu tonnenschweren Flugzeugen, die über ein satellitenbasiertes Netzwerk global betrieben werden können. Abhängig von der eingesetzten Flugplattform kommen dabei unterschiedliche Bodenkontrollstationen zum Einsatz. Miniaturisierte Flugsysteme, die in der lokalen Umgebung des Fernpiloten genutzt werden, verfügen zumeist über ein portables Steuergerät, über das Flugführungskommandos versendet und Sensordaten empfangen werden können (Abb. 4 rechts). UAVs der Kategorie MALE und HALE werden typischerweise von mindestens zwei Personen (Fernpilot und Nutzlast-Operateur) aus einer großen, stationären Bodenkontrollstationen aus ferngesteuert und -bedient (Abb. 4 links). Als bewaffnete Versionen (*Unmanned Combat Aerial Vehicle*, UCAV) werden solche Systeme auch zur Bekämpfung von Boden- und Luftzielen eingesetzt.

[6] Abbildungen übernommen aus [32, S. 2, 39, 41–47].
[7] Diese Aufgaben werden unter dem Begriff ISTAR-Operationen (Intelligence, Surveillance, Target Acquisition and Reconnaissance) zusammengefasst.

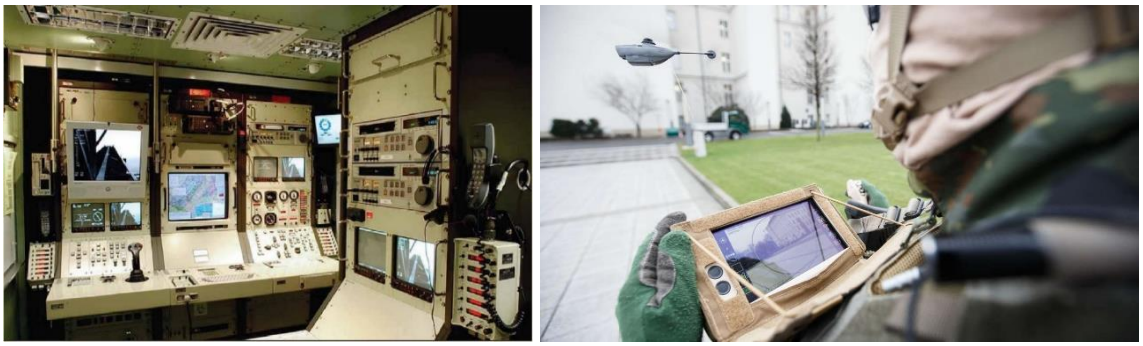


Abb. 4 Stationäre Bodenkontrollstation für ein UAS aus der HALE-Kategorie (links) im Vergleich zum portablen Steuergerät eines microUAS (rechts), übernommen aus [49, S. 27, 50]

Eingliederung taktischer und strategischer UAVs in bestehende Infrastrukturen und Verfahren

Aufgrund ihrer Abmessungen, benötigen taktische und strategische UAVs für Start und Landung entsprechend präparierte Pisten, die zumeist nur an Flughäfen oder auf Flugzeugträgern vorzutreffen sind. Um diese Infrastrukturen nutzen zu können, müssen sich die Fluggeräte in die dortigen Verfahren eingliedern, wozu auch angeleitete Rollmanöver von und zur Parkposition gehören, die von Einweisern (*Marshaller*) durchgeführt werden. Seit der ersten vollautomatisierten Landung des unbemannten Technologiedemonstrators X-47B auf einem amerikanischen Flugzeugträger im Jahre 2013, verstärkt die U.S. Navy ihre Bestrebungen, diese neue Art von Luftfahrzeugen auch in bestehende Abläufe über und unter Deck zu integrieren. Aufgrund der dort vorherrschenden immensen Geräuschkulisse von bis zu 140 dB, findet die Kommunikation zwischen dem Funktionspersonal (*Aircraft Director*, AD) und der Flugzeugbesatzung während des "chaos dance" [51, S. 4] auf dem Flugdeck vorrangig mittels einer gestischen Choreographie standardisierter Handsignale statt. Um dieses Verfahren auch für die Abwicklung von unbemannten Fluggeräten beibehalten zu können, wurde anfangs ein zusätzlicher Sicherheitspilot hinzugezogen, der die Anweisungen des AD interpretieren konnte und sie mittels einer funkbasierten Fernsteuerung (*Control Display Unit*, CDU), in entsprechende Steuerungssignale an das UAV übermittelte (Abb. 5 links).

Abb. 5 Start eines UCAV vom Typ X-47B von einem amerikanischen Flugzeugträger (rechts). Neben dem Funktionspersonal (gelbe Kleidung) ist auch der Sicherheitspilot zu sehen (grüne Kleidung). Die linke Abbildung zeigt das dabei verwendete Steuergerät (CDU). Abbildungen übernommen aus [52, 53].

Aufgrund der nun zusätzlichen Person auf dem ohnehin schon turbulenten Deck, suchte die U.S. Navy nach neuen Lösungen, um die Anzahl des notwendigen Personals wieder zu reduzieren und experimentierte mit dem *on-Deck Intelligent Aircraft Body Language Observer* (DIABLO). Dieser modifizierte

Einweiserstab (Abb. 6) enthält im Inneren neben einer inertialen Messeinheit ein kleines Controllerboard, welches die durchgeführten Gesten klassifiziert und entsprechende Steuerungskommandos an das UAV kabellos übermittelt [54]. Um die Fehleranfälligkeit zu reduzieren, wurde ein endlicher Automat mit acht Zuständen integriert, der in jeder Interaktionsphase nur bestimmte Anweisungen zulässt.

Abb. 6 Funkbasierte Koordination der Rollbewegungen eines unbemannten Tankflugzeugs vom Typ MQ-25 mittels DIABLO, übernommen aus [55]

Nachschubversorgung

Die Eingliederung unbemannter Flugsysteme in die Nachschubversorgung ist weit fortgeschritten. 2015 wurde eine Luftbetankung des UCAV vom Typ X-47B demonstriert [56], sechs Jahre später fungierte bereits Boeings MQ-25 selbst als fliegende Tankstelle und führte eine vollautomatisierte Luftbetankung einer F/A-18 durch [57]. Auch der Transport von Material auf taktischer Ebene mittels unbemannter Multikopter rückt zunehmend in den Vordergrund des Militärs [58].

2.1.3 Kommerzielle Nutzung von UAS

Das weltweit steigende Interesse an UAS spiegelt sich auch auf dem zivilen deutschen Markt wider (Abb. 7). Während sich der Bedarf an UAS im privaten Sektor bereits einem Sättigungsniveau anzunähern scheint, wird dem kommerziellen Bereich ein beachtliches Wachstumspotential für die kommenden Jahre zugesprochen.

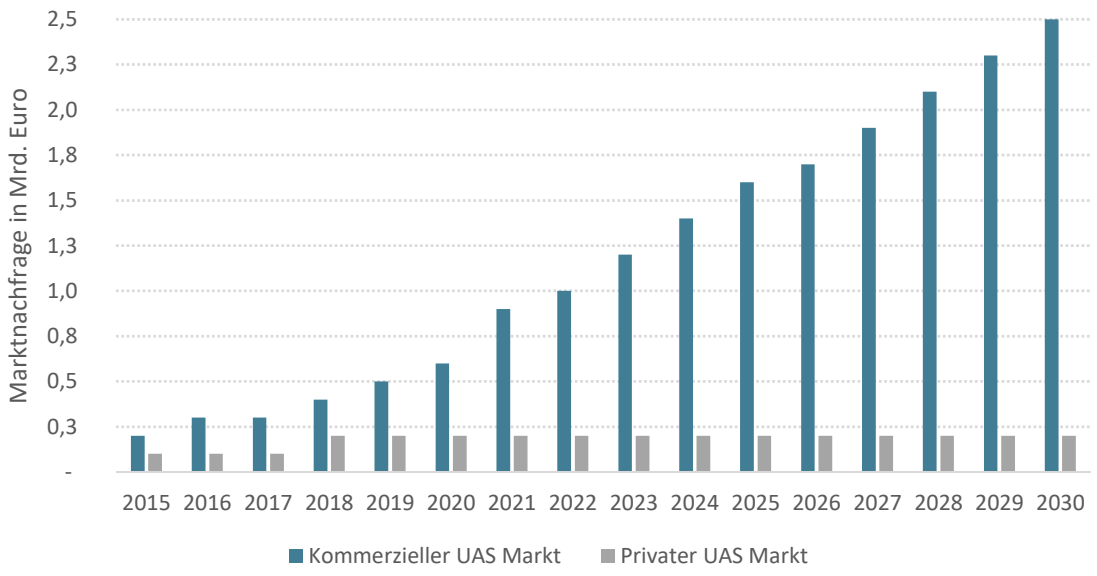


Abb. 7 Prognose der Nachfrageentwicklung auf dem Markt für privat und kommerziell genutzte UAS in Deutschland bis 2030, eigene Darstellung basierend auf [59]

2019 lagen die Hauptanwendungen von kommerziell genutzten UAS vorranging in der Erzeugung von Luftaufnahmen für Vermessungs-, Kartierungs- und Inspektionsaufgaben von Gebäuden und Infrastrukturen, wie Windanlagen und Sendemasten [60, S. 6]. Mit dem Voranschreiten der Entwicklungen insbesondere bei schwebefähigen Plattformen hinsichtlich größerer Nutzlastkapazitäten [61–63], längerer Flugzeiten [64] und der beginnenden Integration dieser Systeme in den urbanen Luftraum [65], finden weitere Nutzlasten den Weg auf die Fluggeräte und ermöglichen damit neuartige Anwendungen.

Welches Potential in UAS für zukünftige kommerzielle Anwendungen liegen könnte, zeigt der Forschungsbericht der *Flying High Challenge*, einer Initiative aus Vertretern von Städten, Behörden und der Wissenschaft, die den urbanen Einsatz von unbemannten Flugsystemen in Großbritannien langfristig fördern möchte und die in Machbarkeitsstudien mögliche Anwendungsfälle für den nutzbringenden Einsatz von UAVs in fünf britischen Stadtregionen[8] in den nächsten 20 Jahren untersucht hat. Fünf potentielle Haupteinsatzgebiete konnten dabei identifiziert werden [66]. Neben der Überwachung, Inspektion und dem Transport von Waren, wird auch im Personentransport (Lufttaxi, Luftkrankenwagen) als auch im Einsatz von UAVs mit robotischen Fähigkeiten (Reparatur von Infrastrukturschäden, Beseitigung von Müll, Unterstützung bei Katastrophen) Potential gesehen. Erste Unternehmen bieten bereits luftgestützte Überwachungssysteme für das eigene Zuhause an [67] oder stellen fliegende Reinigungsroboter für Glasfassaden und Solarpanelen zur Verfügung [68]. Großstädte zeigen zunehmend Interesse daran, Großereignisse mit computergesteuerten UAV-Formationen statt mit klassischer Pyrotechnik zu zelebrieren [69].

Die COVID-19-Pandemie hat den Einsatz von UAS im urbanen Raum zusätzlich beschleunigt. So wurden Multikopter mitunter in China, Dubai, Indonesien, Frankreich und den Vereinigten Staaten nicht nur zur Identifikation von Abstandsunterschreitungen im Rahmen von „*Social Distancing*“ oder zur Messung von erhöhten Körpertemperaturen genutzt, sondern auch zur Desinfektion von ganzen Straßenzügen [70]. Vor allem im BOS-Kontext (Behörden und Organisationen mit Sicherheitsaufgaben) stellen UAS zunehmend ein wichtiges Werkzeug bei der Wahrnehmung hoheitlicher Aufgaben dar.

[8] Bradford, London, Preston, Southampton und West Midland

So setzt beispielsweise die bayerische Polizei seit 2015 UAS zur Überwachung von Großveranstaltungen, der Suche nach Vermissten oder als kostengünstige und wetterunabhängige Alternative zu bemannten Hubschraubern ein [71]. Feuerwehren helfen sie nicht nur, sich einen Überblick vom Ort des Geschehens zu verschaffen, sondern sind mittlerweile auch in der Lage, bei der Brandbekämpfung von schwer erreichbaren hochgelegenen Stockwerken aktiv zu unterstützen [72, 73].

Auch abseits der Städte sind UAS wichtige Werkzeuge zum Schutz von Mensch und Natur geworden. An deutschen Küsten verkürzen sie den Transport von Schwimmhilfen zu Hilfsbedürftigen [6], während sie in Australien dabei helfen, Haie aus der Luft zu erkennen, um von Menschen genutzte Strandabschnitte rechtzeitig räumen zu können [74]. Laut einer repräsentativen Umfrage unter 420 landwirtschaftlichen Betriebsleitern in Deutschland, nutzte 2018 fast jeder zehnte Bauer UAS zur Wildrettung oder zum bedarfsgerechten Einsatz von Düngern, Wasser und Pflanzenschutzmitteln im Rahmen von *Precision Farming* [75]. Auch die Verteilung von Saatgut zur Wiederaufforstung kann über luftgestützte Lösungen effizienter erfolgen als über eine manuelle Aussaat [76].

Beinahe alle großen Versanddienstleister und IT-Großkonzerne bieten mittlerweile luftgestützte Dienstleistungen in ihrem Portfolio, sei es durch Kooperationen mit UAS-Herstellern [77, 78] oder durch Eigenentwicklungen [79, 80]. Die für einen Regelbetrieb notwendigen Technologien, unter anderem zur elektronischen Fernidentifikation und Flugplanaufgabe, dynamischen Routenplanung, Flugüberwachung, Kollisionsvermeidung (*Sense and Avoid*) sowie ein vernetztes UAS Flugverkehrssystem (*UAS Traffic Management*, UTM) sind vorhanden oder befinden sich bereits im Aufbau [81, 82]. Ähnlich verhält es sich mit der Kommerzialisierung von elektrisch angetriebenen VTOL-Flugtaxis. Mit den ersten erfolgreichen Flugerprobungen und der beginnenden Kooperationen von Herstellern, Infrastrukturbetreibern und Behörden zum Aufbau von innerstädtischen Start- und Landeplattformen nimmt die Vision der urbanen Luftmobilität (*Urban Air Mobility*, UAM) zunehmend Gestalt an [83, 84].

2.1.4 Remote Access

Der steigende Automatisierungsgrad und die Vernetzung von UAVs mit der Umwelt über ein übergeordnetes UTM-System [82, 85] erfordern zukünftig nicht mehr das Vorhandensein eines dedizierten Piloten in der Nähe des Einsatzortes. Stattdessen ist anzunehmen, dass aus ökonomischen Gründen geschultes Personal in einem dislozierten Kontrollzentrum Monitoring-Funktionen für eine Flotte von hochautomatisierten UAVs wahrnehmen und nur bei Fehlfunktionen oder Annäherungen an operationelle Systemgrenzen aus der Ferne unterstützend eingreifen wird [86]. Ähnliche Konzepte der Fernsteuerung und -überwachung wurden bereits für den Bereich der Flugsicherung (*Remote Virtual Tower*) vorgestellt und teilweise bereits in den operativen Einsatz[9] überführt [87–90].

2.1.5 Grenzen funkbasierter Kommunikation

Der zunehmende Einsatz funkbasierter Technologien, die einen ähnlichen Bereich des elektromagnetischen Spektrums nutzen, erhöht die Wahrscheinlichkeit von Datenlinkstörungen aufgrund von Überlappungen und Interferenzen. Neben dem Antrieb und der Kontrollflächen, gehört der *„Command and Control“* (C2) Datenlink zu den kritischen Systemen von UAS. Im kommerziellen Bereich nutzen

[9] Am Flughafen Saarbrücken (2018) und London City Airport (2020) sind keine Fluglotsen mehr physikalisch vor Ort, sondern regeln den Flugverkehr mithilfe eines Kameramastes und einer redundant ausgelegten Datenverbindung von Leipzig bzw. von Swanwick aus.

viele Hersteller von kleinen UAS für die Übertragung der Kontroll- und Videodaten lizenzfreie Frequenzbänder[10]. Diese werden jedoch auch von anderen funkbasierten Geräten genutzt (bspw. von vielen WLAN-Geräten), sodass insbesondere in dicht besiedelten Gebieten mit Interferenzen gerechnet werden muss [29, S. 38]. Erschwerend kommt hinzu, dass je nach Flughöhe auch noch weitere Signale (bspw. aus dem Mobilfunknetz) vom Fluggerät aufgefangen werden können, welche in Kombination mit der meist gesetzlich beschränkten, schwachen Ausgangsleistung zu weiteren Signalüberlappungen führen kann.

Beim Betrieb unter BVLOS-Bedingungen werden typischerweise drei Arten von Datenverbindungen zwischen UAV und BKS [92] genutzt, die jeweils über einen oder mehrere Signalwege (terrestrisch, zellular, via Satellit) auf unterschiedlichen Frequenzbändern realisiert werden. Über den sicherheitskritischen *Control and Non-Payload Communication Link* (CNPC) werden schmalbandige Flugsteuerungsanweisungen und Nutzlastparameter zum Fluggerät hochgeladen (Uplink) sowie Telemetrie- und Systemdaten heruntergeladen (Downlink). Über einen separaten und weitaus breitbandigeren *Payload Communication Link* (PCL) können Nutzlastdaten (z.B. Videoströme) heruntergeladen werden. In Tab. 1 ist eine typische Datenlinkkonfiguration für ein kleines UAS dargestellt.

Tab. 1 Typische Datenlinkkonfigurationen für kleine UAS unter BVLOS-Bedingungen

	Control und Non-Payload Communication Link (CNPC)[11]		**Payload Communication Link (PCL)[12]**
Funktion	**Uplink für Kontrolldaten**	**Downlink für Telemetriedaten**	**Downlink für Nutzlastdaten**
Daten	Steuereingaben für Flugplattform Steuerung der Nutzlast	Flug- und Systemparameter	Sensordaten
Übertragungsweg	Terrestrisch, Zellular		
Übertragungsfrequenz	902 – 928 MHz		1,755 – 1850 MHz
Typische Datenraten	ca. 56 Kbit/s		ca. 10,7 Mbit/s

Weil auch der CNPC-Link aufgrund von Interferenzen oder Abschattungen zeitweise oder längerfristig ausfallen kann (*Link-Loss*), verfügen solche Systeme meist über einen speziellen Betriebsmodus (bei kommerziellen Systemen als *Fail-Safe-Mode* bezeichnet), der nach einem kurz- oder längerfristigen Verlust des Uplinks ein vordefiniertes Verhalten des UAVs auslöst. Je nach Konfiguration kann das Fluggerät dabei die Wegpunktnavigation selbsttätig fortführen, eine vorher festgelegte Flugroute abfliegen (z.B. zur Wiedererlangung der Funkverbindung), eine automatische Landung am Startort einleiten (sofern die Energiereserven für einen Rückflug ausreichen) oder eine Sicherheitslandung mittels eines auswurfbaren Fallschirms durchführen (sofern verfügbar). In einer verschärften Form kann auch ein *Flight Termination System* aktiviert werden, welches einen kontrollierten Absturz des Fluggerätes herbeiführt, wenn unter gar keinen Umständen ein zugewiesener Flugbereich verlassen werden darf. Damit derartige Sicherheitsmechanismen jedoch greifen können, müssen sie vor jedem Flug entsprechend konfiguriert werden.

Welche Folgen eine fehlerhafte Konfiguration dieser Schutzmechanismen in Kombination mit einem Datenlinkausfall haben kann, zeigt ein Flugunfall aus dem Jahre 2016 [94]. Die BKS eines Starrflüglers vom Typ *VITO Oculus*[13] meldete direkt nach dem Abflug vom niederländischen Militärflugplatz Weelde (EBWE) mehrere Verbindungsunterbrechungen zum Fluggerät. Dem Piloten gelang es noch rechtzeitig, den Befehl für eine Warteschleife (*Holding-Pattern*) zu übermitteln, bevor nach einigen Vollkreisen die Datenverbindung komplett abbrach und der Autopilot des UAV automatisch in einen

[10] In Deutschland gehören u.a. die Frequenzen im Bereich 2,4 – 2,5 GHz sowie 5,470 – 5,875 GHz zu den sog. ISM-Bändern (Industrial, Scientific, Medical), siehe [91, S. 689].
[11] uAvionix microLink UAX-90045-01 [93].
[12] L-3 Communications Mini UAV Data Link [49, S. 21].
[13] Flügelspannweite 3,80 m, maximale Abflugmasse (MTOW) 55 kg, Reisegeschwindigkeit 110 km/h.

Landemodus umschaltete. Nach mehreren unterbrochenen Landeversuchen stieg das UAV unerwartet auf eine Flughöhe von 4000 Fuß und nahm einen südwestlichen Kurs ein, passierte mehrere kontrollierte Lufträume, überflog eine Landesgrenze und stürzte letztendlich aufgrund von Treibstoffmangel 110 nautische Meilen entfernt auf einem französischen Acker in der Nähe von Dizy-le-Gros ab. Als Ursache für den Link-Loss wurde ein falsch gecrimptes Antennenkabel identifiziert, wohingegen die erfolglosen Landeversuche auf Logikfehler in der Steuerungssoftware zurückzuführen waren. Vorfälle wie dieser zeigen, dass auch ausgeklügelte Sicherheitsmechanismen nur so stark sind wie ihr schwächstes Glied, in diesem Fall ein fehlerhaftes Kabel. Umso wichtiger sind redundante Zugriffsmöglichkeiten auf technische Systeme, vor allem dann, wenn sie derart mobil sind. Während der Warteschleifen in der Nähe des Startflugplatzes befand sich das UAV im Sichtbereich des Fernpiloten und war zudem mit einem stabilisierten Kamerasystem (*Gimbal*) ausgerüstet. Wäre das UAV in der Lage gewesen, über diese Schnittstelle eine Verbindung zum Fernpiloten aufzubauen, hätte dieser eventuell die Auslösung des Notlandesystems (Fallschirm) kommandieren und womöglich das Fluggerät retten können. Funkverbindungen sind anfällig für Störungen und bieten viel Spielraum für Angriffe und Manipulationen [95]. Insofern ist es nicht verwunderlich, dass die Europäische Agentur für Flugsicherheit (*European Union Aviation Safety Agency*, EASA) in Link-Losses das zweitgrößte Risikopotential (neben Kollisionen in der Luft) von UAS sieht [96].

Neben der ingenieurstechnischen Problemstellung, die sich auf die Definition notwendiger Anforderungen an ein visuelles Mensch-UAV-Kommunikationssystem bezieht, damit es in derartigen dynamischen Einsatzumgebungen nutzbringend verwendet werden kann, ergeben sich im gleichen Zuge wissenschaftliche Fragestellungen, die die nonverbale, berührungsfreie und direkte Mensch-UAV-Interaktion als solche betreffen. Welche Besonderheiten müssen bei dieser Form der lautlosen Verständigung im Vergleich zu verbalen Kommunikationsverfahren beachtet werden? Welche nonverbalen Kommunikationsmittel stehen grundsätzlich zur Verfügung und wie können sie genutzt werden, um Informationen zu transportieren? Welchen Einfluss hat die Umgebung auf deren Einsatz? Um erste Anknüpfungspunkte bezüglich der visuellen Interaktion mit technischen Kommunikationspartnern zu identifizieren, kann es von Vorteil sein, sich zunächst anzusehen, wie und unter welchen Umständen Menschen – speziell in der Domäne der Luft- und Raumfahrt – sich mit anderen menschlichen Kommunikationspartnern ohne Lautsprache über den optischen Kanal verständigen.

2.2 Nonverbale interpersonelle Kommunikation

Menschen interagieren und kommunizieren auf vielerlei Kanälen miteinander und mit ihrer Umwelt (Abb. 8). Als nonverbal werden dabei diejenigen Kommunikationskanäle bezeichnet, die ohne eine gesprochene Sprache genutzt werden können[14]. Der visuelle Bereich der nonverbalen Kommunikation basiert hierbei auf zwei Prinzipien: dem Raumnutzungsverhalten im Zusammenspiel mit anderen Individuen (*Proxemik*) und den „intentionalen Bewegungen der Hände, des Rumpfes oder auch des Gesichts“ [97, S. 7], die in kommunikativer Absicht verwendet werden (*Kinesik*). Kendon definiert die Bewegungen der Hände, die annähernd zeitgleich zum gesprochenen Wort durchgeführt werden, als „Gestikulation“. Sprache und Gestikulation stellen nach seinem Verständnis zwei Seiten einer Medaille dar, wobei Gesten dabei einen „integralen Akt des Äußerns“ einnehmen [98, S. 131].

[14] Das schließt den auditiven Kanal nicht gänzlich aus, denn eine Kommunikation kann auch über Vokale oder musikalische Signale erfolgen.

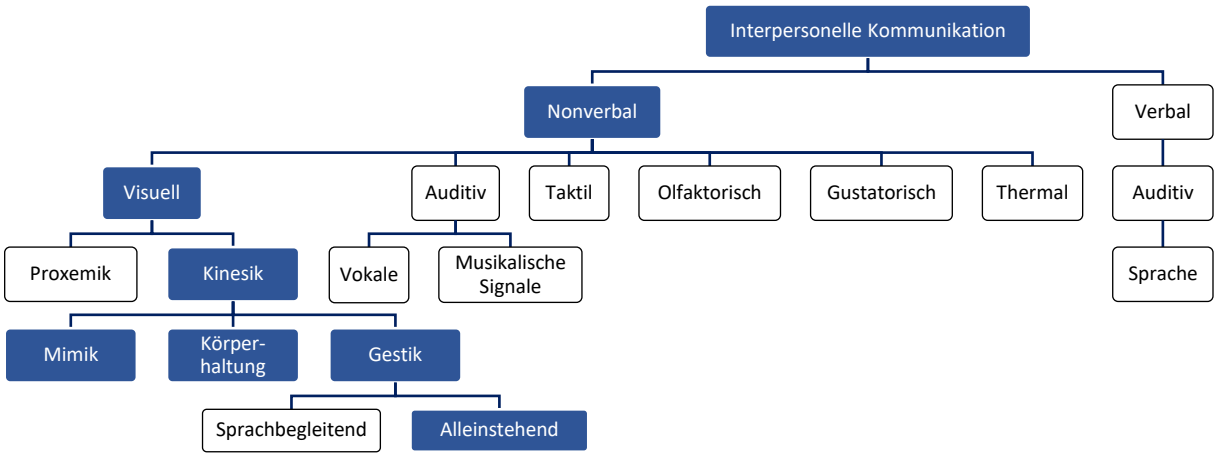


Abb. 8 Interpersonelle Kommunikationskanäle nach [99]. Die blau hinterlegten Elemente beziehen sich auf den visuellen Kanal und werden in dieser Arbeit als Kommunikationsmittel verwendet.

Solche sprachbegleitenden Gesten verstärken für gewöhnlich das Gesagte, sind für sich genommen aber nicht immer eindeutig interpretierbar, da deren Bedeutung nicht nur kulturspezifisch ist, sondern auch von weiteren Faktoren wie soziale Identität, Geschlecht, Alter, Beruf, Bildung und Religion beeinflusst wird [100, S. 124]. Die Gestikulation ist tief in allen Kulturen der Welt verwurzelt und findet in der frühkindlichen Entwicklung noch vor den ersten sprachlichen Äußerungen statt [101, S. 172-178]. Einige Sprachwissenschaftler, wie beispielsweise Bühler mit seiner „Ausdruckstheorie", sind sogar von der prinzipiell gleichen „Leistungsfähigkeit von Gesten und Lauten" [102, S. 40, 103, S. 37] überzeugt. Tatsächlich konnte in einer Vergleichsstudie [104] nachgewiesen werden, dass das Produktionstempo bei Laut- und Gebärdensprache bezogen auf die Produktionsdauer pro Satzinhalt auf einem ähnlichen Niveau liegt, da mit Gebärden mehr Informationen übertragen werden als durch ein Lexem[15] der Lautsprache [105, S. 236]. Produktion bezeichnet in sprachwissenschaftlichen (linguistischen) Kontext die Generierung eines sprachlichen Ausdrucks mittels des entsprechenden Werkzeugs, also des Sprechapparates, des Körpers oder der Hände [106, S. 46].

Sprechen und Hören basieren auf aufeinanderfolgenden (sequentiellen) Vorgängen, wohingegen der visuelle Sinn mehrere, gleichzeitig auftretende Reize auch simultan verarbeiten und voneinander unterscheiden kann. Daher nutzen Gebärdensprachen häufig Formen, die verschiedene linguistische Informationen gleichzeitig übermitteln [106, S. 49]. Man schätzt die Bandbreite der menschlichen Retina auf etwa 8,75 Mbit/Sekunde, was in etwa der Größenordnung einer Ethernet-Verbindung der ersten Generation entspricht [107, S. 1428]. Aus dieser Art der Signalerzeugung und -verarbeitung ergeben sich nach Broszinky-Schwabe [100, S. 118] grundlegende Unterschiede in der verbalen und nonverbalen Kommunikation.

Beginn und Ende von Äußerungen

Verbale formulierte Äußerungen sind zeitlich diskret, verfügen also über einen definierten Anfang und ein Ende, sodass unterschieden werden kann, wann Signale gesendet werden und wann nicht. Nonverbale Signale hingegen werden permanent übermittelt, solange sich beide Gesprächspartner visuell wahrnehmen können. Der Übergang zwischen relevanten und irrelevanten Signalen kann daher fließend verlaufen. Gemäß Watzlawicks erstem Axiom kann man also bei dieser Form der Kommunikation „nicht *nicht* kommunizieren" [108, S. 59].

[15] Ein Lexem ist die grundlegende Einheit eines Vokabulars einer Sprache.

Steuerbarkeit

Lautsprachliche Äußerungen können bewusst eingesetzt und deren Wirkung abgeschätzt werden. Im Allgemeinen trifft das auch auf nichtsprachliche Signale zu, im Speziellen existieren aber auch nonverbale Körpersignale, die nicht direkt steuerbar sind, wie beispielsweise Schwitzen, Erröten oder unbewusste Veränderungen der Hand- oder Körperstellung.

Hierarchie

Nonverbale, gestische Kommunikation hat keine hierarchische Struktur. Anders als bei verbalen Äußerungen, bei denen Wörter aus einzelnen Vokalen und Konsonanten gebildet werden, ergibt die Kombination einzelner Gesten keine neue Bedeutung, sondern stellt unterschiedliche Aspekte eines Sachverhalts dar.

2.2.1 Klassische Ansätze zur Klassifizierung gestischer Äußerungen

Für die Klassifizierung von gestischen Äußerungen existieren in der Literatur mannigfaltige Ansätze, die die Gestikulation aus jeweils unterschiedlichen Blickwinkeln betrachten. Im Folgenden werden diejenigen Konzepte, auf die sich diese Arbeit stützt, kurz vorgestellt.

Klassifizierung anhand des Gestikulationsortes und des Zeitverlaufs

Zur räumlichen und zeitlichen Klassifikation gestischer Äußerungen wird oft Bezug auf zwei Konzepte von McNeill genommen [109]: den „Gestenräumen“ (*gesture spaces*) und den „Gestenphasen“ (*gesture phases*). Das Konzept der „Gestenräume“ unterteilt den Raum um eine Person in mehrere bevorzugte Artikulationsregionen für bestimmte Typen von Gesten (Abb. 9). Bildhafte (ikonische) Gesten werden vorrangig im zentralen Gestenraum (*Center-Center*) produziert, wohingegen hinweisende (deiktische) Gesten weiter außen in der Peripherie vorzutreffen sind. In der extremen Peripherie nimmt die Gestendichte im Allgemeinen dagegen stark ab.

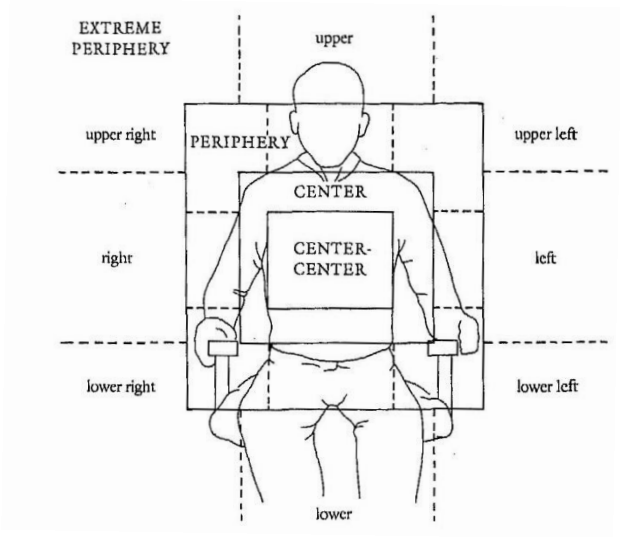


Abb. 9 Typische Gestenräume eines Erwachsenen nach McNeill, übernommen aus [109, S. 89]

Das Konzept der „Gestenphasen“ bezieht sich auf den zeitlichen Verlauf einer gestischen Äußerung und unterteilt diese in vier sukzessive Phasen: *preparation*, *stroke*, *hold* und *retraction*. Aus einer Ruheposition heraus wird der Körper in eine bequeme Ausgangsstellung gebracht (preparation), in der die Ausführung der eigentlichen, bedeutungstragenden Handbewegung (stroke) ermöglicht wird. Bei deiktischen Gesten folgt hierauf meist eine statische Phase, in der die Bewegung unterbrochen und

die Handstellung für eine bestimmte Zeitdauer gehalten wird (hold). Die Rückkehr in die Ruheposition (retraction) definiert das Ende einer gestischen Äußerung.

Klassifizierung anhand der Funktion

Ekman und Friesen untersuchten die Funktion von Gesten und fanden neben der redebegleitenden Verstärkung des Gesagten (*Illustrators*) vier weitere Funktionen von nonverbalen Signalen [110, 111]. Neben Körpersignalen, die den emotionalen Zustand nach außen tragen (*Emotional Expressions*), existieren auch größtenteils unbewusste Handbewegungen ohne klar erkennbaren Zweck, bei denen ein anderer Teil des eigenen Körpers oder Gesichts berührt, gestreichelt, gekratzt oder in anderer Art und Weise manipuliert wird (*Manipulators*). Diese gewohnheitsmäßigen Bewegungen können vermehrt bei Nervosität auftreten. *Regulators* hingegen richten sich vorrangig an einen anderen Gesprächspartner und steuern den Kommunikationsfluss während einer Konversation, bspw. durch eine einladende Handgeste, um jemandem das Wort zu erteilen, oder ein zustimmendes Kopfnicken. Die für diese Arbeit relevanteste Klasse von nonverbalen Äußerungen werden nach dieser Systematik als *Emblems* bezeichnet. Diese wortähnlichen, „autonomen Gesten" [112] sind innerhalb einer Gemeinschaft mit präzisen Bedeutungen belegt und können auch sprachersetzend genutzt werden. Klassische Beispiele hierfür sind die im westlichen Kulturkreis gängige „Daumen-hoch" Geste als positive nonverbale Rückmeldung oder das „Victory-Zeichen" als gestisches Synonym für Frieden [113, S. 82].

Klassifizierung anhand der Abhängigkeit zu einer Sprache

Eine Klassifizierung von Gesten anhand deren Einbettung in eine Sprache erlaubt das System von Morris [114, S. 33-50]. Dabei werden gestische Äußerungen zunächst nach ihrer kommunikativen Funktion sogenannten *sekundären* oder *primären Gesten* zugeordnet (Abb. 10).

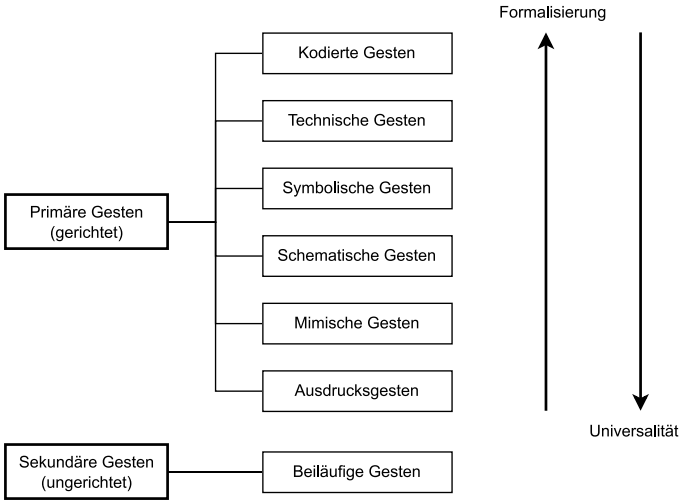


Abb. 10 Taxonomie von Morris [114] zur Einteilung von Gesten nach ihrer kommunikativen Funktion und Grad der Formalisierung

Sekundäre Gesten stellen beiläufige, „mechanische Handlungen mit unbeabsichtigter Aussagraft" [114, S. 33] dar und beziehen sich auf unbewusst durchgeführte, zufällige Bewegungen, die nicht direkt an eine andere Person in kommunikativer Absicht gerichtet sind. Anders verhält es sich mit der Gruppe der primären Gesten, die im Kern eine ausschließlich kommunikative Funktion erfüllen. *Ausdrucksgesten* bilden dabei eine speziesüberreifende Gruppe von universalen Gesichts- und Handsignalen, die Emotionen und Gemütszustände transportieren können. Während *mimische Gesten* möglichst realistisch „Personen, Dinge oder Vorgänge imitieren" [114, S. 38], stellen *schematische Gesten* „stan-

dardisierte Kürzel“ [114, S. 41] dieser Imitationen dar, die jeweils nur ein markantes Merkmale hervorheben und aufgrund der reduzierten Form bei „Nichteingeweihten“ bereits zu Verständigungsproblemen führen können. *Symbolische Gesten* versinnbildlichen abstrakte Eigenschaften und Konzepte, für die es keine materielle Entsprechung in der realen Welt gibt. Kulturell bedingte Unterschiede in der Deutung der jeweiligen Darstellungen erzeugen einen Interpretationsspielraum, der zu Missverständnissen führen kann. *Technische Gesten* stellen weiter spezialisierte Handzeichen dar, die nur innerhalb eines bestimmten Tätigkeitsfeldes oder Fachgebiets zur Kommunikation genutzt werden (z.B. internationales Winkeralphabet[16]) und deren Bedeutung nur für einen kleinen Spezialistenkreis bekannt ist. Darüber hinaus können sie für sich stehen und müssen in keine Syntax eingebettet sein. Im Gegensatz dazu stehen *kodierte Gesten*, welche Bestandteil eines streng formalisierten Zeichensystems sind und nach festgelegten Prinzipien miteinander verknüpft sind, d.h. erst in einem syntaktisch korrekten Gesamtgefüge ihre Bedeutung offenbaren.

2.2.2 Technische Handzeichensysteme in der Luft- und Raumfahrt

In vielen technischen Domänen wurden eigene Handzeichensysteme entwickelt, da äußere Umstände eine verbale Kommunikation verhindern. So existieren spezielle Handzeichensysteme für die Verständigung im Sägewerk [115] oder für die Einweisung von Kran- [116] und Kraftfahrzeugführern [117]. In der zivilen Luftfahrt schreiben die „*Rules of the Air*“ der *International Civil Aviation Organization* (ICAO) vor, welche international standardisierten Handsignale (*Marshalling* oder *Handling Signals*) vom Bodenpersonal (*Signalman*) als auch von der Flugbesatzung während Roll- und Schwebemanövern zu verwenden sind [118]. Auch die nonverbale Verständigung bei Notfällen (*Standard Emergency Hand Signals*), beispielsweise bei Feuer an Fahr- oder Triebwerksteilen, ist hier geregelt.

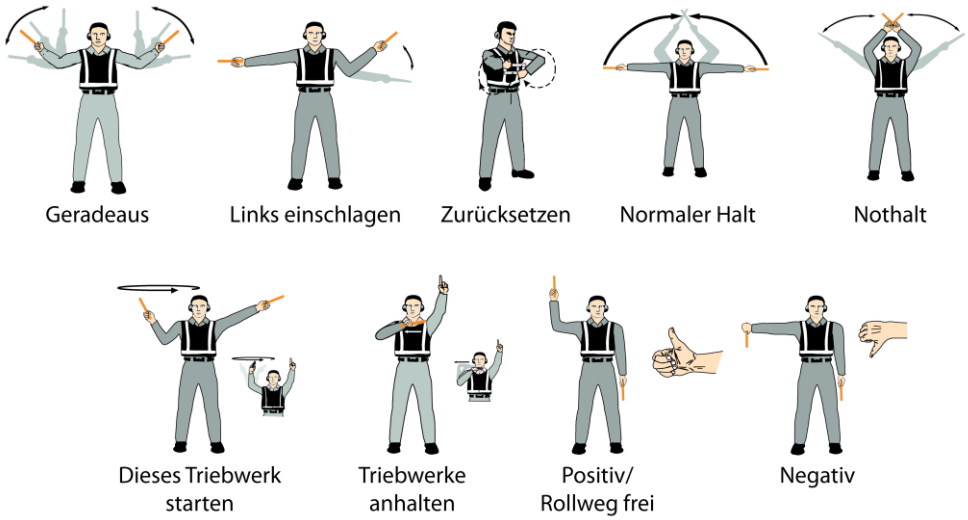


Abb. 11 Auszug aus den ICAO Marshalling Signals für Rollbewegungen, übernommen aus [118, S. 43-49]

Abb. 11 zeigt einen Auszug der darin enthaltenen Handzeichen für die Einweisung von Luftfahrzeugen bei Rollbewegungen am Boden. Bei genauer Betrachtung dieser Handsignale können folgende Darstellungselemente identifiziert werden:

- Periodische vertikale, horizontale und kreisförmige Bewegungen eines oder beider Arme
- Zeigen auf bestimmte Körperteile mit einer Hand/einem Finger
- Berühren eines Körperteils mit einer Hand/einem Finger

[16] Internationales System zur optischen Nachrichtenübermittlung in der Schifffahrt, bei dem einzelne Buchstaben und Ziffern über definierte Armstellungen kodiert werden.

- Kodierung eines Signals über die Handform (Positiv/Negativ)
- Entfernung der Hände zueinander repräsentiert einen qualitativen Wert einer Entfernung (Normaler Halt)
- Die Ausführungsgeschwindigkeit der Bewegung gibt die Stärke der Steuergröße vor (bspw. beim Einlenken)

Betrachtet man zudem die Einweiserzeichen für schwebefähige Luftfahrzeuge (Abb. 12), fällt auf, dass einige Elemente eine hohe Ähnlichkeit aufweisen (Normaler Halt vs. Aufwärts), deren Bedeutung jedoch aufgrund der unterschiedlichen Flugphasen (Rollen vs. Schweben) eindeutig bestimmt werden kann.

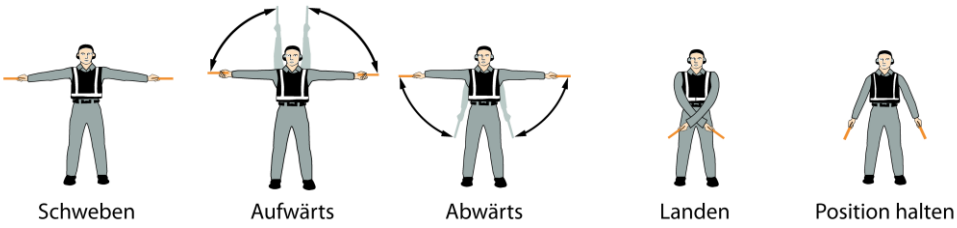


Abb. 12 Auszug aus den ICAO Marshalling Signals für schwebefähige Luftfahrzeuge, übernommen aus [118, S. 49-51]

Im militärischen Kontext werden je nach Teilstreitkraft (Heer, Marine, Luftwaffe), Einsatzumgebung (zu Land, zu Wasser, in der Luft) und Tätigkeit (z.B. Lande-, Rolleinweisung, Vorflugkontrolle) unterschiedliche Handzeichensysteme genutzt [119–121]. So wird beispielsweise während der schnellen Betankung und Aufmunitionierung von fliegenden Einheiten in der Nähe des Operationsraumes an bodengebundenen *Forward Arming and Refueling Points* (FARP) auf die Nutzung funkbasierter Kommunikationssysteme weitestgehend verzichtet, um deren Position durch EM-Abstrahlung nicht zu offenbaren [122, S. 11]. Für den visuellen Austausch numerischer Informationen (z.B. zur benötigten Treibstoffmenge) zwischen Flugbesatzung und Bodenpersonal bei laufenden Triebwerken (*hot refueling*), greift die U.S. Navy auf ein Handzeichensystem zurück, bei dem vertikale und horizontale Fingerstellungen in vorgegebene Zahlengruppen kodiert werden (Abb. 13).

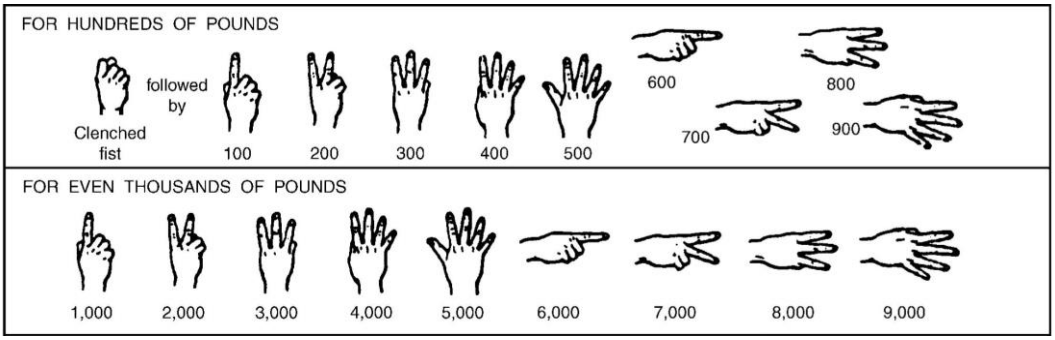


Abb. 13 Handsignale für Kraftstoffmengen der U.S. Navy, übernommen aus [120, S. 92]

Um jedoch eine effektive Zusammenarbeit unterschiedlicher Teilstreitkräfte zu ermöglichen, ist ein gemeinsamer Grundwortschatz an standardisierten Handzeichen für eine „nonverbale Einsatzsprache“ notwendig. Für die NATO-Mitgliedsstaaten gibt beispielsweise das *Standardization Agreement* 3117 [123] das zu verwendende Handzeichensystem vor.

Einfluss der Umgebungsbedingungen auf die nonverbale Darstellung

Nicht alle nonverbalen Darstellungsformen können in jeder Situation genutzt werden. Ein dicker Tauch- oder Raumanzug kann nicht nur die Bewegungsfreiheit, sondern auch die Anzahl physiologisch durchführbarer Gesten unter Umständen stark einschränken. Freizeittaucher [124], Kampf- und Rettungstaucher [125] sowie Höhlentaucher [126] haben einen gemeinsamen Grundwortschatz an 31 Handsignalen, insbesondere für sauerstoff- und druckbezogene Probleme, ergänzt durch weitere einsatzspezifische Gesten. So verwendeten beispielsweise Rettungstaucher der Apollo Missionen ein eigenes Handzeichensystem [127] zur visuellen Kommunikation mit der Bergemannschaft und den Astronauten im Inneren der Landekapsel (Abb. 14 links). Höhlentaucher nutzen darüber hinaus eine eigene Syntax, um bestimmte Sachverhalte in Form von satzähnlichen Konstruktionen zu formulieren. Fragen werden beispielsweise durch eine spezielle Geste (gekrümmter Zeigefinger) eingeleitet, gefolgt vom gewünschten Parameter und einer Zeigegeste auf den Adressaten. Abb. 14 (rechts) zeigt einige solcher tauchgangbezogenen Fragen.

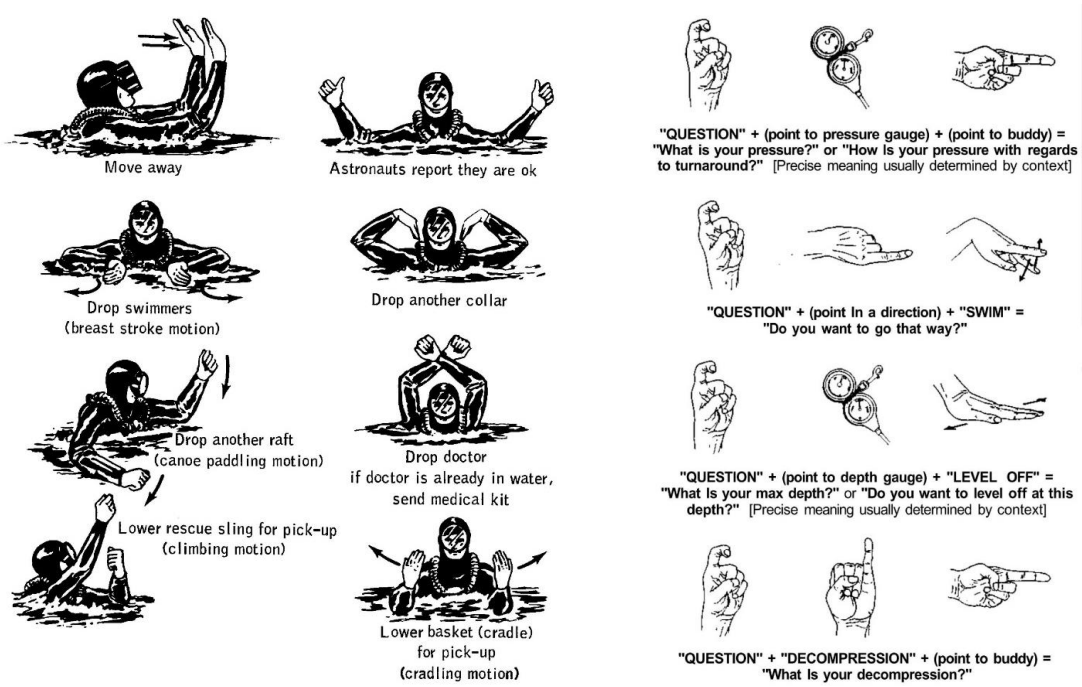


Abb. 14 Handsignale der Apollo-Rettungstaucher (links) und Handzeichensystem von Höhlentauchern (rechts), übernommen aus [126, S. 19, 127, S. 132]

Mit dem Beginn der ersten Außenbordeinsätze (*Extra Vehicular Activity*, EVA) mussten auch Raumfahrtorganisationen geeignete Handzeichensysteme entwickeln, um im Falle eines Ausfalls des funkbasierten Interkom-Systems die Verständigung der Besatzung untereinander sicherstellen zu können. Ein entsprechendes Gestenvokabular muss die motorisch stark eingeschränkte Bewegungsfreiheit der Astronauten im Raumanzug und den Einfluss von reduzierter Gravitation miteinbeziehen. Für zukünftige Missionen auf der Marsoberfläche hat die NASA daher untersucht, welche Kommunikationsabsichten in einem solchen Funkausfallszenario relevant sein könnten und welche Darstellungsformen sich eignen würden, um diese möglichst eindeutig zu übermitteln. Das dabei entwickelte Handzeichensystem enthält viele Elemente aus dem Tauchwesen, welche an die gehende Fortbewegungsart angepasst wurden. Einen Auszug aus diesem Handzeichenvokabular zeigt Abb. 15.

Abb. 15 Auszug aus der Studie der NASA zur extraterrestrischen Kommunikation mittels Handsignalen, übernommen aus [128]

2.2.3 Visuelle Boden-Luft Kommunikation

Die bisher vorgestellten Handzeichensysteme werden vorrangig zur interpersonellen Kommunikation im Nah- und Nächstbereich genutzt. Bei größeren Interaktionsentfernungen oder Höhenunterschieden sinkt die Sichtbarkeit von kleinteiligen Darstellungen für den Signalempfänger, sodass auf eine großräumigere Darstellungsform oder eine gänzlich alternative Form der visuellen Kommunikation zurückgegriffen werden muss. Für die visuelle Boden-Luft-Verständigung existieren in der zivilen und militärischen Luftfahrt mehrere Verfahren, die sich größtenteils darin unterscheiden, ob neben dem eigenen Körper noch weitere Hilfsmittel zur Verfügung stehen, wie groß die Distanz zwischen beiden Kommunikationspartnern ist und unter welchen Umgebungsbedingungen (Tag/Nacht) die Interaktion stattfindet.

Mit Hilfe von Lichtsignalen

In der Nähe von Flughäfen findet die Kommunikation zwischen der Luftfahrzeugbesatzung und dem diensthabenden Fluglotsen in aller Regel über eine Sprechfunkverbindung statt. Sollte jedoch eines der beteiligten Sprechfunkgeräte ausfallen, schreibt die ICAO ein festgelegtes Kommunikationsverfahren über Lichtsignale vor. Dazu sendet der Fluglotse vom Kontrollturm mithilfe einer „Lichtkanone“ durchgängig oder mit Unterbrechungen weiße, grüne und rote Lichtsignale in Richtung des Luftfahrzeugs. Abhängig von dessen aktueller Flugphase (auf Parkposition, auf dem Weg zu/von der Piste, in der Platzrunde) sind die Lichtsignale mit standardisierten Anweisungen kodiert (Abb. 16). Durch die teils sehr hohen Lichtstärken von bis zu 200.000 Candela können diese Signale noch bis zu einer Entfernung von 6,6 Kilometern unter klaren Bedingungen auch am Tage wahrgenommen werden [129]. Bei besonders sicherheitskritischen Anweisungen (z.B. „Nicht landen!“), beispielsweise aufgrund von Schäden an der Infrastruktur, kann auch rot leuchtende Pyrotechnik zum Einsatz kommen.

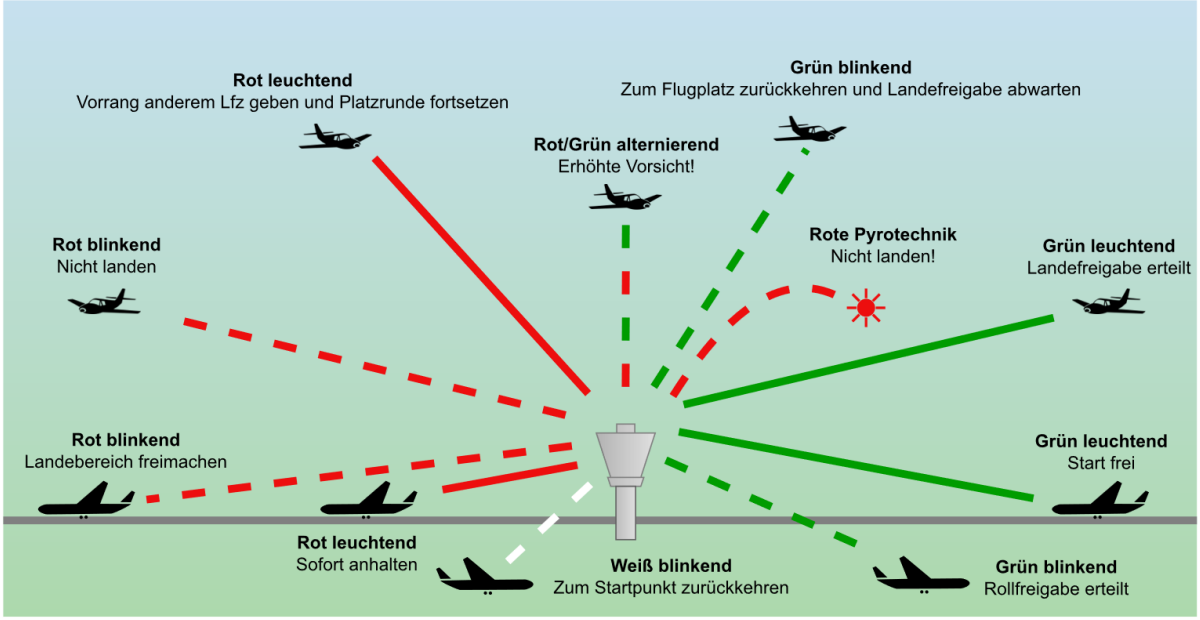


Abb. 16 Visuelles Alternativverfahren zur Übermittlung von Flugfreigaben aus dem Kontrollturm an ein Luftfahrzeug mittels Lichtsignalen, eigene Darstellung basierend auf [118, 40-41]

Mit Hilfe von graphischen Symbolen

Im Gegensatz zu zivilen Anwendungen, können im militärischen Kontext Licht- oder Rauchsignale zu viel Aufmerksamkeit erzeugen und dadurch Rückschlüsse auf die eigene Position nicht nur aus der Luft, sondern auch vom Boden aus zulassen. Ein alternatives Verfahren, welches weitaus mehr Informationen transportieren kann, bietet die Kommunikation mittels großformatiger graphischer Symbole (Abb. 17), die mit verfügbaren Mitteln auf dem Boden gebildet und von fliegenden Einheiten gelesen werden können.

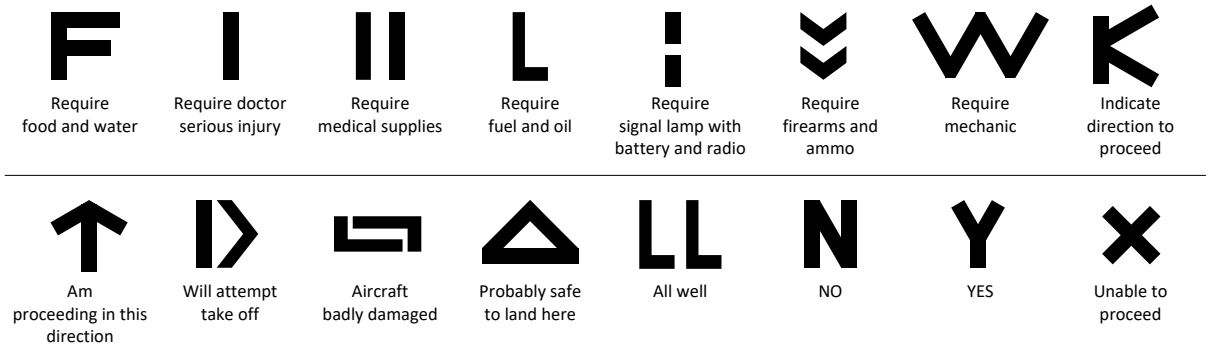


Abb. 17 Symbolische Boden-Luft-Signale der U.S. Army, eigene Darstellung basierend auf [119, S. 80]

Mit Hilfe des eigenen Körpers

Im zivilen Kontext existieren nur wenige, international standardisierte Verfahren zur visuellen Kommunikation mit einem fliegenden Interaktionspartner und diese beschränken sich zudem primär auf die Signalisierung einer Notlage. Das bekannteste Zeichen ist sicherlich das SOS-Signal, welches in Schriftform ausgelegt oder mithilfe des Morsealphabets auf akustischem oder visuellem Wege übermittelt werden kann. Ein entsprechendes nonverbales Seenotsignal existiert auch in der Schifffahrt und wird über ein langsames, synchrones Anheben und Absenken beider Arme übermittelt (sog. „müde Fliege“). Im europäischen Gebirgsraum kommen die „alpinen Notsignale“ zum Einsatz. Um die Bergrettung auf sich aufmerksam zu machen, müssen jeweils im Wechsel sechs akustische oder visu-

elle Zeichen beliebiger Art (Rufen, Leuchten, Winken) innerhalb einer Minute abgesetzt werden, gefolgt von einer Minute Stille. Falls dieses Signalmuster von der Bergwacht wahrgenommen wird, erfolgt eine Antwort mit drei Zeichen pro Minute, gefolgt von einer Minute ohne Signalisierung. Ein weiteres Signal bezieht sich auf den Vorbeiflug oder Anflug eines Rettungstransporthubschraubers und soll der Besatzung die Identifikation der hilfesuchenden Personen erleichtern. Werden beide Arme nach oben ausgestreckt und zu einer „Y-Pose“ geformt, bedeutet das „Ja, ich brauche Hilfe!“. Um anzuzeigen, dass alles in Ordnung ist, sollte die „N-Pose“ genutzt werden, bei der der linke Arm heruntergenommen wird und die Körperform somit den Buchstaben „N“ andeutet (Abb. 18). Auf großräumige Winkbewegungen sollte in diesem Fall verzichtet werden, um die Besatzung nicht zu verunsichern.

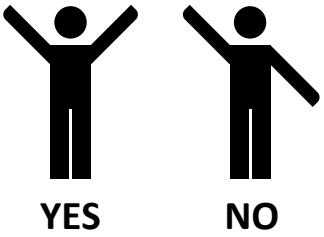


Abb. 18 Internationale Notsignale, eigene Darstellung basierend auf [130]

Denn je nach Anwendungskontext kann dasselbe Signal mit teilweise widersprüchlichen Bedeutungen belegt sein. So kann das großräumige Winken mit einem oder beiden Armen von einem Boot aus als Gruß (Schifffahrt), vom Wasser aus jedoch als Hilferuf und im militärischen Kontext sogar als Aufforderung zum Landeabbruch (vgl. Abb. 19 unten links) verstanden werden.

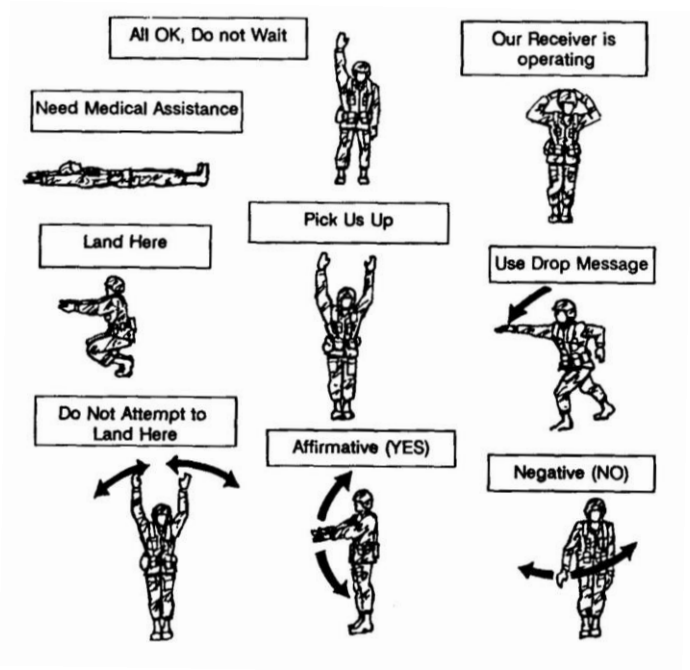


Abb. 19 Visuelle Notsignale der U.S. Army, übernommen aus [131, S. 59]

2.2.4 Visuelle Luft-Luft und Luft-Boden Kommunikation

Befinden sich beide Interaktionspartner in der Luft, erfolgt die Kommunikation im Allgemeinen über redundante Sprechfunkverbindungen. Dennoch ereignen sich jährlich im europäischen Luftraum Störungen im Luftverkehrssystem, die auf Kommunikationsprobleme zurückzuführen sind. Kann beispielsweise über einen Zeitraum von mehreren Minuten trotz mehrmaliger Anrufe durch die Flugsicherung keine Sprechfunkverbindung zu einem Luftfahrzeug hergestellt werden, spricht man von einem *Prolonged Loss of Communication* (PLOC)[17]. Entsteht durch diesen Kommunikationsausfall ein Sicherheitsrisiko für Andere (bspw. durch den unerlaubten Einflug in einen kontrollierten Luftraum mit hohem Verkehrsaufkommen), wird die zuständige militärische Luftraumüberwachungszentrale informiert, die zur Luftraumsicherung zwei Abfangjäger (Alarmrotte) zum fragwürdigen Luftfahrzeug schickt, um sich vor Ort ein Bild von der Situation zu verschaffen. Mittels international festgelegter visueller Luftfahrzeugsignale (*Aircraft Signals*) und spezieller Flugmanöver [118, S. 39] wird versucht, eine rudimentäre Kommunikation mit der Flugbesatzung aufzubauen und diese beim Bestehen einer Notwendigkeit zu einem nahegelegenen Flughafen zu eskortieren. So kann der Pilot des abfangenden Flugzeugs (*Interceptor*) beispielsweise mittels „Flügel-Wackeln" (*„Folgen Sie mir!"*), dem Aus-/Einfahren des Landefahrwerks (*„Landen Sie hier!"*) oder durch ein abruptes Wegdrehen vom abgefangenen Luftfahrzeug (*„Sie dürfen weiterfliegen."*) seine Intentionen visuell übermitteln. Der Pilot des abgefangenen Luftfahrzeugs kann wiederum, zum Beispiel durch das zweimalige Ein-/Ausschalten der Landebzw. der Navigationslichter, mitteilen, dass die Anweisung verstanden wurde oder durch das Einfahren des Fahrwerks beim Überflug über eine Piste anzeigen, dass der gewählte Flughafen oder Flugplatz ungeeignet für eine Landung ist. Auf diese Weise kann mit den bereits verfügbaren technischen Hilfsmitteln eine rudimentäre Kommunikation aufgebaut werden.

Die bodengerichtete visuelle Kommunikation aus der Luft ist aufgrund der für einen sicheren Flug notwendigen Mindestflughöhen und -geschwindigkeiten oft nur sehr eingeschränkt möglich (insbesondere für Starrflügler). Daher kommen in der visuellen Luft-Boden-Kommunikation meist deutlich sichtbare Flugmanöver wie das „Flügelwackeln" zur positiven Rückmeldung beziehungsweise ein Vollkreis nach rechts als negative Rückmeldung zum Einsatz. Schwebefähige Plattformen können darüber hinaus gut sichtbare Nick-, Gier- und Rollbewegungen in kommunikativer Absicht einsetzen.

2.2.5 Gebärdensprachen

Weit vor dem Beginn der modernen Luftfahrt entwickelten sich in verschiedenen Kulturen und Regionen der Welt eigenständige und teils komplexe, nonverbale Kommunikationsformen, insbesondere bei Gehörlosen und Schwerhörigen. Weltweit existieren rund 200 dieser sogenannten Gebärdensprachen, von denen bisher aber nur etwa zehn Prozent erforscht sind [133]. Im Gegensatz zur Pantomime, bei der versucht wird, mit Hilfe des ganzen Körpers in Form eines Gebärden- und Mienenspiels jeweils eine Handlung eines einzelnen Akteurs auf künstlerische Weise in Realzeit darzustellen, werden bei Gebärdensprachen nur Teile des Oberkörpers als linguistische Symbole genutzt [106, S. 33-34], die nicht zwangsläufig an konkrete oder ikonische Inhalte gebunden sind. In diesem System aus Symbolen ist die Bedeutung und das Zusammenspiel jener Elemente kulturspezifisch, sodass auch, wie bei gesprochenen Sprachen, regionale Dialekte auftreten können. Gebärdensprachen stellen somit keine „gebrochene" Form einer gesprochenen Sprache dar, sondern haben eine eigene linguistische Struktur, welche von der Struktur gesprochener Sprachen unabhängig ist [106, S. 13-14]. Weit verbreitet ist die

[17] Laut EUROCONTROL [132, S. 35-36] entstanden im Zeitraum 2013-2017 die meisten PLOCs (rund 33%) in Zusammenhang mit Frequenzwechseln (falsch eingestellt, vergessen zu wechseln, keine Aufforderung zum Wechsel von ATC erhalten, etc.).

American Sign Language (ASL), die offiziell Anwendung vor allem in Nordamerika findet, aber auch in anderen Regionen der Welt wie Lateinamerika, Afrika und Südostasien praktiziert wird. Im deutschsprachigen Kulturraum wird in der Gehörlosengemeinschaft mithilfe der *Deutschen Gebärdensprache* (DGS) kommuniziert. Diese ist unter anderem in Deutschland, Belgien und Luxemburg rechtlich als eigenständige Sprache anerkannt[18] [134–136] und wird auch in der Linguistik aufgrund ihrer komplexen Struktur und Grammatik als vollwertige Sprache angesehen. In ihrer linguistischen Struktur weist die DGS viele grammatikalische Elemente einer gesprochenen Sprache auf. Aufgrund der visuellen Informationsverarbeitung unterscheidet sich diese Grammatik jedoch in einigen Bereichen, sodass in der Regel eine zweijährige Ausbildung für eine sichere Beherrschung gefordert wird, beispielsweise im Rahmen einer Dolmetschertätigkeit. Um den Umfang dieser Arbeit nicht zu überdehnen, werden nicht alle linguistischen Facetten von Gebärdensprachen vollumfänglich dargestellt. Stattdessen sollen sie – am Beispiel der DGS – zur systematischen Beschreibung der menschlichen Kommunikationsmittel (Artikulatoren) dienen. Der nachfolgende Abschnitt orientiert sich größtenteils an den Arbeiten von Papaspyrou et al. [99] und Boyes Braem [106].

2.2.6 Phonologie visueller Handzeichensysteme am Beispiel der Deutschen Gebärdensprache

Die Phonologie ist die Lehre von den kleinsten bedeutungsunterscheidenden Elementen einer Sprache, den Phonemen. Die Phoneme der DGS lassen sich in zwei Klassen unterteilen: den Handzeichen (manuelle Komponente) und allen anderen zur Kommunikation beitragenden nonverbalen Äußerungen (nonmanuelle Komponenten), wie z.B. der Haltung des Kopfes bzw. Oberkörpers, der Blickrichtung oder der Bewegung des Mundes [99, S. 9]. Abb. 20 zeigt die zugrundeliegende Systematik und die Hauptphoneme mit den entsprechenden Parametern. Zum besseren Verständnis sind in der Übersicht jedoch nur die für diese Arbeit relevanten Komponenten dargestellt, eine detaillierte Übersicht befindet sich im Anhang A1.

[18] Durch die Verankerung im Gesetz (§§ 6, 9 Behindertengleichstellungsgesetz) ergeben sich weitreichende Ansprüche für Gehörlose, wie bspw. das Recht auf eine schulische Ausbildung in Gebärdensprache oder das Recht auf einen Dolmetscher bei Amtsvorgängen und vor Gericht.

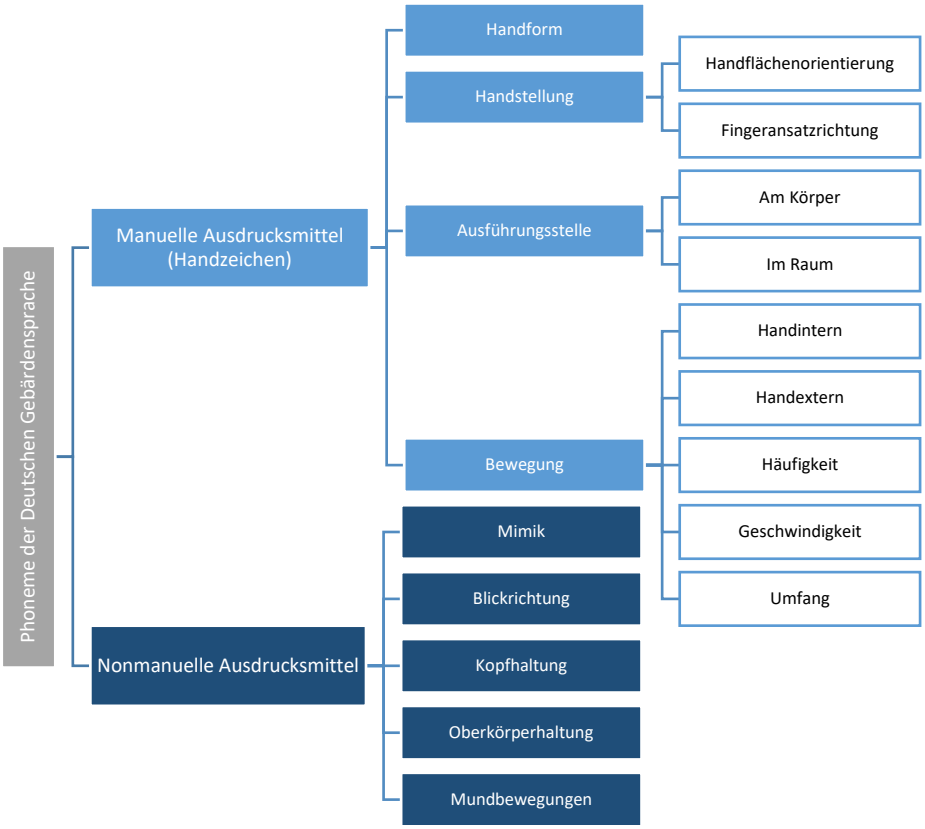


Abb. 20 Hauptphoneme der Deutschen Gebärdensprache nach Papaspyrou et al. [99]

Hauptkomponenten des manuellen Kanals

Der manuelle Kanal wird mittels der vier Hauptkomponenten *Ausführungsstelle*, *Handform*, *Handstellung*, *Bewegung* sowie weiteren Unterparametern beschrieben. Gebärden, welche mit unterschiedlichen Bedeutungen belegt sind, sich jedoch nur in einem Parameter unterscheiden, werden als Minimalpaar bezeichnet [99, S. 18]. Darüber hinaus werden Gebärden für gewöhnlich im sogenannten *Gebärdenraum* geformt [106, S. 23]. Dieser entspricht einem gedachten Halbkreis, mit dem Zentrum im Rumpf einer Person. Zur besseren Verortung einer Gebärde wird der Gebärdenraum in eine frontale, horizontale und vertikale Körperebene aufgeteilt [99, S. 33] und die Lage der Hand in jeder Ebene beschrieben (Abb. 21). Je näher eine Gebärde am Rumpf ausgeführt wird, desto „leiser“ ist sie. Umgekehrt führen großräumige, weiter vom Rumpf entfernte Gebärden zu „lauten“ Äußerungen, die auch von weiter entfernt stehenden Adressaten wahrgenommen werden können. Der Großteil der Bewegungen findet jedoch aus ergonomischen Gründen vor dem Körper in der neutralen Ausführungsstelle statt.

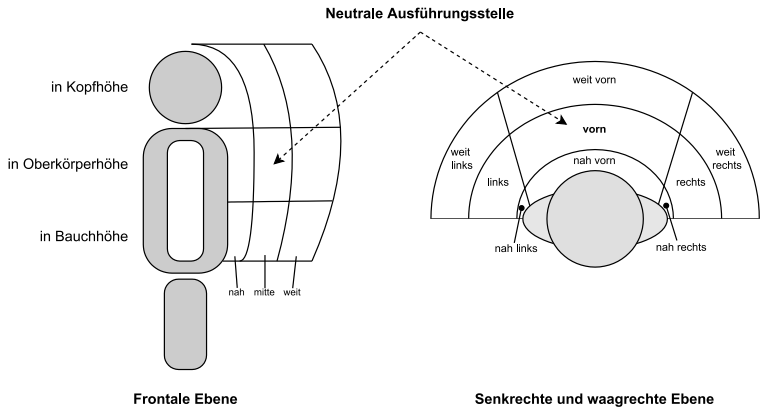


Abb. 21 Ausführungsstellen im Raum, eigene Darstellung basierend auf [99, S. 47]

Gebärden können nicht nur frei im Raum platziert (*Ausführungsstelle im Raum*), sondern auch in der Nähe von oder auf anderen Körperstellen ausgeführt werden (*Ausführungsstelle am Körper*). Die *Handform* stellt eine weitere Hauptkomponente des manuellen Kanals dar und beschreibt die Stellung einzelner Finger einer Hand zueinander. Abb. 22 zeigt die sechs Grundhandformen, die in allen bisher erforschten Gebärdensprachen vorkommen [137]. In der DGS wurden bislang 30 Handform-Phoneme gefunden, zusätzlich existiert auch ein Fingeralphabet zum Buchstabieren von Eigennamen oder anderen Elementen, welche in der DGS nicht als eigenständige Gebärden enthalten sind [99, S. 20].

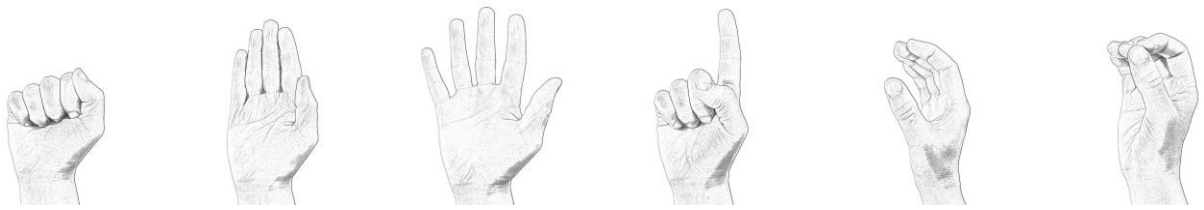


Abb. 22 Die sechs Grundhandformen (v.l.n.r.): Faust, Flachhand, Spreizhand, gestreckter Zeigefinger, C-Hand sowie O-Hand, eigene Darstellung basierend auf [106, S. 22]

Die Handformen der linken und rechten Hand können nicht beliebig miteinander kombiniert werden, sondern unterliegen sprachspezifischen Kombinationsregeln. Zwei Regeln finden jedoch in allen Gebärdensprachen Anwendung. So besagt beispielsweise die Regel der Dominanz, dass *„falls zwei Hände eine unterschiedliche Handform aufweisen, die dominierende ‚starke' Hand sich bewegt und die ‚passive' Hand stillsteht"* [106, S. 28]. Zusätzlich besagt die Regel der Symmetrie, dass beide Hände die gleiche Handform aufweisen müssen, sofern sie sich gleichzeitig bewegen [106, S. 27].

Die *Handstellung* ist ein weiteres Merkmal und wird aus der Perspektive des Sprechers beschrieben. Dazu wird als Parameter jeweils neben der Ausrichtung der Handinnenfläche (*Handflächenorientierung*) auch die Orientierung des Handrückens (*Fingeransatzrichtung*) bestimmt. Die dynamischen Merkmale einer Gebärde während der Produktion werden über die Dimension *Bewegung* beschrieben. Als Parameter werden sowohl die Bewegungen einzelner Finger (z.B. Reib-, Krümm- oder Knickbewegungen) als auch Bewegungen mit Handgelenksbeteiligung erfasst, beispielsweise wenn die Hand geschüttelt, gedreht oder gebeugt wird [99, S. 53]. Die Hand kann sich auch als Ganzes zwischen zwei oder mehreren Ausführungsorten auf unterschiedlichen Arten (geradlinig, bogenförmig, wellenartig, etc.) sowie mit unterschiedlichen Geschwindigkeiten und Anzahl an Wiederholungen bewegen [99, S. 51]. *Bewegung* wird als wichtigste Komponente bei der Analyse der Gebärdensprache angesehen [106, S. 26]. Eine kompakte Übersicht der entsprechenden manuellen Parameter mit einigen exemplarischen Ausprägungen zeigt Tab. 2.

Tab. 2 Hauptkomponenten, Parameter und Ausprägungen des manuellen Kanals nach [99]

Hauptkomponente	Parameter	Beispiele für Ausprägungen
Ausführungsstelle	Am Körper	Kopf, Gesicht, Rumpf, Arm, Hand
	Im Raum	Waagrecht/senkrecht: nah/weit/fern, links/vorn/rechts Frontal: nah/Mitte/weit, Kopfhöhe, Oberkörperhöhe, Bauchhöhe
Handform	-	Faust, Flachhand, O-Hand, C-Hand, gestreckter Zeigefinger, Spreizhand
Handstellung	Handflächenorientierung	Links, rechts, oben, unten, rechts oben, rechts unten, etc.
	Fingeransatzrichtung	Frontal: oben, unten, links, rechts, etc. Waagrecht: nach vorn, vorn rechts, links, hinten links, etc. Senkrecht: nach oben, vorn oben, vorn unten, hinten unten, etc.
Bewegung	Handintern	Reib-, Fingerspiel-, Krümm-, Knick-, Schüttel-, Dreh-, Beugebewegung, etc.
	Handextern	Geradlinig, Bogen, Welle, Zickzack, Kreis, Spirale
	Häufigkeit	Einmalig, doppelt, mehrfach
	Geschwindigkeit	Niedriges, normales, hohes Tempo
	Umfang	Verkleinert, normal, vergrößert

Nonmanuelle Komponenten

Informationen können aber auch über die Haltung und Ausrichtung des Oberkörpers oder des Kopfes als auch über die Blickrichtung übermittelt werden. Der Gesichtsausdruck (Mimik) kann mit Hilfe der sechs Basisemotionen (Angst, Ekel, Freude, Trauer, Überraschung, Wut) auch in adverbialer Funktion genutzt werden, um die eigene Einstellung zum Gesprächsinhalt zu transportieren oder um Imperative sowie Fragen zu formulieren [99, S. 71]. Der *Blickkontakt* ist dabei „das wichtigste Signal der Gesichtsmimik“ [100, S. 126].

Bilderzeugungstechniken für ikonische Gebärden in der DGS

Ein weiteres Klassifizierungssystem bezieht sich auf die Bildhaftigkeit von Gesten. Dabei wird zwischen *ikonischen*, *deiktischen*, *metaphorischen* und *pantomimischen* Darstellungen unterschieden. Die DGS bietet mehrere Erzeugungstechniken für bildhafte (ikonische) Gebärden [99, S. 82]. Die Hand kann stellvertretend für einen Gegenstand genutzt werden (*Hand als Substitutor*) oder eine andere Hand repräsentieren, die einen Gegenstand verwendet (*Hand als Manipulator*). Auch kann die Hand als Zeichenwerkzeug dienen (*Hand als Skizze*), um eine imaginäre Spur oder Silhouette im Raum zu erzeugen. Als Begrenzungsanzeiger können Hände dienen, um die Größe oder Ausdehnung von Objekten anzuzeigen (*Hand als Maß*). Am bekanntesten ist sicherlich die Hand in der Funktion eines Hinweispfeils (*Hand als Index*), der auf ein reales oder imaginäres Objekt hinweist. Zeigegesten können in der Funktion als Pronomen (ich, du), Lokativpronomen (hier, dort) als auch Demonstrativpronomen (dieses, jenes) verwendet werden [138, S. 395].

2.3 Zusammenfassung

Dieses Grundlagenkapitel hat aufgezeigt, dass zeitweise beabsichtigte oder unbeabsichtigte Unterbrechungen von funkbasierten Kommunikationsverbindungen in der Luft- und Raumfahrt auftreten können und geeignete Alternativverfahren zur Verständigung bei Funkausfällen entwickelt wurden. Die Wechselmöglichkeit auf eine visuelle Kommunikationsform (vor allem per Handzeichen) stellt insbesondere in der bemannten Luftfahrt eine elementare Komponente einer resilienten Kommunikation dar. Je nach Einsatzumgebung, Teilstreitkraft, Flugphase und Interaktionsabstand werden hierfür verschiedenste technische Handzeichensysteme mit teils mehrdeutigen Signalisierungsarten zur zwischenmenschlichen Verständigung genutzt, deren Bedeutung sich erst im jeweiligen Nutzungskontext ergibt. Die steigende Nachfrage nach unbemannten Flugsystemen und der Wunsch, diese sicher und nahtlos in bestehende menschenzentrierte Prozesse und Verfahren zu integrieren, verlangt nach einer visuellen Mensch-UAV-Kommunikationsschnittstelle, welche ähnliche Funktionen bereitstellen kann, um eine alternative Verständigung per Handzeichen beim Fehlen von funkbasierter Infrastruktur zu ermöglichen. Komplexere Handzeichensysteme (z.B. Gebärdensprachen) bieten einen systematischen Zugang zu den nonverbalen Kommunikationsfähigkeiten des Systems „Mensch“ und können bei der Abbildung der relevanten Körpersignale und Interaktionsprozesse in geeignete Computermodelle helfen, die für eine technische Umsetzung benötigt werden.

Nachdem in diesem Kapitel vorrangig die nonverbale Mensch-Mensch-Interaktion behandelt wurde, geht das nachfolgende Kapitel nun speziell auf den aktuellen Stand der Forschung in Bezug auf die nonverbale Interaktion von Menschen mit technischen, flugfähigen Systemen ein und zeigt auf, welche Besonderheiten bei dieser Form der Mensch-Maschine-Kommunikation beachtet werden müssen und in welchen Bereichen noch Forschungsbedarf besteht.

3 Stand der Forschung

Das gestiegene Interesse an der kommerziellen Nutzung von unbemannten Flugsystemen auch außerhalb des militärischen Kontextes (vgl. Kapitel 2.1.3) und die allgemeine Verfügbarkeit insbesondere von kleinen, kostengünstigen Multikoptersystemen haben das Anwendungsspektrum und den Nutzerkreis von UAS signifikant erweitert und einen neuen Teilzweig der Mensch-Maschine-Interaktion (*Human-Machine Interaction*, HMI) hervorgebracht, der in der Literatur unter dem Begriff der *Human-Drone Interaction* (HDI) zusammengefasst wird. Darunter fallen unterschiedliche Anwendungen, bei denen Menschen UAVs als Werkzeuge zur luftgestützten Inspektion, Vermessung oder Dokumentation beispielsweise von schwer zugänglichen Infrastrukturen wie Baustellen [139], Windrädern [140] oder Pipelines [141] nutzen oder sie als Transportvehikel für zeitkritische Medikamente [142] oder Blutproben [143] einsetzen. Bei derlei Anwendungen erfolgt der Informationsaustausch zwischen Systemnutzer und Fluggerät über eine klassische funkbasierte Schnittstelle, auf die mithilfe einer stationären oder mobilen Bodenkontrollstation zugegriffen werden kann. Insofern konzentriert sich die Forschung im Bereich der *konventionellen HDI* auf die Entwicklung neuer Verfahren für die fliegende Komponente (u.a. Energieversorgung [144], Kollisionsvermeidung [145], lokale Navigation [146]), die Funkstrecke [147] oder auf die Benutzerschnittstelle der Bodenkontrollstation (z.B. zur effizienten Missionsplanung und -durchführung [148]). Bei diesen funkbasierten UAV-Anwendungen spielt die Position und Entfernung des Nutzers in Relation zum UAV nur eine untergeordnete Rolle, solange ein sicherer Flugbetrieb gewährleistet werden kann und die Betriebsgrenzen der Funkstrecke oder andere einschränkende Faktoren wie rechtliche Vorgaben (z.B. Betrieb nur innerhalb der Sichtweite des UAS-Betreibers) beachtet werden.

Die fortschreitende Miniaturisierung und Automatisierung der Fluggeräte ermöglichte die Entwicklung neuartiger Anwendungen, die zunehmend auch den nutzbringenden Einsatz von UAVs in der direkten Nähe von Menschen erlaubt. So werden kleine UAVs mittlerweile auch als persönliche Begleiter bei Indoor-/Outdoor Aktivitäten genutzt [149] oder mit Schwimmhilfen [6] und Defibrillatoren [150] ausgestattet, um vor Ort helfen zu können. Aus der räumlichen Annäherung an den Nutzer erwächst aber auch die Notwendigkeit für eine lokale und direkte, wechselseitige Kommunikationsmöglichkeit. Nicht immer können diese Interaktionen über konventionelle HDI-Verfahren abgewickelt werden, weil nicht davon ausgegangen werden kann, dass beim Nutzer ein zum UAV kompatibles Kommunikationsgerät vorliegt. Hierzu gehört beispielsweise die luftgestützte Suche nach Vermissten [151] oder die automatisierte Evakuierung von Personengruppen aus Gefahrengebieten mittels UAVs [152, 153]. Auch existieren Anwendungsszenarien, bei denen das Mitführen eines weiteren technischen Gerätes nicht praktikabel ist und/oder auf die Abstrahlung elektromagnetischer Signaturen zum Selbstschutz verzichtet werden muss. Beispiele hierfür finden sich unter anderem in der kleinräumigen, kollaborativen Zusammenarbeit von Mensch und UAV bei der Instandsetzung von Hochspannungsleitungen [154] oder dem Einsatz von UAVs zur Nahbereichsaufklärung.

In den letzten Jahren hat sich daher innerhalb der HDI ein weiterer Forschungszweig herausgebildet, der eine möglichst intuitive Interaktion mit fliegenden Systemen auch ohne den signaltechnischen Umweg über ein zusätzliches Kommunikationsgerät wie eine BKS anstrebt und der unter anderem[19] als *collocated Human-Drone Interaction* (cHDI) bezeichnet wird. Die nachfolgenden Abschnitte enthalten einen Überblick über den gegenwärtigen Stand der Forschung in der cHDI mit einem Fokus auf

[19] In der Literatur finden sich auch andere Bezeichnungen für dieses Forschungsgebiet wie *situated HRI* [155] oder *natural HDI* [156].

dem Informationsaustausch über den optischen Kanal als auch den dabei eingesetzten Verfahren zur bildgestützten Erfassung von Körpermerkmalen und deren Klassifizierung. Darüber hinaus werden Möglichkeiten aufgezeigt, wie Systemantworten über einen visuellen Rückkanal an den Nutzer übermittelt werden können und welche Motive bei der Zusammenstellung eines gestischen Vokabulars verfolgt werden. Abschließend werden gegenwärtige Forschungslücken in der cHDI identifiziert und geeignete Forschungsfragen formuliert, die zur Schließung dieser beitragen können.

3.1 Collocated Human-Drone Interaction

Die direkte Interaktion mit unbemannten Flugsystemen hat viele Parallelen zur Mensch-Roboter-Interaktion (*Human-Robot Interaction*, HRI) [157]. Zu den bestehenden Herausforderungen bei der Interaktion mit bodengebundenen Systemen (u.a. Kollisionsvermeidung, Erkennung von Nutzereingaben, Identifikation von Absichten), kommen bei der Interaktion mit fliegenden Systemen weitere spezifische Faktoren hinzu:

- Die dreidimensionale Bewegungsfähigkeit der Flugplattform erlaubt einen größeren Bewegungsspielraum bei der Vermeidung von Hindernissen und zur Auflösung von Verdeckungen, führt jedoch auch zu einem komplexen Bewegungsverhalten, welches bei der Interaktion berücksichtigt werden muss.
- Aufgrund von Bauraum-, Gewichts- und Energielimitierungen ist die Auswahl an Nutzlasten (z. B. Sensoren) stärker eingeschränkt.
- UAVs werden vorrangig im Freien verwendet, wodurch sich variable Interaktionsdistanzen in hochgradig dynamischen Umgebungen ergeben.
- Ihre Eigengeschwindigkeit sowie rotierende Bauteile können eine Gefahr für die körperliche Unversehrtheit des Nutzers darstellen.
- Multisensorische Erfahrungen (hohe Schallpegel durch Antriebselemente, umherfliegende Partikel und kleine Objekte aufgrund von starken Abwinden) beeinflussen zusätzlich die Interaktion.

Abhängig von der Anwendung und des Automatisierungsgrades des Flugsystems, kann der Mensch bei der Interaktion mit einem UAV unterschiedliche Rollen einnehmen [158, S. 167440], bei denen er unter anderem auftreten kann als:

- Aktiver Steuerer,
- Empfänger einer UAV-Dienstleistung,
- Sozialer Companion sowie als
- Supervisor.

Als *aktiver (Fern-) Steuerer* behält er die volle Kontrolle über das Fluggerät bei und kann es als Werkzeug zur Erreichung seiner Ziele nutzen, beispielsweise zum Transport von Lasten oder zur Inspektion von Infrastrukturen. Daneben kann der Mensch aber auch als *Empfänger einer UAV-Dienstleistung* fungieren, beispielsweise dann, wenn er aufbereitete Aufklärungsergebnisse erhält oder durch das Abwerfen eines Rettungsrings vor dem Ertrinken gerettet wird [159]. Verfügt das UAV zudem über erweiterte (soziale) Fähigkeiten, kann der Mensch von der Gesellschaft des fliegenden *„Social Companion“* profitieren [160]. Hochautomatisierte Systeme benötigen im Allgemeinen nur minimale menschliche Eingaben. Für die Einsatzplanung des fliegenden Systems, aber auch bei der Annäherung an Systemgrenzen kann der Mensch in der Funktion als *Supervisor* auftreten. Aufgrund der hohen

Mobilität von UAVs kann ein Mensch aber auch ungewollt in die Interaktion mit einem solchen System geraten. Cauchard et al. bezeichnen diejenigen Menschen, die durch die Anwesenheit eines UAVs beeinträchtigt werden, als *„affected bystanders“* [2, S. 78].

In der Literatur werden unterschiedliche Aspekte der cHDI untersucht, welche basierend auf der Systematik von Tezza und Andujar [158, S. 167439] in vier Hauptforschungsfelder gruppiert werden können. Unter *neuartigen Anwendungen* werden jene Arbeiten gebündelt, welche den aktuellen Nutzungsraum von UAVs erweitern und möglicherweise zukünftige Entwicklungsrichtungen innerhalb der cHDI aufzeigen. Mit *Interaktionsmodalitäten* werden all jene Arbeiten zusammengefasst, die sich mit der Frage beschäftigen, welche Kanäle für die direkte Interaktion im Nahbereich genutzt werden können, während andere der Frage nachgehen, welche *Kommunikationsinhalte* dabei zwischen Mensch und UAV ausgetauscht werden können. Unter dem Begriff der *Proxemik* wird subsumiert, wie Mensch und UAV den Raum während einer Interaktion nutzen.

Aufgrund der Schwebefähigkeit, des geringen Integrationsaufwandes und des allgemein einfacheren Betriebs werden in der cHDI vorwiegend Multikoptersysteme als Versuchsplattformen für Mensch-Maschine-Untersuchungen genutzt. Kollokierte Anwendungen sind natürlich auch mit anderen UAV-Konfigurationen wie unbemannten Luftschiffen, Helikoptern, Hybridsystemen oder Starrflüglern (z.B. bei Rollmanövern am Boden) möglich. Angesichts der vielfältigen Anwendungsmöglichkeiten dieser UAV-Variante beziehen sich die nachfolgenden Kapitel primär auf den Einsatz von Multikoptersystemen in der cHDI.

3.1.1 Neuartige Anwendungen

Während digitale Sprachassistenten heute beim Finden und Verwalten von Informationen unterstützen, sehen Theiß et al. in unbemannten Kleinfluggeräten die zukünftigen *„Personal Spatial Search Assistants“* [161], d.h. persönliche Assistenten, die bei raumbezogenen Problemstellungen helfen sollen. In einem Laborversuch konnten sie zeigen, dass Probanden mit der Hilfe eines kleinen, assistierenden UAVs Suchaufgaben schneller lösen konnten als ohne dessen Hilfe. Neben der Mobilität kann auch das erzeugte Propellergeräusch und der Luftstrom in interaktiver Weise genutzt werden, um Menschen mit Seheinschränkungen bei der Orientierung als auch der Lokalisation von Objekten und Wegen zu helfen [162]. UAVs als „greifbare Benutzerschnittstellen“ (*Tangible User Interfaces*) eröffnen zudem neuartige, taktile Interaktionskonzepte, bei denen die physische Interaktion mit dem Fluggerät im Vordergrund steht [163] und zur Verbesserung der Wahrnehmung des eigenen Körpers beitragen kann [164].

In der cHDI besteht ein Interesse daran, UAVs vermehrt in den Lebensalltag einzubinden. Baytaş et al. prägten den Begriff der „*social drones*“ [165, S. 2] und beziehen sich damit auf eine Gattung von „autonomen, frei fliegenden, verkörperten Agenten, die in Umgebungen agieren, welche von Menschen bevölkert werden und welche entweder direkt mit dem Agenten interagieren oder lediglich als unbeteiligte Dritte anwesend sind“ (frei übersetzt nach [166]). Das kann beispielsweise ein fliegender „Lauf-Buddy“ sein, der seinen Nutzer beim Sport begleitet [20] und diesen mit Jubelgeräuschen zu Höchstleistungen motiviert [160]. Eine elementare Grundvoraussetzung für die erfolgreiche Integration in das öffentliche Leben ist jedoch eine soziale Akzeptanz der Fluggeräte, welche unter anderem durch eine weiche, anthropomorphe Formensprache, multimodale Interaktionsmodi als auch eine Sensibilität für Distanz und Nähe erreicht werden kann [167].

Wie im Bereich der Interaktion mit bodengebundenen technischen Systemen, geht auch in der cHDI der Trend langfristig zu Companion-Systemen [168], welche ihrem Nutzer nicht nur bei bestimmten Tätigkeiten assistieren, sondern sich auch nach dessen Fähigkeiten, Bedürfnissen, Vorlieben und

Wünschen richten und sich in der Rolle eines kompetenten und vertrauenswürdigen Partners präsentieren [169, S. 335]. Bevor jedoch die Vision eines *„Companion Flying Robot“* [170] in die Realität umgesetzt werden kann, müssen zunächst die notwendigen kognitiven und perzeptiven Fähigkeiten ihren Weg auf die Fluggeräte finden. Daher werden unter anderem auch alternative und nutzerzentrierte Schnittstellen für die direkte Mensch-UAV-Interaktion von der Forschergemeinde gefordert [2, S. 79].

3.1.2 Interaktionsmodalitäten

Das technische Subsystem, worüber Nutzer mit einem interaktiven System interagieren können, wird allgemein als Benutzungs- oder Benutzerschnittstelle (*User Interface,* UI) bezeichnet[20]. Bei natürlichen Benutzerschnittstellen (*Natural User Interface*, NUI) benötigt der Nutzer keine zusätzlichen Geräte, um mit einem technischen System in Interaktion zu treten, sondern kann stattdessen den eigenen Körper (z.B. Berührungen, Stimme, Körperbewegungen) direkt für Systemeingaben nutzen. Kommerziell erhältliche UAVs verfügen über ausreichend (Rechen-) Leistung, um zusätzliche Eingabegeräte wie aktive und passive Sensoren als auch Ausgabegeräte in Form von kleinen Displays oder Lautsprechern mitführen zu können. Mit der Fähigkeit, diese frei im Raum positionieren zu können, hat nach Funk das Zeitalter der „fliegenden Benutzerschnittstellen“ [172] begonnen, wodurch auch Ungeübten eine intuitive Bedienung dieser Systeme ermöglicht werden kann. Die dazu notwendigen Nutzereingaben können gemäß der Taxonomie von Jones et al. [173] hierbei in „Low-Level“-Anweisungen, z.B. Richtungsanweisungen zur Veränderung der UAV-Position im Raum, und „High-Level“-Anweisungen (z.B. Aktivierung intelligenter Funktionen wie einem Folgemodus) unterteilt werden. Für die Übermittlung beider Eingabearten existieren sowohl unimodale als auch multimodale Ansätze.

Unimodale Ansätze

Spracheingabesysteme sind eine Form eines NUI und sind in der HRI und HMI etabliert, sodass der Gedanke naheliegend ist, diese Schnittstelle auch auf die cHDI zu übertragen. In der Praxis zeigt sich jedoch, dass dies nicht ohne weiteres umgesetzt werden kann, wenn angenommen werden muss, dass dem Nutzer keine weiteren Kommunikationsgeräte zur Verfügung stehen (z.B. in Form eines Mobiltelefons). Für die Untersuchung der gerätefreien, sprachbasierten Mensch-UAV-Interaktion statteten Jokisch et al. [174] nicht wie gewöhnlich den Nutzer, sondern einen kleinen Quadkopter mit einem Mikrofon aus und versuchten, die Spracheingaben des in der Nähe befindlichen Nutzers während des Schwebeflugs zu erkennen. Dabei zeigte sich, dass die Erkennung dieser Sprachbefehle von einem fliegenden System aus wesentlich schwieriger umzusetzen ist als von einem bodengebundenen System. Hauptursächlich dafür sind die kontinuierlichen Fluglagekorrekturen des verwendeten Multirotorsystems, die zu schnellen Motordrehzahländerungen führen und damit eine selektive oder adaptive Filterung des Audiosignals ohne Wissen über die akustischen Abstrahlcharakteristika des Fluggerätes und das aktuelle Flugmanöver erschweren. Aktuelle Multirotorsysteme erzeugen in einem Meter Entfernung einen Lärmpegel von bis zu 80 dB. Für eine zuverlässige Erkennung von Sprachbefehlen von derartigen Plattformen müssen Spracherkennungssysteme diese Schwelle meistern [175, S. 24].

Neben den Herausforderungen der Signalaufbereitung, sind akustische Eingabeschnittstellen für fliegende UAVs mit aerodynamischer Auftriebserzeugung auch aus ergonomischen Gesichtspunkten nachteilig. Je näher das UAV beim Nutzer schwebt, desto stärker tritt der sogenannte *Lombard-Effekt* [176] auf, der den Nutzer aufgrund des Propellerlärms und der Abwinde zwingt, lauter zu sprechen und die Tonhöhe anzuheben. Dies könnte bei einer länger andauernden Interaktion die Stimmbänder

[20] Nach DIN EN ISO 9241-110 umfasst eine Benutzungsschnittstelle "alle Bestandteile eines interaktiven Systems (Software oder Hardware), die Informationen und Steuerelemente zur Verfügung stellen, die für den Benutzer notwendig sind, um eine bestimmte Arbeitsaufgabe mit dem interaktiven System zu erledigen." [171, S. 7].

belasten und das Interaktionserlebnis negativ beeinträchtigen. Neben dem Schallpegel trägt auch die Frequenzverteilung maßgeblich zum Lärmempfinden bei. Das akustische Emissionsspektrum elektrisch betriebener Multirotor UAVs enthält viele hochfrequente Anteile, welche eine hohe „psychoakustische Belästigung“ (*psychoacoustic annoyance*) darstellen [177, 178] und damit einem Anstieg des Stress-Niveaus bei Umstehenden hervorrufen können [167, S. 86]. Elektrisch betriebene Propeller sind aktuell das primäre Hauptauftriebserzeugungsverfahren von schwebefähigen UAVs, sodass zum gegenwärtigen Zeitpunkt die Empfehlung lautet: „Don’t talk to noisy drones“ [174].

Weitaus gängiger in der Literatur sind hingegen visuelle Schnittstellen, die aktive oder passive bildgebende Sensoren nutzen, um Nutzereingaben zu erfassen. Im Gegensatz zu Mikrofonen, können diese Sensoren sehr effektiv von schädlichen Störeinflüssen der Auftriebserzeugung (Bewegungsunschärfe oder Wobble-Effekte durch Resonanzschwingungen) durch vibrationsreduzierende Dämpfer und eine motorisierte, kardanische Aufhängung (*Gimbal*) entkoppelt werden. Darüber hinaus lassen sie sich im Vergleich zu akustischen Verfahren auch an variierende Umgebungsbedingungen, wie wechselnde Beleuchtungssituationen [175, S. 26], einfacher anpassen. Bei visuellen Kommunikationsverfahren werden üblicherweise Körpersignale wie die Blickrichtung [179], Zeigegesten [180] oder Hand- und Armgesten [181] in Nutzereingaben kodiert. Die technische Umsetzung der Funktionalität reicht dabei von einer reinen Simulation [182] über eine Verarbeitung auf der Bodenkontrollstation (Offboard-Prozessierung) [183] bis hin zur Umsetzung auf dem Fluggerät selbst (Onboard-Prozessierung) [184].

Taktile Schnittstellen am UAV stellen eine weitere Eingabemodalität dar. Tsykunov et al. entwickelten zur Fernsteuerung eines UAV-Schwarms einen fliegenden Joystick namens „DroneStick“ [185], der aus einem kleinen Quadkopter besteht, an dem ein elastisches Wendelkabel mit einem kleinen, vibrationsfähigen Handgriff befestigt ist. Über diesen kann der Nutzer den vor ihm schwebenden Quadkopter frei im Raum verschieben, welcher zugleich seine Fluglageänderung über eine Funkstrecke an den UAV-Schwarm übermittelt. Diese Methode bietet den Vorteil, dass Nutzer kein persönliches Steuergerät benötigen, weil die Benutzerschnittstelle selbst zum Nutzer fliegt und eine taktile Rückmeldung erlaubt. Aufgrund der notwendigen Nähe zu den Propellern, kommen bei dieser Art der Steuerung jedoch nur kleine, für den Nutzer ungefährlichen UAVs in Frage. Daneben erlauben auch fliegende Touchdisplays [163] eine direkte, taktile Interaktion mit einem UAV.

Multimodale Ansätze

Multimodale Ansätze kombinieren mehrere Eingabeverfahren (typischerweise visuelle mit akustischen), um dem Nutzer eine natürliche Interaktion zu ermöglichen. Die verwendeten Modalitäten verfügen für gewöhnlich jedoch nicht über denselben Funktionsumfang, sodass sie nicht alternativ verwendet werden können. Nonverbale Äußerungen werden meist auf deiktische Funktionen beschränkt und können theoretisch auch verbal kompensiert werden („Fliege zu dem Haus links vom Baum.“), wenn in der jeweiligen Situation Referenzpunkte enthalten sind. Der umgekehrte Fall, d.h. die vollständige gestische Formulierung eines verbalen Befehls, wird dagegen selten umgesetzt [186, 187]. Multimodale Ansätze werden üblicherweise in explorativen Studien evaluiert [188]. Abhängig vom untersuchten Nutzerkreis schwanken auch die Präferenzen bei der Wahl einer bevorzugten Modalität. Das Spektrum reicht dabei von einer klaren Bevorzugungen einer bestimmten Modalität (beispielsweise des visuellen Kanals in [156]) bis hin zu einer gleichverteilten Präferenz über alle zur Verfügung gestellten Modalitäten hinweg (vgl. [189]).

3.1.3 Kommunikationsinhalte

Kommunikation bezieht sich im Allgemeinen auf die Übermittlung beziehungsweise den Austausch von Informationen in Form von Mitteilungen, denen eine Absicht zugrunde liegt und somit ein bestimmtes Kommunikationsziel angestrebt wird [190, 191, S. 64]. In der cHDI finden sich zumeist zwei Kategorien solcher Mitteilungen (Low- und High-Level Instruktionen), mit denen unterschiedliche Ziele verfolgt werden. Low-Level-Eingaben gehören zu den am meisten untersuchten Instruktionen und beziehen sich üblicherweise auf die relative Positionierung der Flugplattform im Raum durch manuelles Verschieben [185] oder einfache, singuläre akustische [192] oder visuelle (Richtungs-) Instruktionen [193, 194]. Auch die Übermittlung absoluter Positionen durch die Nennung des entsprechenden Ortes [173] oder durch den Gebrauch von Zeigegesten [180] finden sich in der Literatur. Einige Arbeiten gehen auch der Frage nach, wie Objektbeschreibungs- oder Zielauswahleingaben auf natürliche Art und Weise an ein fliegendes UAV übermittelt werden können. Medeiros et al. [195] entwickelten eine Methode zur Zieldesignation mittels dreidimensionaler Zeigegesten zur Fensterauswahl bei UAV-unterstützten Feuerwehreinsätzen, während Monajjemi et al. [196] ein Verfahren vorgestellt haben, um einzelne UAV nonverbal zu einer Gruppe zusammenzufassen bzw. sie aus dieser wieder zu entfernen. Eine eindeutige Definition, welche Eingabe noch als Low-Level oder schon als High-Level betrachtet werden muss, existiert bislang nicht in der cHDI. Patrona et al. [197, S. 2] sehen beispielsweise Low-Level Befehle im Bereich der Flugsteuerung (*Control*) verortet, während High-Level-Befehle ihrer Ansicht nach die Grundlage für die Kommunikation mit einem fliegenden System (*Handling*) darstellen.

High-Level-Eingaben werden meist mit dem Zugriff auf höherwertige UAV-Funktionalitäten in Verbindung gebracht, um beispielsweise bestimmte Automatisierungsfunktionen (Starten, Landen, Aufzeichnung [198]) zu aktivieren oder Subsysteme zu schalten, wie Antrieb oder Bremsen [199, 200]. Dieser Zugriff wird in der großen Mehrzahl der Arbeiten durch singuläre Eingaben nachgebildet. Die darüber hinaus gehende Verkettung mehrerer Instruktionen zur Beschreibung von höherwertigen Kommunikationszielen, beispielsweise zur Übermittlung von Auftragsdaten in einem dialogähnlichen Vorgang, wurde bislang nur in wenigen Arbeiten aus dem Bereich der cHDI thematisiert [201, 202]. Abb. 23 zeigt beispielhaft ein solches nonverbales Vokabular, welches in Verbindung mit einer festgelegten Syntax auf Basis einer formalen Sprache genutzt wurde, um einfache Auftragsbeschreibungen gestisch an ein UAV zu übermitteln. Für eine möglichst intuitive Kommunikation werden jedoch Verfahren benötigt, die sich an natürlichen Sprachen orientieren und damit auch einen gewissen Spielraum bei der syntaktischen Varianz erlauben [173, S. 1744].

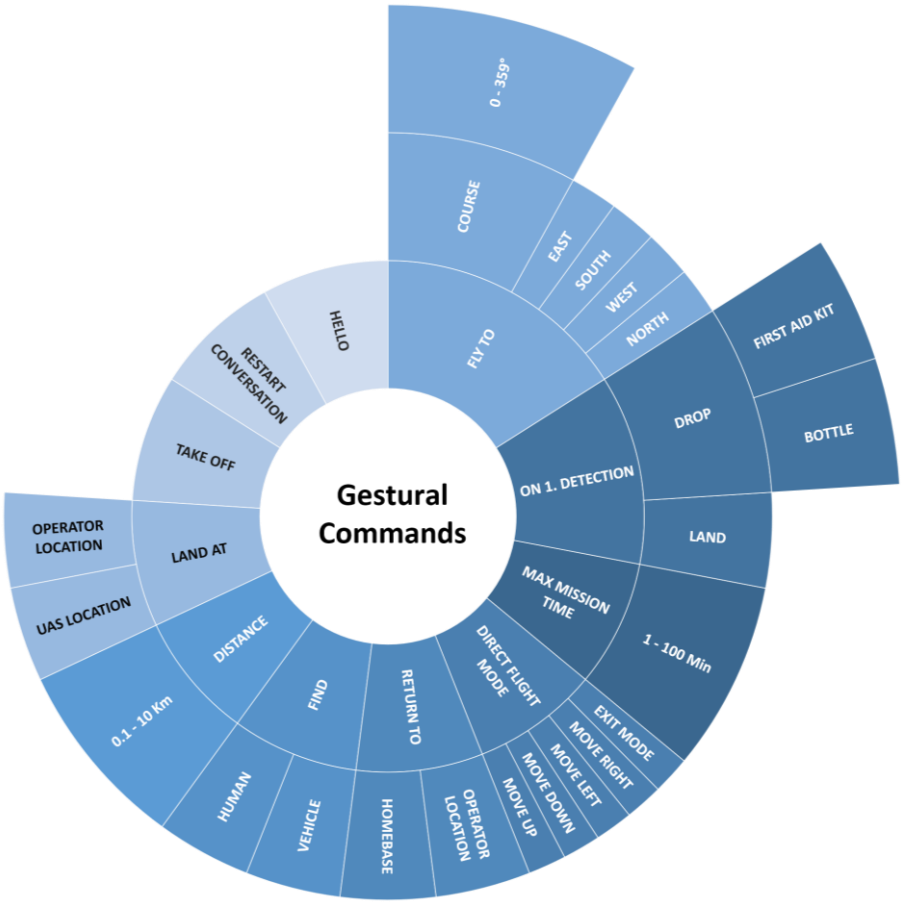


Abb. 23 Gestisches Vokabular zur Übermittlung von Aufträgen an ein UAV mittels einer formalen Sprache, zu lesen von innen nach außen, eigene Darstellung übernommen aus [202, S. 324]

Neben den Vorgängen während einer kollokierten Kommunikation, bestehen noch weitere ungeklärte Fragenstellungen, die sich insbesondere durch die Nutzung des visuellen Kanals ergeben. So sehen Cauchard et al. noch ungelöste Probleme beim Mehrfachzugriff auf ein Fluggerät [2, S. 79] bei Verwendung einer visuellen Schnittstelle. Wie soll ein UAV mit mehreren Personen oder Menschenmengen im Sichtfeld umgehen? Wie und wann sollen Menschen und UAV während der Konversation kommunizieren? Neben einer Möglichkeit zur korrekten Erkennung eines potentiellen Interaktionspartners, besteht auch der Bedarf für visuelle Authentifizierungsverfahren, die sicherstellen, dass nur Nutzer mit entsprechenden Zugriffsrechten mit dem UAV interagieren können.

3.1.4 Proxemik

Der Anthropologe Edward T. Hall prägte den Begriff der Proxemik und erkannte, dass Menschen während der Interaktionen mit anderen Individuen unbewusst bestimmte räumliche Abstände[21] bevorzugen und unterteilte diese in eine intime, persönliche, soziale sowie öffentliche Distanzzone [204]. In der Mensch-Roboter-Interaktion wurden ähnliche Effekte im Raumnutzungsverhalten beobachtet und geeignete Annäherungsverfahren vorgestellt, die sich jedoch nicht ohne Weiteres auf die Mensch-UAV-Interaktion übertragen lassen, da mehr Parameter zur Verfügung stehen [205]. Im Gegensatz zu bodengebundenen Robotern, können sich UAVs Menschen nicht nur aus allen Richtungen, sondern auch aus unterschiedlichen Höhen nähern, um eine Interaktion aufzubauen.

[21] Beispiel für Gesellschaften aus dem Nordamerikanischen Raum: intim (< 0,46 m), persönlich (0,46 – 1,22 m), sozial (1,22 – 2,10 m), öffentlich (> 2,10 m) [203, S. 579].

Wojciechowska et al. untersuchten in diesem Zuge verschiedene Annäherungsstrategien mit unterschiedlichen Fluggeschwindigkeiten, Flughöhen, Anflugrichtungen sowie Trajektorien [18]. Die befragten Probanden bevorzugten die Annäherung eines kleinen Quadkopters mit moderater Geschwindigkeit (0,5 m/s) von vorne bis zur sozialen Distanzzone und bei gleichbleibender Flughöhe auf Augenhöhe. Weitere Faktoren, die Einfluss auf das Stressempfinden und damit auf die minimale tolerierbare Interaktionsdistanz haben, sind die Größe sowie der vom Fluggerät erzeugte Abwind und Schallpegel. Daneben spielt es auch eine Rolle, ob die Propeller während der Interaktion in Betrieb sind oder weitere Sicherheitsmaßnahmen getroffen wurden (bspw. Propellerschutz [206] oder eine Glaswand [207]), ob der Nutzer dabei sitzt oder steht [208], sowie in welcher Umgebung die Interaktion stattfindet, d.h. innerhalb eines Gebäudes [209], im innerstädtischen Straßenverkehr [210] oder auf einer offene Fläche [20]. Ein anthropomorphes Design sowie eine verbale Begrüßung können darüber hinaus eine geringere Annäherungsdistanz begünstigen [167]. Mit zunehmender Miniaturisierung wird auch eine weitere Annäherung des UAVs bis hin zum Körperkontakt toleriert. Auda et al. identifizierten geeignete Verfahren und Orte für die Landung von Micro-UAVs auf dem menschlichen Körper [211].

In der Literatur finden sich Hinweise darauf, dass Untersuchungen zur Proxemik in VR- und Laborumgebungen mit stehenden Propellern den Realismus der Interaktion nur in einem sehr begrenzten Maße nachbilden können [205]. Für ein möglichst realistisches Interaktionsszenario werden daher cHDI-Studien mit realen, flugfähigen UAV empfohlen, welche über intelligente Funktionen (z.B. Personenverfolgung, Hindernisvermeidung) verfügen [212, S. 9].

3.1.5 Evaluierungsmethoden

Im Bereich der cHDI finden unterschiedliche Evaluierungsmethoden Anwendung, die sich im Realismus der Interaktion, der Komplexität des Experimentalsystems, dem Gefahrenpotential für die Probanden als auch der Reproduzierbarkeit und Skalierbarkeit unterscheiden [18, S. 174]. In sehr frühen Entwicklungsphasen werden bevorzugt Online-Umfragen genutzt oder Experteninterviews in Design-Workshops geführt, um Inspirationen zu sammeln und erste Ideen zu testen. Den Probanden werden hierbei Abbildungen oder Videos von statischen oder sich bewegenden UAVs und deren möglichen Verhalten präsentiert, welche sie mittels Fragebögen bewerten. Die Immersion ist hierbei gering, sodass das Interaktionserlebnis von der Vorstellungskraft der Probanden abhängt und die Übertragung der Erkenntnisse auf die Realität sehr eingeschränkt möglich ist.

Um den Faktor Vorstellungskraft zu eliminieren, kann die Interaktion in eine virtuelle Umgebung verlegt werden. Bei dieser Art von Studien wird dem Probanden über eine Kuppelprojektion oder VR-Brille ein virtuelles UAV präsentiert, welches über ein von Algorithmen oder Menschen erzeugtes Verhalten verfügen kann. Diese Methode hat den Vorteil, dass Interaktionssituationen leicht reproduziert werden können, wodurch die Aussagekraft der Messdaten gesteigert wird. Auch können gefahrlos Situationen nachgestellt werden, die für den Probanden unter realen Bedingungen potentiell gefährlich sein könnten. Nachteilig ist, dass virtuelle Umgebungen primär den visuellen und auditiven Kanal des Probanden bedienen, sodass nicht alle multisensorischen Erfahrungen während der Interaktion nachgebildet werden können.

Eine realitätsnähere Möglichkeit, die Interaktion mit technischen Systemen zu evaluieren, bilden sogenannte *„Wizard of Oz“*-Experimente (WoZ) mit realen Fluggeräten. In Anlehnung an das Kinderbuch „Der Zauberer von Oz“ von Lyman Frank Baum wird bei dieser Art von Experiment dem Probanden die Interaktion mit einem intelligenten System suggeriert. In Wirklichkeit handelt es sich dabei um einen Prototyp in einem frühen Entwicklungsstadium, bei dem die Funktionalität von einem

im Hintergrund agierenden Menschen (dem Zauberer) simuliert wird. Diese Methode wird vorrangig zur Evaluierung von neuartigen und noch nicht implementierten Interaktionskonzepten genutzt.

Den maximalen Realismus erfahren Probanden bei der Interaktion mit realen Fluggeräten, deren Verhalten ausschließlich von Algorithmen gesteuert wird und ein Sicherheitspilot lediglich bei Gefahrensituationen in das Geschehen eingreift. Wenn der Einfluss der Umgebung miteinbezogen werden soll, können Experimente dieser Art auch im Freien durchgeführt werden. Hierbei sind jedoch die Versuchsbedingungen nicht immer reproduzierbar (Wind- und Wettereinflüsse, unterschiedliche Beleuchtungssituationen). Diese Evaluierungsmethode erfordert zudem den meisten Integrations- und Organisationsaufwand, liefert aber ein authentisches, realistisches Bild der Interaktion.

3.1.6 Zusammenfassung und Literaturübersicht

Die lokale und natürliche Interaktion wird auch in der Mensch-UAV-Interaktion zunehmend wichtiger, um kollaborative Anwendungen zu ermöglichen, bei denen sich beide Interaktionsteilnehmer in räumlicher Nähe befinden. Dazu wurden unterschiedliche Modalitäten für alternative Schnittstellen vorgeschlagen, bei denen auf eine funkbasierte Kommunikation verzichtet werden kann. Der Einsatz von akustischen Schnittstellen wird aufgrund der Geräuschkulisse der in der cHDI meistgenutzten Multikoptersysteme gegenwärtig nicht empfohlen. Visuelle Schnittstellen bieten hier wesentlich mehr Vorteile und Möglichkeiten im Vergleich zu anderen Eingabemodalitäten. Jedoch beschränken sich die Kommunikationsinhalte dieser Schnittstelle oftmals nur auf die Übertragung richtungsbezogener oder hinweisender Informationen, obwohl diese Interaktionsform wesentlich mehr Potential böte (vgl. nonverbale interpersonelle Kommunikation in Kapitel 2.2). Darüber hinaus ist auch das Raumnutzungsverhalten von Mensch und insbesondere schweren und hochautomatisierten UAVs während solcher nonverbalen Interaktionsvorgänge noch wenig erforscht, ebenso wie der nonverbale Kommunikationsvorgang an sich. Für eine realistische Untersuchung werden daher vollintegrierte Experimentalsysteme gefordert. Tab. 3 zeigt die untersuchten Arbeiten im Bereich der cHDI als Übersicht.

Tab. 3 Literaturübersicht der betrachteten Arbeiten in der cHDI

Referenz	Untersuchungs-inhalt					Umgebung	Evaluierungs-methode	Prozessierung	Instruktionen	UAV-MTOM
	Steuerung	Proxemik	Kommunikation	Neuartige Anwendung	Optimierung	Online, Labor, VR, Feld	Explorativ, Prototyp	WoZ, offboard, onboard	Low-Level (L), High-Level (H)	Sehr leicht, Leicht, Mittel, Schwer[22]
2010										
Jones et al. [173]	●	-	●	-	-	VR	Explorativ	WoZ	L+H	-
2011										
Ng, Sharlin [17]	●	-	●	-	-	Labor	Explorativ	WoZ	L+H	Leicht
2012										
Song et al. [200]	-	-	●	-	-	Labor	Prototyp	Offboard	H	-
2013										
Pfeil et al. [213]	●	-	-	-	-	Labor	Prototyp	Offboard	L	Leicht
2014										
Nagi et al. [214]	●	-	-	-	-	Labor	Prototyp	Offboard	L	Leicht
2015										
Cauchard et al. [156]	●	-	-	-	-	Feld	Explorativ	WoZ	L+H	Mittel
Obaid et al. [215]	-	-	●	●	-	Online	Explorativ	-	-	-
Monajjemi et al. [216]	-	-	●	-	-	Feld	Prototyp	Onboard	L	Leicht
2016										
Monajjemi et al. [155]	●	●	-	-	-	Feld	Prototyp	Offboard	L	Leicht
Schelle, Stütz [201]	-	-	●	●	-	Feld	Prototyp	Offboard	L	Schwer
2017										
Sun et al. [194]	●	-	-	-	-	Feld	Prototyp	Onboard	L	Mittel
Yeh et al. [167]	-	●	-	-	-	Labor	Explorativ	-	-	Leicht
Colley et al. [210]	-	●	●	-	-	Feld	Explorativ	WoZ	-	Mittel
Peshkova et al. [189]	●	-	-	-	●	Labor	Explorativ	WoZ	L	Leicht
2018										
Mohaimenianpour, Vaughan [184]	●	-	-	-	-	Feld	Prototyp	Onboard		Leicht
Abioye et al. [186]	●	-	●	-	-	VR	Prototyp	Offboard	H	-
Schelle, Stütz [202]	-	-	●	-	-	Feld	Prototyp	Onboard	H	Schwer
Khamis et al. [179]	●	-	-	●	-	Labor	Explorativ	Offboard	L	Leicht
2019										
Wojciechowska et al. [18]	-	●	-	-	-	Labor	Explorativ	Onboard	-	Leicht
Abioye et al. [175]	-	-	●	-	-	Labor	Explorativ	Onboard	H	-
Cauchard et al. [217]	-	-	●	●	-	Feld	Prototyp	Onboard	H	Mittel
2020										
Gromov et al. [180]	●	-	-	-	-	Labor	Explorativ	Offboard	L	Sehr leicht
2021										
Theiß et al. [161]	-	-	-	●	-	Labor	Prototyp	Offboard	-	Leicht
Lieser et al. [208]	-	●	-	●	-	Labor	Prototyp	Offboard	-	Sehr leicht
Tsykunov et al. [185]	●	-	-	-	-	Labor	Prototyp	Offboard	L	Leicht
Palossi et al. [218]	●	●	-	-	●	Labor	Prototyp	Onboard	-	Sehr leicht
Schelle, Stütz [219]	-	-	●	-	-	Labor	Explorativ	-	H	-
Medeiros et al. [195]	-	●	-	-	-	Feld	Prototyp	Offboard	-	Mittel
Marinov et al. [193]	●	●	-	-	-	Labor	Prototyp	Offboard	L	Sehr Leicht

●: relevant; -: nicht relevant; n/a: nicht angegeben

[22] Einteilung nach maximalen Abfluggewicht: sehr leicht (< 100 g), leicht (< 0,5 Kg), mittel (0,5 - 5 Kg), schwer (5 - 25 Kg)

3.2 Bildbasierte Gestenerkennungssysteme in der cHDI

Die ersten Systeme zur Erkennung von Freiraumgesten ohne zusätzliche Geräte in Form von Datenhandschuhen oder optischen Markern gehen bis in die 90er Jahre des letzten Jahrhunderts zurück [220]. Seitdem wurde eine Vielzahl von Ansätzen für die Interaktion mit Computern (HCI), Maschinen (HMI) sowie Robotern (HRI) vorgestellt. Die cHDI kann dabei zwischen der stationären HCI und der mobilen HRI positioniert werden. Nachdem die kollokierte Interaktion im Fokus dieser Arbeit steht, werden im Folgenden diejenigen Gestenerkennungssysteme näher beschrieben, die für die Datenerfassung eine Sensorik an Bord des UAVs nutzen. Ferner werden nur solche Verfahren in Betracht gezogen, welche keine speziellen optischen Marker zur Personendetektion und Gestenerkennung voraussetzen und zudem auf mobilen Trägerplattformen echtzeitfähig sind. Abb. 24 zeigt eine die thematische Eingrenzung dieses Kapitels.

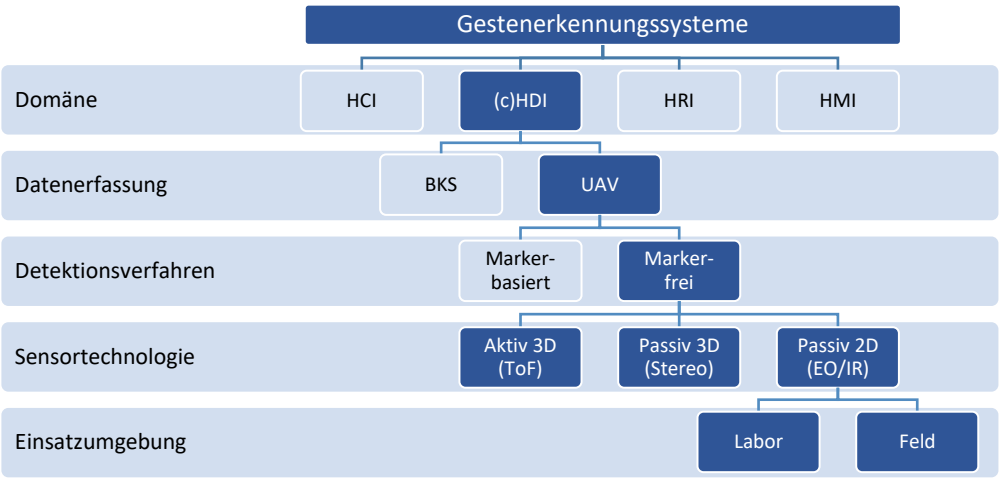


Abb. 24 Einordnung der untersuchten Gestenerkennungssysteme im Kontext der HDI

Verarbeitungskette

Grundsätzlich ähneln sich Gestenerkennungssysteme in den wesentlichen Verarbeitungsschritten, sodass eine generische Bildverarbeitungskette (*Image Processing Pipeline,* IPP) nach folgendem Muster aufgebaut ist (Abb. 25). Über einen bildgebenden Sensor wird der Mensch mitsamt der Umgebung physikalisch erfasst und muss zunächst in den Bilddaten erkannt (Personendetektion) und von störenden Bildanteilen wie Objekten im Vorder- und Hintergrund isoliert werden (Segmentierung). Aus den gefilterten Daten werden im nächsten Schritt die Positionen von relevanten Körpermerkmalen (Kopf, Hände, Finger) identifiziert und auf ein Körpermodell abgebildet (Körpermodellierung). Basierend auf dem Körpermodell werden nun informationstragende Merkmale, wie Winkel, Längenverhältnisse oder höherdimensionale Merkmale erzeugt (Merkmalgenerierung), die eine Zuordnung der Beobachtungen zu einer Klasse aus einer vorgegebenen Menge an Klassen im darauffolgenden Schritt ermöglicht (Klassifizierung). Diese klassifizierten Daten werden im nachgeschalteten Schritt als Entscheidungsgrundlage für das weitere Systemverhalten herangezogen (Interpretation).

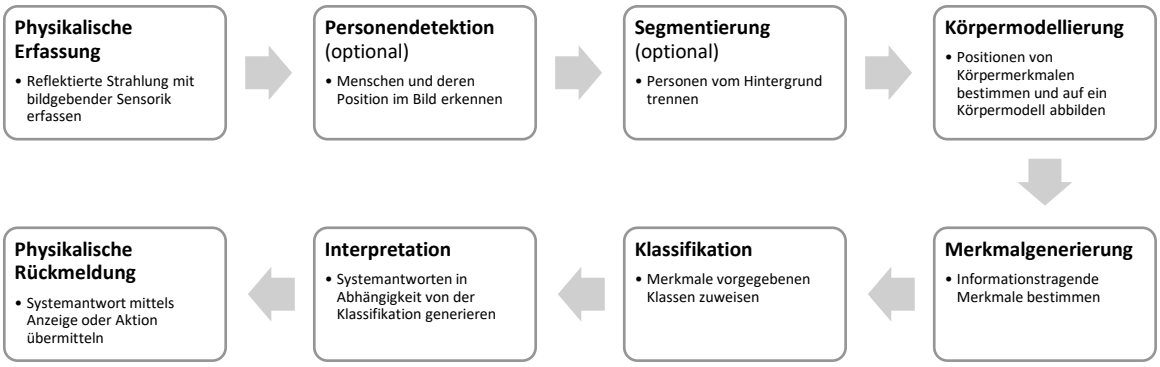


Abb. 25 Wesentliche Verarbeitungsschritte eines generischen Gestenerkennungssystems

Dabei kann die erkannte Geste entweder direkt einer Systemreaktion zugeordnet werden (z.B. Flugrichtungsänderung des UAV) oder mit anderen Modalitäten in einem multimodalen Kontext kombiniert betrachtet und weiter analysiert werden. Am Ende der generischen Verarbeitungskette steht eine Systemantwort, die sich dem Nutzer als sichtbare Reaktion in Form einer Bewegung oder Anzeige auf einem Ausgabegerät zeigt. Wie zuverlässig und robust (Laboreinsatz vs. realistische Umgebung) ein Gestenerkennungssystem ist, bestimmt jedoch nicht nur der Algorithmus, sondern auch die Qualität der Eingangsdaten und damit die verwendete Sensortechnologie.

3.2.1 Sensortechnologien zur Erfassung von Bewegungen

In der HDI werden sowohl aktive als auch passive Kamerasysteme zur zwei- und dreidimensionalen Erfassung von Körperbewegungen eingesetzt. Aktive Systeme beleuchten die Umgebung im Nahinfrarotbereich, um auch Daten im Dunkeln liefern oder zusätzliche Entfernungsinformationen in Form von Punktwolken oder Tiefenkarten bereitstellen zu können. Typische Vertreter sind *Time-of-Flight* (ToF) Kameras, die anhand der Signallaufzeit von kurzen Lichtpulsen oder dem Phasenunterschied zwischen ausgesendeten und empfangenen modulierten Lichtsignalen die Entfernung bestimmen.

Aktive Sensoren

Mit der Vorstellung der ersten Kinect stand für die breite Allgemeinheit nicht nur ein vollintegrierte, aktive Tiefenkamera zur Verfügung, sondern beinhaltete auch ein umfassendes Softwarepaket zur Personenerkennung und -verfolgung mitsamt einem Skelettmodell mit 20 Gelenkpunkten [221]. Dadurch revolutionierte der kostengünstige Sensor nicht nur die Unterhaltungsindustrie, sondern wurde auch für die Forschung ein wichtiges Werkzeug zur Analyse natürlicher Interaktionsformen und ermöglichte die Entwicklung neuartiger Mensch-Maschine-Anwendung. Seitdem haben sich berührungslose Interfaces in vielen Bereichen des Lebens etabliert, sei es zur Steigerung des Nutzererlebnisses im Unterhaltungsbereich oder als assistierende Technologien, um beispielsweise Sterilität im Operationssaal bei der Bedienung von Bildschirmen sicherzustellen [222]. Aktive Kamerasysteme[23] wurden seitdem stetig miniaturisiert und erfüllen die Leistungs- und Gewichtsbeschränkungen für den Einsatz auf kleinen UAVs.

Die Nutzung aktiver Tiefensensoren für UAVs ist jedoch nicht ganz unproblematisch. Viele der Verfahren, die Tiefeninformationen nutzen, um Personen zu erkennen und Posen zu bestimmen, wurden

[23] Vgl. Microsoft Kinect V2 und Azure Kinect DK, Intel Realsense R200 und D455, Asus XTION 2.

originär für stationäre Anwendungen entwickelt und setzen oftmals auf *Background-Subtraction*-Verfahren zur Segmentierung. So stellten Naseer et al. fest, dass die Personenerkennung von OpenNI[24] in Zusammenspiel mit einer auf einem Quadkopter montierten Kinect versagte [223, S. 625] und erst widerhergestellt werden konnte, nachdem die Eigenbewegung der Flugplattform in einem Vorverarbeitungsschritt herausgerechnet wurde. Neben einer eingeschränkten Reichweite, haben aktive Tiefensensoren zudem einen weiteren gravierenden Nachteil, der gegen einen realitätsnahen Einsatz unter Tageslicht spricht. Die für die Projektion des Lichtmusters verwendeten Emitter arbeiten oftmals in einem Wellenlängenbereich, der in das Spektrum der Sonne fällt, sodass bei Anwendungen unter Tageslicht Überlagerungen entstehen, die zu falschen Entfernungsangaben führen [224]. Zwar gab es bisher Bemühungen, den Einfluss der Sonnenstrahlung durch modifizierte Beleuchtungsverfahren [225] oder stärkere Emitter zu kompensieren, die Reichweitenproblematik bleibt jedoch weiterhin erhalten. Insofern finden sich aktive Sensoren vorrangig bei cHDI-Anwendungen in Innenräumen und bei relativ kurzen Interaktionsdistanzen (vgl. [223]).

Passive stereoskopische Kamerasysteme

Passive Kamerasysteme in stereoskopischer Anordnung arbeiten nach dem Prinzip der Triangulation, d.h. ein Raumpunkt muss in den Bildebenen der Kameras mittels Korrespondenzanalyse gefunden werden, um anhand der Disparität die Entfernung zu einem Objekt errechnen zu können. Dafür ist es jedoch unbedingt notwendig, dass die intrinsischen und extrinsischen Kameraparameter bekannt und beide Bilder rektifiziert, d.h. von geometrischen Verzerrungen befreit, sind. Auch muss sichergestellt sein, dass die Belichtung und der Transport der Daten auf beiden Kameras synchronisiert erfolgt. Die Reichweite und Genauigkeit der Tiefenbestimmung solcher Kamerasysteme hängt maßgeblich von der Auflösung der verwendeten Bildsensoren, aber auch vom Basisabstand, d.h. dem horizontalen Abstand beider Kameras zueinander, ab. Nachdem der eingebaute Bildsensor die maximale Auflösung vorgibt, kann die Reichweite üblicherweise nur über den Basisabstand vergrößert werden. Je weiter jedoch beide Kameras voneinander entfernt sind, desto schwieriger wird eine pixelgenaue Ausrichtung. Vibrationen des Antriebs können zu Oszillationen führen, die eine zuverlässige Tiefenmessung erschweren. Gängige, kommerziell erhältlichen und vollintegrierte, stereoskopische Systemen nutzen daher gewöhnlicherweise feste Basisabstände im Bereich von 50 – 120[25] mm und erreichen damit eine effektive Reichweite von etwa 16 Metern [226]. Verfahren, die auf der Korrespondenzanalyse beruhen, haben jedoch allgemein den Nachteil, dass Entfernungen zu homogenen, merkmalsarmen Oberflächen ohne zusätzliche Hilfsbeleuchtung (z.B. durch strukturierte Lichtmuster wie bei *Active Stereo Vision*) nur sehr unzureichend bestimmt werden können.

Passive monokulare Kamerasysteme

Passive elektrooptische Kamerasysteme sind auf eine vorhandene Beleuchtungsquelle angewiesen und bilden die Umwelt zweidimensional in Form von Intensitätswerten ab. Durch den einfachen Aufbau sind sie jedoch wesentlich kompakter und leichter als aktive Tiefensensoren oder passive Multikamerasysteme, sodass sie beinahe auf jedem UAV in einfacher oder mehrfacher Ausführung für unterschiedliche Zwecke zu finden sind. Zudem verfügen sie im Vergleich zu anderen Technologien über die höchste Auflösung und können auch einfacher mit variablen Brennweiten ausgerüstet werden, wodurch die effektive Interaktionsdistanz erweitert werden kann. Das Spektrum passiver Kamerasysteme ist jedoch nicht nur auf den visuellen Bereich begrenzt. Für Mensch-UAV-Interaktionen auch in harschen Umgebungsbedingungen werden Sensoriken benötigt, die es dem UAV erlauben, Menschen

[24] Ein in der Literatur oft genutztes, kommerzielles Softwarepaket zur Posen- und Gestenerkennung.
[25] Intel Realsense D435 (50 mm), Stereolabs Zed 2i (120 mm).

auch bei Dunkelheit und durch Rauch und Nebel erkennen zu können. Ungekühlte thermische Sensoren, die im langwelligen Infrarotspektrum (*Longwave Infrared*, LWIR) empfindlich sind, rücken daher zunehmend in den Fokus der cHDI, beispielsweise für Gestenerkennungssysteme, die auch bei erschwerten Sichtbedingungen [227] funktionieren müssen, oder zur hochautomatisierten Vermisstensuche [201].

3.2.2 Körpermodelle

Der menschliche Körper ist ein komplexes biomechanisches System, u.a. aus Knochen, Gelenken, Muskeln und Sehnen, welches eine sehr große Anzahl an Konfigurationen erlaubt. Möchte man die Körperstellung und damit die Orientierung aller Gliedmaßen im Raum (Pose) mit Methoden des maschinellen Sehens (*Computer Vision*) bestimmen, befindet man sich im Bereich der *Human Pose Estimation* (HPE). Soll die Pose einer Person bestimmt werden, spricht man von *Single-Person* HPE, bei mehreren Personen von einer *Multi-Person* HPE. Viele Anwendungen in der virtuellen Realität, bei der Bewegungsanalyse oder der Aktivitätserkennung (*Action Recognition*) sind auf eine zuverlässige Posenschätzung angewiesen. Durch seine innere Struktur stellt der menschliche Körper die HPE jedoch vor einige Herausforderungen. Neben den kinematischen Abhängigkeiten und Einschränkungen einzelner Gliedmaßen erschwert der symmetrische Aufbau des Körpers zudem eine eindeutige Links-/Rechts-Zuordnung, insbesondere bei zweidimensionalen Erfassungsmethoden. Verdeckungen und Mehrdeutigkeiten können zudem entstehen, wenn sich Körperteile einer oder mehrerer Personen überlappen. Neben rein merkmalbasierten Methoden existieren auch modellbasierte Verfahren, die geeignete Köpermodelle nutzen, um die physikalischen Möglichkeiten des menschlichen Körpers zu beschreiben und die Posenschätzung zu verbessern. Nach Gong et al. werden in der HPE vorrangig drei Körpermodelle [228, S. 10] genutzt.

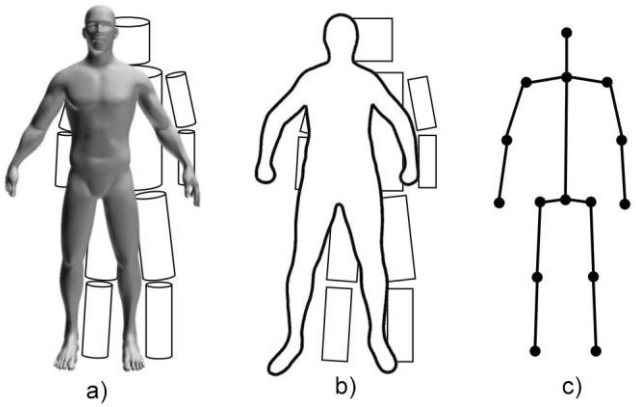


Abb. 26 In der HPE verwendete Körpermodelle: a) Volumetrisches Modell, b) Planares, formbasiertes Modell sowie c) Skelettbasiertes, kinematisches Modell, übernommen aus [229, S. 4]

Volumetrische Modelle bilden den menschlichen Körper mithilfe von miteinander verbundenen geometrischen Formen (Zylinder, Konen, etc.) oder mittels Polygonnetzen (Mesh) als dreidimensionale, flexible Oberflächen ab (Abb. 26a). Mit Modellen wie SCAPE[26] können nicht nur Posen, sondern auch Oberflächenmerkmale (Kleidung, Muskeln, etc.) detailliert nachgebildet werden, wozu jedoch aufwendige 3D-Scans erforderlich sind. Diese Repräsentation wird vorwiegend für die realistische Animation von virtuellen Charakteren benötigt. Für ein echtzeitfähiges Gestenerkennungssystem ist dieser hohe Detailgrad jedoch meist nicht notwendig.

[26] Shape Completion and Animation of People [230].

Formbasierte Modelle stellen eine einfachere Möglichkeit zur Repräsentation von artikulierten Posen dar (Abb. 26b). Dabei wird der menschliche Körper als Satz von starren (*Pictorial Structure Model* [231]) oder verformbaren Körperteilen (*Deformable Structure Model* [232]) abgebildet, die über federartige Verbindungen miteinander in einer Baumstruktur verbunden sind. Daneben kann auch die Körpersilhouette als Grundlage zur Beschreibung einer Pose genutzt werden [233].

Kinematische Modelle hingegen orientieren sich am Skelett und bilden den menschlichen Körper als kinematische Kette von verbundenen Gelenken innerhalb eines gerichteten Graphen ab und sind oftmals an einer „Strichmännchen-Darstellung" erkennbar (Abb. 26c). Trotz der fehlenden Informationen über die Form und Struktur einzelner Gliedmaßen, können über die räumliche Anordnung der Gelenke und der Orientierung der Verbindungsstücke eine Vielzahl von echtzeitfähigen Anwendungen realisiert werden.

3.2.3 Methoden zur luftgestützten Erfassung von Körpermerkmalen

Einfache Gestenerkennungssysteme zur Interaktion mit UAVs können bereits umgesetzt werden, wenn bekannt ist, wo sich die Hände in Relation zu einem Referenzpunkt am Körper, beispielsweise dem Gesicht, befinden (vgl. [214]). Insofern ist nicht immer eine aufwendige Schätzung der Pose notwendig. Aus der Sicht der Bildverarbeitung besteht der menschliche Körper aus einer Menge von visuellen „Objekten", die nach einem bestimmten Muster aufgebaut sind und folglich auch mit Objektdetektionsverfahren in Bilddaten gefunden werden können. Objektdetektoren für Einzelbilder, die auf Pixel- und Kantenebene nach vorgegebenen Mustern (*Templates*) suchen (z.B. HOG[27] oder YOLO[28]), können darauf trainiert werden, Gesichter und Hände zu erkennen, selbst wenn sie in nur sehr niedrigen Auflösungen vorliegen [155]. Auch noch wesentlich kleinere Merkmale, wie Augenbewegungen, können als Körpersignale interpretiert werden. So nutzten Khamis et al. einen CLNF[29]-Detektor, um anhand der Pupillenposition innerhalb der Augenkonturen die Blickrichtung zu bestimmen und damit eine UAV-gestützte Interaktion zu ermöglichen [179]. Soll jedoch nicht nur festgestellt werden, wo sich ein Gesicht im Bild befindet, sondern auch um wessen Gesicht es sich dabei handelt, müssen zusätzliche Gesichtsmerkmale gefunden, beschrieben, normalisiert und mit einem entsprechenden Klassifikator mit einer Datenbank abgeglichen werden. Costante et al. kombinierten dazu einen LPBH-Bildbeschreiber[30] mit einer *Support Vector Machine* (SVM) als Klassifizier, um Gesichter mit einem personalisierten Satz von nutzerspezifischen Gestenklassifikatoren zu verknüpfen und dadurch ein personalisiertes Gestenerkennungssystem auf einem UAV zu integrieren [238].

Für die Erkennung von Körperpartien, die für gewöhnlich nicht von Kleidungsstücken verdeckt werden (Gesicht, Hände) eignen sich aber auch klassische, farbbasierte Verfahren [214], die nicht angelernt werden müssen. Dabei kann eine biologische Tatsache ausgenutzt werden: aufgrund des im Blut befindlichen Hämoglobins ist der Farbton der Haut bei den meisten Menschen ähnlich (vgl. *„Flesh Tone Line"* zur Farbkalibrierung von Farbbildkameras mittels Vektorskop [239]) und unterscheidet primär in der Helligkeit und Farbsättigung durch die Pigmentierung. In einem geeigneten Farbmodell, wie beispielsweise HSV (*Hue, Saturation, Value*), kann eine einfache farbtonbasierte Segmentierung mittels Schwellwertverfahren und morphologischer Basisoperationen umgesetzt werden. Über eine weitere Konturenanalyse der isolierten BLOBs[31] können anschließend zusätzliche Merkmale, wie die

[27] Histograms of Oriented Gradients [234].
[28] You Only Look Once [235].
[29] Constrained Local Neural Fields [236].
[30] Local Binary Patterns Histogram [237].
[31] Binary Large Objects.

Position der Fingerspitzen, lokalisiert werden. Das Ergebnis eines solchen Verfahrens zeigt beispielhaft Abb. 27. Dabei wurden Konvexitätsdefekte in der Hüllkurve der farbsegmentierten Hand zur Bestimmung der Fingerpositionen genutzt, um eine Fingerzählfunktion für eine visuelle Mensch-UAV-Schnittstelle umzusetzen [202].

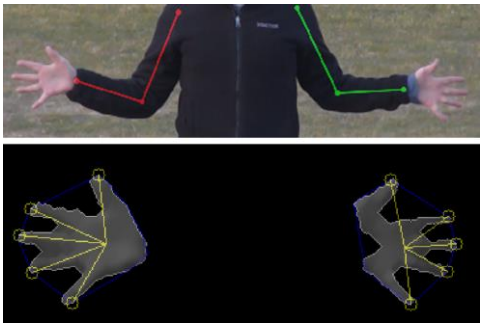


Abb. 27 Lokalisierung der Fingerkuppen (gelbe Kreise) mittels Defekten in der konvexen Hüllkurve (blaue Kontur) der farbsegmentierten Hand (weiße Kontur), eigene Darstellung

Farbbasierte Ansätze geraten jedoch schnell an ihre Grenzen, wenn abrupte Farbveränderungen eintreten (z.B. wechselnde Bewölkung) oder sich gleichfarbige Objekte im Hintergrund befinden.

Formbasierte Verfahren stellen eine robustere Alternative dar, benötigen jedoch in der Regel ebenfalls eine vorgeschaltete Segmentierung zur Trennung eines Menschen von der Umgebung. Stehen Tiefendaten zur Verfügung, kann über ein einfaches *Clipping*-Verfahren mit festen oder adaptiven Schwellwerten der Mensch aus der Szenerie isoliert und die Position der Extremitäten mittels Extrempunktanalyse lokalisiert werden. Ein solches Verfahren wurde beispielsweise in [240] vorgestellt und setzt auf folgenden Algorithmus (Abb. 28): mittels eines HOG-basierten Objektdetektors wird eine Person im Infrarotbild des Tiefensensors lokalisiert und der relevante Bildausschnitt (*Region of Interest*, ROI) auf das Tiefenbild übertragen (b) sowie auf den Oberkörperbereich beschränkt (c). Nach der Entfernung des Hintergrunds mittels Medianfilterung (d) werden die verbleibenden Daten binarisiert (e) und deren Kontur bestimmt (f). In dieser Kontur werden zunächst Kandidaten gesucht, die maximal weit vom Körperzentrum entfernt sind und sich gleichzeitig außerhalb einer definierten Körperbegrenzung (g) befinden. In einem anschließenden Schritt werden die verbleibenden Kandidaten einem Körperteil mittels Plausibilitätsprüfung zugewiesen (h).

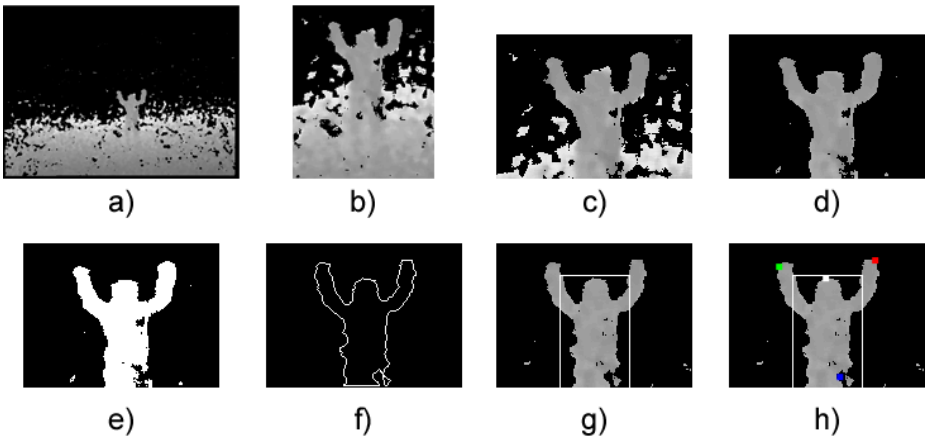


Abb. 28 Lokalisierung der Extremitäten in stark verrauschten Tiefendaten eines stereoskopischen Sensors aus einem schwebenden UAV (Beschreibung der Verarbeitungsschritte im Text). Abbildung h) zeigt die erkannten Extrempunkte: **grün** = linker Arm, **rot** = rechter Arm, **blau** = Hüfte, **weiß** = Kopf. Eigene Darstellung übernommen aus [240, S. 180].

Diese Methode eignet sich für einfache Gestenerkennungssysteme, die lediglich die Position der Hände zur Kodierung von Informationen nutzen. Sollen darüber hinaus weitere Körpermerkmale als Kommunikationsmittel verwendet werden, wird eine detailliertere Analyse benötigt, um beispielsweise die Orientierung und Position einzelner Gliedmaßen aus einer Punktwolke mittels geodätischer Distanzen auf ein Skelettmodell zu übertragen (Abb. 29).

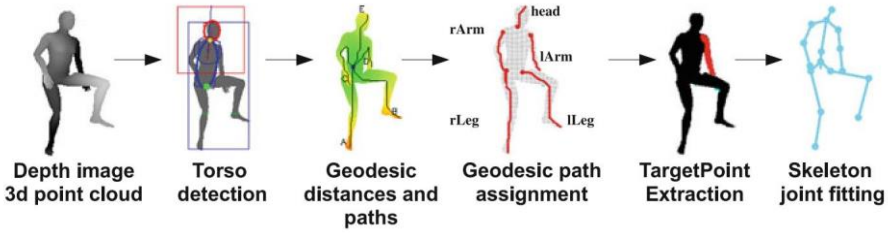


Abb. 29 Posenbestimmung mittels Tiefendaten und geodätischen Distanzen, übernommen aus [241, S. 325]

Generative und Deskriptive Methoden

Generative Methoden passen ein Modell des menschlichen Körpers an die Beobachtungen an. Dafür wird üblicherweise eine initiale Kalibriergeste vor der Interaktion benötigt, um die Länge der Gliedmaßen abschätzen zu können und das Skelettmodell auf den Nutzer abzustimmen. Oftmals wird hierfür die sogenannte *Psi-Pose* verwendet, bei der beide Arme seitwärts ausgestreckt werden und zusammen mit dem Rumpf den Buchstaben T nachbilden. Mit dieser Methode können zwar alle Posen abgebildet werden, schnelle Bewegungen können jedoch zu Trackingfehlern führen. Zudem stellt die notwendige Kalibrierpose eine unnatürliche Komponente innerhalb der Interaktion dar.

Deskriptive Methoden gehen den umgekehrten Weg und extrahieren relevante Merkmale aus den Bilddaten und gleichen diese mit zuvor erlernten Posen ab. Für die Klassifikation werden SVMs, Entscheidungsbäume (*Decision Trees*) oder künstliche neuronale Netze genutzt [241, S. 322]. Bei diesem Verfahren wird keine Kalibrierpose benötigt, dafür können nur diejenigen Posen erkannt werden, die auch im Trainingsdatensatz enthalten waren.

Top-Down und Botton-Up-Ansätze

In der HPE existieren prinzipiell zwei Herangehensweisen, um Posen in Einzelbildern zu bestimmen. Bei klassischen *Top-Down*-Ansätzen wird der Suchraum zunächst mit einer Personendetektion verkleinert, bevor anschließend nach Körpermerkmalen gesucht wird, die zu einem Körpermodell zusammengesetzt werden können. Diese Ansätze eignen sich, wenn einige wenige Personen in den Bilddaten abgebildet sind. Bei teilweisen Verdeckungen oder dicht nebeneinanderstehenden Menschen können diese Verfahren jedoch schnell versagen, wenn der Personendetektor nicht anschlägt. Darüber hinaus skalieren diese Methoden schlechter, d.h. je mehr Personen im Bild sichtbar sind, desto rechenintensiver werden diese Verfahren [242, S. 133332].

Einen anderen Weg gehen sogenannte *Bottom-Up*-Ansätze. Dabei wird bereits im ersten Schritt versucht, anatomische Körpermerkmale, vorzugsweise Gelenke (engl. *joints*), im Gesamtbild ohne eine vorgeschaltete Personenerkennung zu lokalisieren. Erst in den darauffolgenden Schritten werden die Beobachtungen zu einem kinematischen Körpermodell zusammengesetzt. Der große Vorteil dieser Methode liegt darin, dass auch eine Detektion möglich ist, wenn nur Teile des Körpers zu sehen sind.

Convolutional Pose Machines

Die Veröffentlichungen von umfangreichen annotierten Bilddatensätzen mit artikulierenden Menschen (z.B. *LSP*[32], *COCO*[33], *MPII*[34], *Human3.6*[35]) als auch der kostengünstige Zugriff auf kommerzielle Hochleistungsrechner im Rahmen von *Cloud-Computing* ermöglichten die Weiterentwicklung von künstlichen neuronalen Faltungsnetzwerken (*Convolutional Neural Network*, CNN) zur Objektdetektion. *DeepPose* [247] gilt dabei als Vorreiter bei der Lokalisierung von anatomischen Merkmalen mittels CNN und ebnete den Weg für die nachfolgenden „*Convolutional Pose Machines*" (CPM) [248]. CPMs erzeugen aus einem Eingangsbild (Abb. 30a) für jedes antrainierte Körpermerkmal (Gelenk, Knöchel, etc.) eine eigene Konfidenzkarte (*Part Confidence Map*), die beschreibt, mit welcher Wahrscheinlichkeit ein Pixel im Bild zum entsprechenden Körpermerkmal gehört (vgl. Abb. 30b). Mit einem breit aufgestellten Trainingsdatensatz kann eine hohe Generalisierung erreicht werden, sodass CPMs klassische Methoden in der Detektionsleistung, insbesondere in realistischen „in-the-wild" Umgebungen, schnell übertrafen. Jedoch stellten überlappende Gliedmaßen in Multi-Personen-Szenarien in Einzelbildern nach wie vor eine große Herausforderung dar, weil eine eindeutige Zuordnung der detektierten Körpermerkmale zu einem Körperteil einer Person nicht immer möglich war.

Cao et al. lösten dieses Problem durch die Einführung von sogenannten *Part Affinity Fields* (PAFs) als Teil ihres *OpenPose* Algorithmus [249]. PAFs sind zusätzliche Merkmalsbeschreiber, die die Vorzugsrichtung eines bestimmten Körperteils (z.B. linker Oberarm) von einem anatomischen Merkmal (z.B. linke Schulter) zum anderen (z.B. linker Ellenbogen) in Form eines Vektorfeldes kodieren (vgl. Abb. 30c). Durch eine Integration über dieses Feld kann eine „Affinitäts-Metrik" umgesetzt und damit die Zuordnung von zwei zu einem Körperteil gehörenden Körpermerkmalen wesentlich vereinfacht werden. Um anschließend aus der Menge von detektierten Körperteilen zusammenhängende Körpermodelle bilden zu können, formulierten sie die Körperteil-Skelett-Zuordnung als *Maximum Matching Problem* für bipartite Graphen (vgl. Abb. 30d), die sie mittels der Ungarischen Methode [250] lösten.

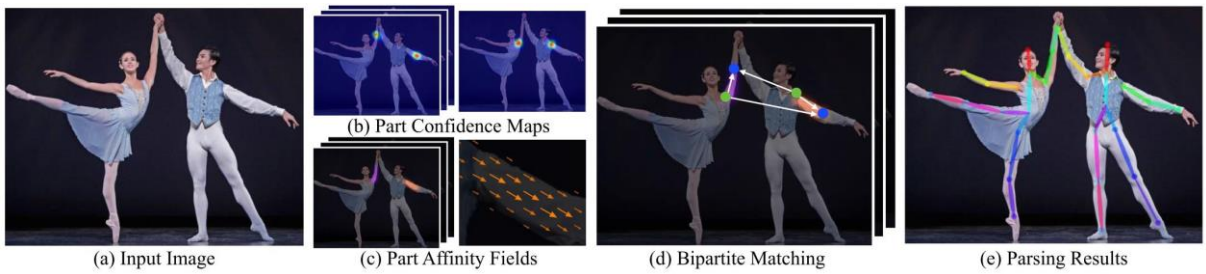


Abb. 30 DNN basierte Pipeline von OpenPose zur Posenschätzung mittels Part Confidence Maps und Part Affinity Fields, übernommen aus [251, S. 2]

Weitere DNN basierte HPE-Methoden folgten, die mit alternativen Netzwerktopologien jeweils unterschiedliche Optimierungen vornahmen. Groos et al. steigerten mit *EfficientPose* [252] die allgemeine Detektionsgenauigkeit, während *CrowdPose* [253] und *PifPaf* [254] entwickelt wurden, um die Erkennungsleistung bei großen Menschengruppen und kleinen Abbildungsmaßstäben zu verbessern. Bazarevsky et al. stellten in *BlazePose* [255] eine vereinfachte Netzwerkstruktur vor, um eine echtzeitfähige Posenschätzung auf leistungsarmen Mobilgeräten wie Smartphones zu ermöglichen. Palossi et al. präsentierten eine weitere Netzwerktopologie namens *FrontNet* [218], welche mit einigen wenigen Hundert Milliwatt und einer speziellen *Parallel Ultra-Low Power* (PULP) Prozessorarchitektur auf extrem

[32] Leeds Sports Pose Dataset [243].
[33] Microsoft Common Objects in Context [244].
[34] Max Planck Institute for Informatics Human Pose Dataset [245].
[35] 3.6 Million Accurate 3D Human Poses Dataset [246].

ressourcenarmen Plattformen, wie beispielsweise Nano UAVs betrieben werden kann. Für eine weitere Aufzählung von HPE-Varianten sei an dieser Stelle auf die Übersichtsarbeit von Chen et al. verwiesen [229].

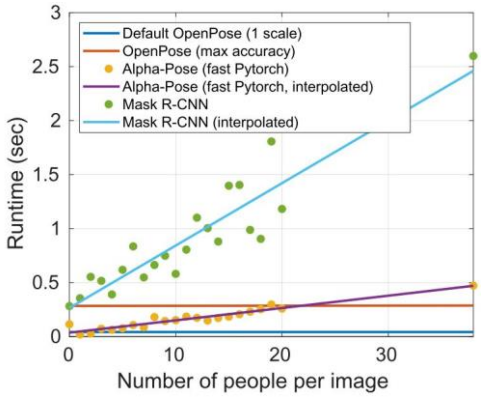


Abb. 31 Inferenzzeiten verschiedener DNN-Ansätze zur Multi-Person HPE im Vergleich, übernommen aus [251, S. 9]

Trotz der unzähligen Alternativen an DNN basierten HPE-Verfahren setzte sich OpenPose in der HDI-Forschung als beliebtes Werkzeug zur Erfassung von Körpermerkmalen durch (vgl. [193, 256]), was durch mehrere Gründe begünstigt wurde. Durch die zeitnahe Veröffentlichung der Implementierung als *Open Source Framework*[36] wurde die Weltgemeinschaft früh in die Weiterentwicklung der Software einbezogen. Die ausführliche Dokumentation und die bereitgestellten Konfigurationen für mehrere Rechnerarchitekturen (x64, ARM) und Programmiersprachen (C++, Python) erleichterten weiterhin die Integration der Methode in andere Anwendungen. Im Vergleich zu anderen Verfahren bietet der Bottom-Up-Ansatz von OpenPose zudem den großen Vorteil, dass die Inferenzzeit, d.h. die Durchlaufdauer der Daten durch das Netzwerk, unabhängig von der Anzahl der Personen im Bild ist (vgl. Abb. 31) und die Netzwerkarchitektur (*Multi-Branch Feedforward CNN*) eine echtzeitfähige, parallele Prozessierung auf Grafikkarten (z.B. mittels *CUDA* oder *OpenCL*) erlaubt. Einen Auszug der in der cHDI eingesetzten Methoden zur Erkennung von Körpermerkmalen aus Bilddaten zeigt Tab. 4.

[36] https://github.com/CMU-Perceptual-Computing-Lab/openpose

Tab. 4 Literaturübersicht der betrachteten Verfahren zur visuellen Erfassung von Körpermerkmalen in der cHDI

Referenz	Sensor-technologie	Verfahren					Erfasste Körpermerkmale				Nutzermodell	Bemerkung
	LWIR, IR, EO, Stereo, ToF	Tiefendatenanalyse	Farbanalyse	Formbasiert	Supervised ML[37]	Unsupervised ML[38]	Gliedmaßen und Gelenke	Hände und/oder Finger	Gesicht	Augen/Blickrichtung	2D, 3D	
2012												
Song et al. [200]	Stereo	●	-	●	-	-	●	●	-	-	3D	NATOPS
2013												
Naseer et al. [223]	Stereo	●	-	-	-	-	●	●	-	-	3D	Kinect + OpenNI
2014												
Mantecón et al. [257]	ToF	-	-	-	-	●	-	●	-	-	3D	Handerkennung
Nagi et al. [214]	EO	-	●	-	-	-	-	●	●	-	2D	Armrichtung
Costante et al. [238]	EO	-	●	-	●	-	●	-	-	-	2D	Personalisiert
2016												
Schelle, Stütz [201]	LWIR, IR, Stereo	●	-	●	●	-	-	●	-	-	2D/3D	
2017												
Sun et al. [194]	EO	-	●	-	-	-	-	●	-	-	2D	
Schelle, Stütz [240]	IR, Stereo	●	-	●	●	-	-	●	-	-	3D	
2018												
Schelle, Stütz [202]	EO	-	●	●	-	●	●	●	-	-	2D	OpenPose
Mohaimenianpour et al. [184]	EO	-	-	-	-	●	-	●	●	-	2D	YOLOv2
Khamis et al. [179]	EO	-	-	-	-	●	-	-	●	●	2D	OpenFace
2019												
Cauchard et al. [217]	Stereo	●	-	-	-	-	-	●	-	-	-	Bodenprojektion
2020												
Kassab et al. [258]	EO	-	-	-	-	●	-	●	●	-	2D	Tiny YOLOv2
2021												
Marinov et al. [193]	EO	-	-	-	-	●	●	-	-	-	2D	OpenPose
Bhat et al. [183]	EO	-	-	-	-	●	●	-	-	-	2D	OpenPose Variante
Alon et al. [227]	LWIR, EO, ToF	●	-	●	●	●	●	●	-	-	3D	Realsense SDK

●: relevant; -: nicht relevant; n/a: nicht angegeben

3.2.4 Methoden zur Klassifizierung von Körpersignalen

Visuelle Körpersignale können als räumlich-zeitliche Muster beschrieben werden, die sich in Form, Verlauf und Dauer unterscheiden [168, S. 314] und damit als Klassifikationsproblem im Bildbereich als auch im Zeitbereich interpretiert werden. Abhängig davon, wann die Klassifizierung erfolgt, werden Gestenerkennungsverfahren dabei zu Offline- oder Online-Systemen zugeordnet. Offline-Systeme arbeiten mit bereits aufgezeichneten Daten und können bei der Klassifizierung in der Zeitachse vor und zurück springen, um relevante Gestenereignisse im Datenstrom zu identifizieren.

Online-Systeme hingegen können nur auf die aktuellen Live-Sensordaten zurückgreifen und müssen den Datenstrom kontinuierlich auf mögliche Ereignisse analysieren. Insofern ergeben sich gegenüber Offline-Systemen zusätzliche Herausforderungen (vgl. [259, S. 85]). Gestische Äußerungen gehen zumeist fließend ineinander über, sodass der Anfang und das Ende einer Geste im Voraus nicht bekannt sind. Auch wird von Online-Gestenerkennungssystemen erwartet, dass jede durchgeführte Geste ide-

[37] k-nearest-Neighbour, k-Means, AdaBoost Haar-like Features, Particle Filtering, HOG
[38] DNN (Deep Neural Network), CNN (Convolutional Neural Network), Recurrent Neural Network (RNN)

alerweise nur einmal als solche erkannt wird. Das dabei verwendete Verfahren muss zudem auch innerhalb der Leistungskapazitäten des Systems bleiben, um eine Echtzeiterkennung zu erlauben. Onlinesysteme leiden zudem unter dem *„Immersion Syndrome“* [260, S. 30], d.h. durch die permanente Bewegungsanalyse besteht die Gefahr, dass das System nicht zwischen intentionalen und zufälligen, unbeabsichtigten Gesten unterscheiden kann. Das hat zur Folge, dass der Nutzer während der Interaktion mit dem System nicht in der Lage ist, gleichzeitig mit anderen Geräten oder Personen nonverbal zu interagieren. Daher gliedert sich die Gestenerkennung typischerweise in zwei zeitlich aufeinanderfolgende Phasen: in der ersten Phase muss zunächst erkannt werden, dass überhaupt eine intentionale Geste durchgeführt wurde (*Gesture Spotting*) und in der nachfolgenden Klassifikationsphase wird bestimmt, welche Geste es war. Das Ziel ist es also, diesen kontinuierlichen, unsegmentierten Strom aus Bewegungsdaten in diskrete lexikalische Entitäten zu transformieren.

3.2.4.1 Gesture Spotting

Um intentionale Gesten von ungewollten Bewegungen zu unterscheiden, können sowohl operationale als auch signaltechnische Segmentierungsmethoden genutzt werden. Operationale Segmentierungsmethoden setzen bereits auf der Seite des Gestenproduzenten an und versuchen, eine möglichst eindeutige Trennung von gewollten und ungewollten Gesten mittels eines geeigneten Gesten- und Interaktionsdesigns zu erreichen. Oftmals wird dafür eine neutrale Pose als „Klasse 0“ definiert, die der Nutzer einnehmen kann, ohne ein Signal an das Gestenerkennungssystem zu senden. Verlässt der Nutzer diese, wird die Gestenanalyse aktiviert (vgl. [240]). Auch können spezielle, gestisch relevante Raumbereiche definiert werden, in denen gestische Eingaben verarbeitet werden. So modellierten Mohaimenianpour et al. [184] in ihrem Gestenerkennungssystem zwei quadratische Gestenräume um das detektierte Gesicht als *„home zones“*, um damit eine Art „visuellen Totmannschalter“ zu realisieren. Zur Übermittlung von Flugsteuerungsanweisungen an ein UAV muss der Nutzer mindestens eine Hand in einem dieser Gestenräume belassen und gleichzeitig mit dem anderen Arm die Bewegungsrichtung vorgeben (Abb. 32a).

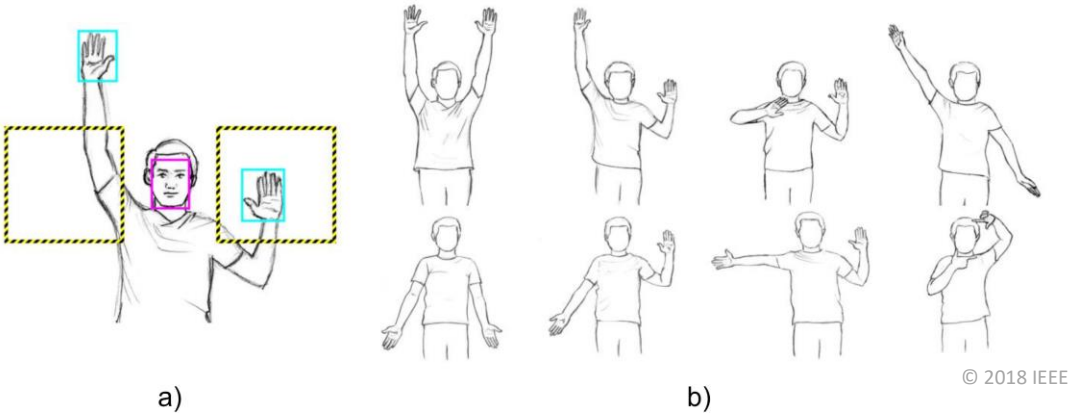


Abb. 32 a) Gestische „home zones“ (gestrichelt) um einen Nutzer, b) das verwendete Gestenvokabular, übernommen aus [184, S. 5227]

Die Nutzung von speziellen Gestenräumen kann die Zuverlässigkeit eines Gestenerkennungssystems verbessern, schränkt jedoch den möglichen Bewegungsspielraum des Nutzers und damit die Natürlichkeit der Interaktion stark ein. Insofern wird bei interaktiven Systemen angestrebt, Gesten vorzugsweise auf der Signalebene zu segmentieren.

Escalara et al. unterteilen Gestensegmentierungsverfahren auf der Signalebene in *direkte* (heuristische) und *indirekte* (erkennungsbasierte) Methoden [261, S. 10]. Direkte Methoden suchen in Low-

Level Bewegungsparametern (z.B. Geschwindigkeits-/Beschleunigungsverläufen) nach markanten Signalwechseln, beispielsweise nach einer steilen Flanke, um einen Gestenwechsel zu erkennen. Erkennungsbasierte Verfahren verwenden stattdessen höherwertige und multidimensionale Verfahren (z.B. DTW, HMM), um den Beginn einer bekannten Signalform im kontinuierlichen Datenstrom zu erkennen. Eine weitere Möglichkeit stellen auch *Sliding-Window*-Ansätze dar, die jeweils nur einen zeitlichen Ausschnitt des Signals analysieren.

3.2.4.2 Regelbasierte und statistische Gestenklassifikationsverfahren

Regelbasierte Gestenerkennungssysteme ähneln klassischen Expertensystemen und modellieren Gesten anhand von Kombinationsregeln von Low-Level Parametern (z.B. Winkelbeziehungen, Geschwindigkeiten, Beschleunigungen) und klassifizieren sie mithilfe von Inferenzmaschinen. Durch den logischen Aufbau können mit dieser Methode auch formelle Beschreibungssprachen mit einer definierten Syntax realisiert werden, wie beispielsweise die *„Gesture Description Language"* von Hachaj und Ogiela [262]. Regelbasierte Erkennungssysteme bieten den Vorteil, dass das Gestenvokabular ohne großen Aufwand erweitert werden kann, weil keine Trainingsphase für die Klassifikation benötigt wird. Zudem können Systementscheidungen einfacher nachvollzogen werden. Die Klassifikationsleistung und Robustheit des Verfahrens (bspw. bei unterschiedlichen Betrachtungswinkeln) hängen maßgeblich davon ab, ob geeignete Merkmale für die Klassifizierung gewählt und wie strikt die Regeln zur Beschreibung einer Geste formuliert wurden. Im Gegensatz zu regelbasierten Verfahren benötigen statistische Verfahren unterschiedlich große und teilweise annotierte Datensätze für die Clusteranalyse und das Training der entsprechenden Klassifizierer. Typische Vertreter sind SVM, Partikelfilter, *Adaptive Naive Bayes Classifier* (ANBC), *Hidden-Markov*-Modelle (HMM), *Dynamic Time Warping* (DTW) oder endliche Zustandsautomaten. Statt binärer Ergebnisse, liefern statistische Verfahren Konfidenzwerte oder Detektionswahrscheinlichkeiten.

3.2.4.3 Statische Darstellungen

Für die Erkennung von Posen reicht es oft aus, die geometrischen Verhältnisse bestimmter Körpermerkmale über einen definierten Zeitraum zu beobachten. Eine einfache Methode bietet die Klassifikation von Gesten anhand ihres Ausführungsortes [194, 201]. Dazu wird der Raum um eine Person in ein System aus zweidimensionalen Rechtecken oder dreidimensionalen Würfeln zerlegt und die Verweildauer der entsprechenden Artikulatoren (zumeist der Hände) darin analysiert. Ähnlich zum Winkeralphabet aus der Schifffahrt kann so die Armstellung in Signale kodiert werden (Abb. 33). Nachteilig an dieser Methode ist jedoch, dass die Anzahl möglicher Kombinationen nicht beliebig erhöht werden kann. Auch das eher technisch anmutende Gestenvokabular kann von Ungeübten als wenig intuitiv wahrgenommen werden. Zudem müssen die Positionen und Dimensionen der Gestenräume an die körperlichen Proportionen des Nutzers angepasst werden. Um auch Bewegungen in der frontalen Ebene (vor und zurück) erkennen zu können, werden darüber hinaus Tiefeninformationen benötigt.

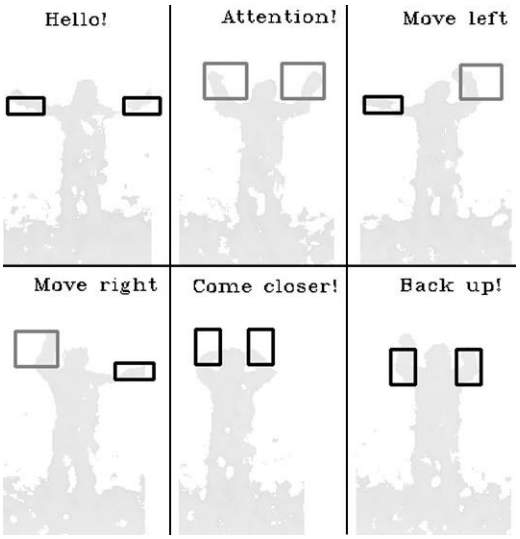


Abb. 33 Gestenklassifizierung anhand von dreidimensionalen Gestenräumen, eigene Abbildung übernommen aus [201, S. 96]

Stehen Trainingsdaten zur Verfügung, können auch statistische Verfahren (z.B. SVM) für das Clustering der Daten und die Klassifizierung gewählt werden.

3.2.4.4 Dynamische Darstellungen

Dynamische Darstellungen, die sowohl räumlich als auch zeitlich variieren, können auch aus der Sicht der Signalverarbeitung als mehrdimensionale Zeitreihen mit diskreten Werten interpretiert und in periodische und nicht-periodische Signale unterteilt werden. Wink- oder Kreisbewegungen stellen periodische Signale dar und können sowohl im Zeitbereich als auch im Frequenzbereich (z.B. über eine schnelle Fourier Transformation) im Datenstrom relativ einfach identifiziert werden [263, S. 968, 264], wohingegen der Beginn und das Ende von nicht-periodische Signale aufgrund der Segmentierungsproblematik schwerer zu bestimmen sind.

Das „Aircraft Signals NATOPS[39] Manual“ der U.S. Navy [120] enthält die standardisierten Marshaller-Handsignale, die auf amerikanischen Flugzeugträgern zur nonverbalen Kommunikation genutzt werden. Der Kodierung der Signale basiert vorwiegend auf einer Mischung aus statischen, periodischen und nicht-periodischen Gesten sowie unterschiedlichen Handformen. Song et al. stellten ein mehrstufiges Verfahren zur visuellen Erfassung dieser Signale vor [200]. Der auf Tiefendaten setzende Ansatz nutzt dabei eine IPP zur Bestimmung der Körperpose mittels eines generativen Körpermodells in Kombination mit einer Partikelfilterung. Parallel dazu findet eine weitere Klassifizierung der Handform mittels eines HOG-basierten Merkmalbeschreibers und einer Multiklassen-SVM statt. Die Ergebnisse beider IPPs bilden die Datengrundlage für das nachgeschaltete kontinuierliche Gestenerkennungsverfahren, welches auf dem LDCRF-Verfahren[40] beruht und um einen zeitlichen Sliding-Window-Ansatz sowie zusätzliche Filterungsschichten erweitert wurde. Mit diesem Verfahren konnten 24 Handsignale mit einer Genauigkeit von 75,37 % erkannt werden. Die Autoren weisen darauf hin, dass das verwendete Verfahren eine Latenz von zwei bis drei Sekunden besitzt und die Erkennungsleistung durch einen Kontextbezug noch weiter gesteigert werden könnte. Je nachdem, welches Signal zuletzt über-

[39] Naval Air Training and Operating Procedures Standardization.
[40] Latent-Dynamic Conditional Random Field [265].

mittelt wurde (z.B. Bremsen anziehen), kann eine Untermenge des Wortschatzes (z.B. alle Rollanweisungen) in der nachfolgenden Phase nicht mehr relevant sein und damit den Suchraum wesentlich verkleinern.

Eine weitere Möglichkeit zur Klassifizierung von dynamischen Einweiserzeichen bietet eine Analyse der Gestenphasen (siehe Kapitel 2.2.1). Choi et al. [266] verwendeten einen farbbasierten Ansatz zur Bestimmung der Handpositionen und wiesen mittels eines geometrischen Regelsatzes jedem Zeitschritt der Handbewegungen einen von vier Bewegungsphasen (*Halte-*, *Dynamik-*, *Verdeckungs-* und *Normalphase*) in Form eines Symbols zu. Mittels der LCS-Methode[41] erzeugten sie aus dieser Sequenz aus Symbolen für jedes Handsignal ein entsprechendes Bewegungsmodell. Das Verfahren lieferte zwar bei dem Testdatensatz aus acht dynamischen Handsignalen eine höhere Erkennungsrate als vergleichbare HMM-Verfahren (99 % vs. 86,5 %), enthielt jedoch keine Gesture-Spotting-Komponente, sodass nur zeitlich getrimmte Daten untersucht wurden.

3.2.5 Zusammenfassung und Literaturübersicht

Für die Erkennung gestischer Nutzereingaben werden in der Literatur unterschiedliche Sensortechnologien, Nutzermodelle sowie vielfältige Merkmalextraktions- und Klassifizierungsverfahren eingesetzt. Monokulare passive Sensoren, die im visuellen Spektrum empfindlich sind und auch Farbinformationen bereitstellen können, stellen dabei die bevorzugte Technologie für den Einsatz auf fliegenden Systemen dar, die im Freien unter Sonnenlichteinfluss betrieben werden. Im Vergleich zu anderen passiven Technologien (LWIR, IR, Stereo) bieten sie eine ausreichend hohe Abbildungsleistung und sind zudem, im Gegensatz zu aktiven Verfahren, auch im Erfassungsbereich nicht limitiert. Für die echtzeitfähige Modellierung der Nutzeranatomie werden kinematische Modelle bevorzugt, die den menschlichen Körper hinreichend genau beschreiben, um eine breite Palette von statischen und dynamischen Darstellungen abbilden zu können. Convolutional Pose Machines, die zudem nach dem Bottom-Up-Detektionsansatz arbeiten (z.B. OpenPose), können die dafür notwendigen Gelenkpositionen aus 2D-Bilddaten extrahieren und damit ein Nutzermodell auch bei teilweiser Verdeckung der jeweiligen Körpermerkmale bereitstellen. Für die Klassifizierung der erfassten Körpermerkmale werden in der Literatur unterschiedliche geometrische, regelbasierte und statistische Verfahren aus dem Bereich der Mustererkennung und Signalverarbeitung genutzt, wobei ein Trend zu holistischen Ansätzen in Verbindung mit Machine Learning Verfahren beobachtet werden kann. Dieses Vorgehen bietet den Vorteil, dass relevante Körpermerkmale nicht manuell modelliert werden müssen, sondern während der Lernphase des Klassifikationssystems in den Gewichtungsfaktoren des Modells abgebildet werden. Der große Nachteil hierbei ist jedoch, dass hierfür ein umfangreicher Satz an geeigneten Trainingsdaten benötigt wird und das Modell zudem mit jeder Klassenerweiterung aufs Neue trainiert werden muss, was nicht nur ressourcen- und zeitintensiv ist, sondern auch das Klassifikationsverfahren wenig flexibel erscheinen lässt.

Von einer visuellen Benutzerschnittstelle wird jedoch eine gewisse Flexibilität erwartet, um verschiedene nonverbale Vokabulare verarbeiten zu können. Insofern werden alternative Klassifikationsansätze benötigt, die auch eine einfache, nachträgliche Anpassung des Vokabulars an den jeweiligen Nutzungskontext erlauben. Tab. 5 zeigt die untersuchten Gestenerkennungsverfahren aus der cHDI in der Übersicht.

[41] Longest Common Subsequence [267].

Tab. 5 Literaturübersicht der betrachteten Gestenerkennungsverfahren in der cHDI

Referenz	Daten	Erkennungs-verfahren				Gesten-typen		UAV		Anzahl implementierter Gesten	Erweiterbar ohne zusätzl. Training	Größe Training/ Test
		Regelbasiert	Klassisch[42]	CNN	LSTM	Statisch	Dynamisch	Am Boden	In der Luft			
2008												
Choi et al. [266]	2D	-	●	-	-	-	●	●	-	8	-	182/202
2009												
Venetsky, Tieman [264]	2D	-	-	●	-	●	●	●	-	20	n/a	n/a
2012												
Song et al. [200]	3D	-	●	-	-	●	●	●	-	24	-	4800/2400
Higgins [268]	2D	●	●	-	-	●	●	●	-	10	n/a	n/a
2014												
Mantecón et al. [257]	3D	-	●	-	-	●	●	●	-	11	-	5280/1320
2016												
Monajjemi et al. [155]	2D	●	●	-	-	-	●	-	●	1	-	-
Schelle, Stütz [201]	2D/3D	●	●	-	-	-	●	-	●	7	-	n/a
2017												
Kawatsu et al. [269]	2D	-	-	●	●	●	-	●	-	7	-	45k/4k
Sun et al. [194]	2D	●	-	-	-	●	-	-	●	3	●	-
2018												
Natarajan et al. [270]	2D	-	●	-	-	●	-	-	●	5		8k
Mohaimenianpour et al. [184]	2D	●	-	-	-	●	-	-	●	8	●	16883
Schelle, Stütz [202]	2D	●	-	-	-	●	-	●	●	19	●	-
2021												
Zhou et al. [271]	2D	-	-	●	-	●	-	●	●	10	-	3024/1296

●: relevant; -: nicht relevant; n/a: nicht angegeben

3.3 Design und Verfügbarkeit von gestischen Vokabularen in der cHDI

Handgesten als Kommunikationsmethode für interaktive Systeme sind attraktiv, weil sie natürlich, universell und direkt sind [272]. Die dazu notwendigen Gestenvokabulare werden von der Forschergemeinde jedoch mit unterschiedlichen Zielvorstellungen entwickelt. Die nachfolgenden Abschnitte beschreiben die gängigsten Motive.

3.3.1 Motive

Gestische Vokabulare werden manchmal nur entwickelt, um den Funktionsnachweis für ein neues Gestenerkennungsverfahren zu erbringen [181, 183]. Oftmals fließen dabei technologische Einschränkungen wie die Erkennungsleistung des vorgestellten Systems in das Design des Vokabulars mit ein, was zu unnatürlichen Gesten führen kann, die für einen potentiellen Nutzer nur schwer zu erlernen sein können [273, S. 636]. In anderen Fällen werden Gestenerkennungssysteme auf der Grundlage von bereits etablierten, domänenspezifischen Handzeichensystemen, wie beispielsweise aus der Luftfahrt

[42] Haar-like Features, Single/Multi Class Support-Vector-Machine (SVM), Dynamic-Time-Warping (DTW), Hidden-Markov-Modelle (HMM), Histogramm-of-Oriented Gradients (HOG)

[274], dem Feuerwehrwesen [227] oder der Beizjagd [17] entwickelt, sodass die Elemente des Gestenvokabulars bereits größtenteils vorgegeben sind. Zwar wurden bereits auch Gebärdenerkennungssysteme, z.B. für die Amerikanische [275], Indische [276] und Deutsche Gebärdensprache [277] vorgestellt, jedoch sind einige Forscher der Auffassung, dass das gegenwärtige Niveau zur rechnergestützten Erkennung von Gebärdensprache in etwa dem Niveau der Spracherkennung vor 30 Jahren entspricht, weil erst allmählich ein Übergang von einer isolierten Einzelgestenerkennung zu einer kontinuierlichen Erkennung von Aufgaben mittels eines kleinen Wortschatzes stattfindet [278, S. 325].

Im Gegensatz zu den technisch motivierten Ansätzen, stehen nutzerzentrierte Ansätze, die physiologische, psychologische und kognitive Faktoren in den Fokus der Interaktion stellen und versuchen, störende Aspekte wie Ermüdungserscheinungen [279] und Mehrdeutigkeiten [280] aufgrund der Wahl der Gesten zu minimieren oder einen maximalen Komfort [281] beim Gestendesign zu erreichen. Dazu muss jedoch zunächst der potentielle Nutzerkreis eingegrenzt werden. Obaid et al. unterteilen Probanden mit Erfahrungen im Umgang mit Robotern sowie gestischen Eingabesystemen in „technische Nutzer“ und Probanden ohne solche Erfahrung in „nichttechnische Nutzer“ [282, S. 383]. Um nichttechnischen Nutzern das Erlernen der Systembedienung zu erleichtern, wird oft ein möglichst intuitives als auch natürliches Gestenvokabular angestrebt. Metaphern können dabei helfen, dieses Ziel zu erreichen.

Metaphern und mentale Modelle

Im Laufe ihres Lebens sammeln Menschen Erfahrungen nicht nur im Umgang mit anderen Menschen, sondern auch bei der Interaktion mit technischen Systemen und entwickeln daraus eigene geistige Vorstellungen darüber, nach welchen Prinzipen diese funktionieren und wie sie bedient werden können. Diese Vorstellungen werden allgemein als *mentale Modelle* bezeichnet und können Menschen dabei helfen, den Nutzen eines Systems zu erkennen und Vorhersagen zu den Auswirkungen von Eingaben auf den zukünftigen Systemzustand treffen zu können [283, S. 7].

Peshkova et al. vertreten die These, dass ein konsistentes Gestenvokabular, bei dem alle Gesten auf derselben Analogie oder Metapher aus dem Alltag basieren, ein entsprechendes mentales Modell beim Nutzer ansprechen kann und ihm somit erleichtert wird, bereits erworbenes Wissen auf eine neue Interaktionssituation – wie beispielsweise mit einem UAV – zu übertragen und sogar unbekannte Gesten selbst herzuleiten. Ein Gestenvokabular wäre somit intuitiv, wenn ein einziger Hinweis genügt, um einen ganzen Satz an kohärenten Gesten einer Metapher zu definieren [284, S. 35]. Basierend auf mehreren explorativen Probandenstudien, bei denen nichttechnische Probanden gebeten wurden, eigene Gesten für die Steuerung von UAVs zu entwickeln, identifizierten die Autoren drei Klassen von zugrunde liegenden mentalen Modellen. Bei *„imitativen“* Modellen repräsentiert der Körper oder einzelne Teile davon (z.B. Hand oder Kopf) das Fluggerät, sodass Rotations- und Positionsveränderungen in eine entsprechende Bewegung des UAVs direkt übersetzt werden. Gesten, bei denen das UAV über einen imaginären Stellvertreter (z.B. Joystick, Marionette) gesteuert wird, gehören zur *„instrumentierten“* Klasse. Beim dritten, von der Mehrzahl der Probanden bevorzugten, mentalen Modell wird das UAV als *„intelligentes Wesen“* [284, S. 38] angesehen und ist somit auch in der Lage, High-Level Anweisungen (z.B. Zeigegesten oder „Folge-mir“ Bewegungen) zu interpretieren. Dieses mentale Modell erlaubt eine natürliche Form der Interaktion, weil auch Menschen untereinander diese Art von Kommunikation nutzen.

3.3.2 Gestendatensätze

Die Literatur enthält viele öffentliche Datensätze, die gestikulierende Menschen in unterschiedlichen Alltagssituationen zeigen. Die meisten davon wurden jedoch primär für die Entwicklung von Erkennungsverfahren von Hand- und Körpergesten oder Handlungen im Bereich der HCI entworfen. Insofern eignen sich diese Datensätze nur bedingt für Anwendungen in der cHDI, da eine oder mehrere der nachfolgenden Anforderungen nicht erfüllt werden:

- Personen sind in unterschiedlichen Abbildungsgrößen dargestellt
- Die verwendete Auflösung erlaubt die Differenzierung einzelner Finger und deren Stellung
- Die Aufnahmen zeigen den ganzen Körper oder zumindest den Oberkörper
- Die Aufnahmen wurden aus einer erhöhten Kameraperspektive außerhalb von Gebäuden aufgenommen
- Die Aufnahmen enthalten auch die Eigenbewegung der Plattform
- Der Bildhintergrund ist komplex und enthält diverse Objekte
- Der Blick der Personen ist in Richtung Kamera gerichtet
- Die dargestellten Körpersignale sind intentional und wurden mit unterschiedlichen Ausführungsgeschwindigkeiten aufgezeichnet

Zwar wurden in der Vergangenheit auch Datensätze für die cHDI entwickelt, aber selbst da finden sich nur einige wenige, öffentlich verfügbare Videodatensätze, die die genannten Anforderungen gänzlich erfüllen. Insbesondere Datensätze, die aus einer erhöhten Perspektive in einer freien Umgebung außerhalb von Gebäuden erzeugt wurden, sind rar und beschränken sich aktuell auf einige wenige. Eines davon ist der „UAV-GESTURE“ Datensatz [285], welcher gestikulierende Einzelpersonen auf einem Feldweg zeigt und 13 Handsignale aus dem NATOPS-Handzeichensystem [120] enthält.

Daneben stellten Patrone et al. einen Gestendatensatz unter dem Namen „AUTH UAV Gesture Language“ zusammen, der aus elf NATOPS Einweisersignalen und vier UAV-spezifischen Kommandos zur Bedienung der Kamera besteht [197]. Zwar suggeriert der Name, eine eigene Sprache zu enthalten, genau genommen handelt es sich jedoch lediglich um ein Gestenvokabular aus 15 Geste-Kommando-Paaren, welche singuläre Gesten auf festgelegte UAV-Funktionen direkt abbilden. Für eine Sprache fehlen jedoch weitere grammatikalische Merkmale, wie eine Syntax. Tab. 6 zeigt eine Übersicht der öffentlich verfügbaren Datensätze, die für Anwendungen in der cHDI entwickelt wurden.

Tab. 6 Literaturübersicht geeigneter Datensätze zur nonverbalen cHDI

Datensatz	Klassen	Personen	Auflösung [Px]	Bildwiederholrate [Hz]	Umgebung		Perspektive		Kamera		Bemerkung
					Indoor	Outdoor	Boden	Luft	Statisch	Dynamisch	
2004											
Schuldt et al. [286]	6	25	160x120	25	●	●	●	-	●	-	2D
2009											
Lin et al. [287]	14	3	640x480	25	●	-	●	-	●	●	Keck, 2D
2011											
Song et al. [288]	24	20	320x240	20	●	-	●	-	●	-	NATOPS, 3D
2012											
Guyon et al. [289]	8-12	20	240x320	10	●	-	●	-	●	-	ChaLearn, 2D/3D
2019											
Perera et al. [285]	13	10	1920x1080	25	-	●	-	●	-	●	UAV-GESTURE, 2D
2021											
Patrona et al. [197]	15	58	1920x1080	30	-	●	●	-	●	-	AUTH, 2D

●: relevant; -: nicht relevant; n/a: nicht angegeben

3.4 Übermittlung von Systemantworten über einen visuellen Rückkanal

Ohne einen Einblick in die internen Vorgänge sind die Ziele und Intentionen von intelligenten und mobilen, technischen Systemen für Außenstehende meist nur schwer abschätzbar und können von Unbehagen bis zu angstähnlichen Zustände führen [2, S. 78]. Auf die Frage, wann sie eine Rückmeldung von einem UAV während der Interaktion auf dem Flugdeck erwarten, hoben hauptberufliche Aircraft Director der U.S. Navy zwei Situationen hervor: zum einen sollte das System signalisieren, wenn jemand als Einweiser erkannt wurde und zum anderen, wenn ein neues Kommando registriert wurde [51, S. 27]. In der Literatur gibt es gleichlautende Forderungen auch von Laienanwendern [156, S. 364, 173, S. 1744]. Welche Möglichkeiten stehen dafür also zur Verfügung? Wie bereits in Kapitel 3.1.2 erläutert, scheiden akustische Verfahren beim gegenwärtigen Auftriebserzeugungsverfahren und den daraus resultierenden hohen Schallpegeln für die kollokierte Interaktion mit UAVs aus. Daher beschränkt sich die nachfolgende Literaturübersicht auf visuell wahrnehmbare, implizite (Flugverhalten) als auch explizite Verfahren (Lichtsignale) zur Kommunikation.

3.4.1 Flugdynamische Kommunikationsverfahren

Die Veränderung des Flugverhaltens von UAVs in kommunikativer Absicht wurde sowohl für große [216] als auch für kleine Interaktionsdistanzen untersucht [290]. Abhängig von der aerodynamischen Konfiguration der fliegenden Plattform (z.B. Luftschiff, Starrflügler, Drehflügler) stehen unterschiedliche Flugmanöver zur Verfügung. Allgemein besitzen Fluggeräte sechs Freiheitsgrade, wodurch kommunikative Signale in Bewegungen (vorwärts/rückwärts, links/rechts, hoch/runter), Rotationen (Rollen um Längsachse, Nicken um Querachse und Gieren um Hochachse) oder Geschwindigkeitsvektoren bzw. Drehraten kodiert werden können. Szafir et al. modifizierten Trajektorien, um die Vorhersagbarkeit des UAV-Flugpfades für einen Beobachter zu verbessern [291], während Colley et al. die Flugcharakteristik modifizierten, um navigatorischen Anweisungen zu übermitteln [210]. Jensen et al. bildeten zwischenmenschliche Begrüßungssignale in Form von nickenden und wackelnden „Drohnen-Gesten“ nach und untersuchten deren Sichtbarkeit in der Mensch-UAV-Interaktion [292]. Die teils sehr subtilen Bewegungen wurden jedoch nicht von allen Probanden wahrgenommen, weil sie die derartigen kommunikativen Manöver nicht von den Flugstabilisierungsbewegungen unterscheiden konnten.

Auch stellten die Autoren fest, dass die UAV-Front nicht immer als solche erkennbar war und Probanden dadurch nicht wussten, wohin der „Blick“ des UAVs gerichtet war.

Einfluss der Flugcharakteristik auf Wahrnehmung

Müller und Muirhead stellten bei Untersuchungen mit ihrem „Joggobot“ [20] heraus, dass minimale Abweichungen von der festgelegten Flugroute aufgrund von aerodynamischen und regelungstechnischen Unzulänglichkeiten dazu beitragen können, dem UAV einen „Charakter“ zu verleihen. Ihrer Ansicht nach könnten diese sogenannten *„secondary movements“* [20, S. 2025] Nutzern erlauben, das UAV nicht nur als eine technische Maschine, sondern als „Buddy“ wahrzunehmen und begünstigen damit eine gewisse soziale Verbundenheit [20, S. 2029-2030]. Der Umfang dieser zusätzlichen Flugbewegungen sollte jedoch mit Bedacht gewählt werden. Zu starke und abrupte Veränderungen von Geschwindigkeiten und Drehraten können nicht nur negative Auswirkungen auf die möglicherweise sensible Nutzlast haben, sondern auch von Menschen als „bedrohliches, aggressives oder unberechenbares Verhalten“ [292, S. 10] gedeutet werden. Die Kodierung von Emotionen in das Flugverhalten ist daher ein wichtiger Forschungsbereich in der cHDI. Cauchard et al. [293] leiteten auf der Grundlage von sieben stereotypischen Persönlichkeiten exemplarisch drei Persönlichkeitsmodelle für UAVs her (*Exhausted Drone*, *Anti-Social Drone* sowie *Adventure Hero Drone*) und bildeten deren Charaktermerkmale auf physikalische Parameter[43] ab. In einer anschließenden Studie konnten sie nachweisen, dass der Großteil der Probanden den emotionalen Zustand und damit die Persönlichkeit nur anhand der Flugcharakteristik korrekt deuten konnte.

Einschränkungen

Neben den Manöver-Einschränkungen, die aus dem Auftrieberzeugungsverfahren resultieren, kann auch der konkrete Anwendungsfall zusätzliche Limitierungen erfordern, was die Sichtbarkeit und Eindeutigkeit der flugdynamischen Signale weiter degradieren kann. Insbesondere Bewegungen um die Nick- und Rollachse führen bei Multikoptersystemen immer zu einer Änderung der Position, was nicht immer gewünscht sein kann, beispielsweise dann, wenn räumliche Begrenzungen einen hochpräzisen Schwebeflug erfordern. Für die Kommunikation von perzeptiven Zuständen (Sieht mich das UAV?) tendieren die Empfehlungen daher zu lichtbasierten Signalisierungsmethoden [292, S. 8].

3.4.2 Lichtbasierte Kommunikationsverfahren

Die zunehmende Komplexität von unbemannten Fluggeräten verlangt auch eine von außen erkennbare Rückmeldung über den momentan aktiven Betriebsmodus. Hersteller von UAVs aller Größenordnungen statten ihre Systeme daher mit einzelnen oder als Matrix angeordneten LEDs an markanten Positionen, wie Bugrad oder Flügelenden, aus, um über eine Farbkodierung den aktuellen Status sichtbar zu machen (Abb. 34). Die Bandbreite an Informationen, die über solche Punktlichtquellen übertragen werden kann, ist natürlich sehr begrenzt und die Bedeutung der entsprechenden Farbe oder des Leuchtmusters muss beim Anwender bekannt oder zumindest erschließbar sein, um die ausgestrahlten Informationen korrekt interpretieren zu können.

[43] Relative Position des UAVs zum Probanden, Geschwindigkeit, Drehraten, Flughöhe, Reaktionszeit und Bereitschaft, menschliche Anweisungen zu befolgen.

Abb. 34 Farbkodierte Betriebsindikatoren (rote Markierungen) eines kleinen Hexakopters (links) sowie eines UAVs der HALE-Kategorie (rechts), rechte Abbildung übernommen aus [55]

Szafir et al. experimentierten mit einem einzeiligen, umlaufenden LED-Ring, um eine bevorstehende Richtungsänderungen des UAVs an umstehende Beobachter zu kommunizieren [294]. Von den vier implementierten Anzeigemetaphern (Abb. 35) konnte die Blinker- und Schubdüsenmetapher die Vorhersagbarkeit der Flugbewegung in exozentrischen Betrachtungssituationen[44] verbessern, während die Blickmetapher (*Gaze*) insbesondere bei egozentrischen Betrachtungssituationen[45] UAV-Intentionen signifikant erhöhen konnte (im Vergleich zu Flugbewegungen ohne zusätzliche Signalisierung).

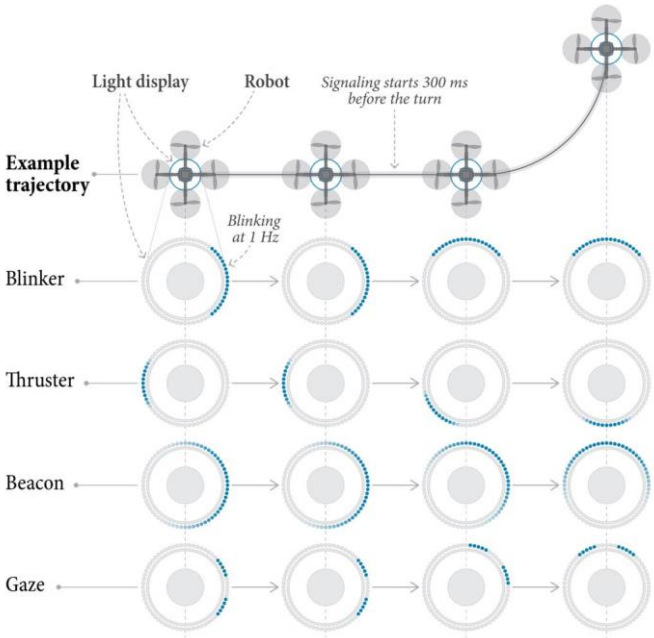


Abb. 35 Visualisierung von vier Anzeigemetaphern zur Richtungsindikation mittels eines umlaufenden LED-Rings, übernommen aus [294, S. 21]

Mehrzeilige Lichtquellen (Abb. 34 rechts) erlauben darüber hinaus auch die Darstellung von abstrakten wie konkreten Symbolen, textueller Information bis hin zu statischen und dynamischen Abbildungen. Weitaus mehr Möglichkeiten bieten hochaufgelöste LCD- und E-Paper-Displays [295, 296]. Insbesondere auf kleinen Flugplattformen beeinflussen derlei schwere Anbauten jedoch die Schwerpunktlage und verkürzen die maximale Flugdauer aufgrund des zusätzlichen Gewichts und Strombedarfs. Auch muss beim Interaktionsdesign beachtet werden, dass Adressaten Informationen aufgrund

[44] Aus Sicht eines außenstehenden Beobachters.
[45] Aus Sicht der Probanden.

der bauartbedingten Blickwinkelabhängigkeit und möglicher Spiegelungen übersehen können [294, S. 20].

Zur optischen Aufwertung von Flugmodellen, aber auch zur Übermittlung von Mitteilungen, werden im Modellbau spezielle LED-Rotorblätter verwendet. Diese Rotorblätter sind auf der Unterseite mit LEDs bestückt und besitzen in der Propellernabe einen Mikrocontroller, der von einem kleinen Akku gespeist wird. Abhängig von der aktuellen Rotordrehzahl, erzeugt dieser zeitlich versetzte Steuersignale für jede einzelne LED, sodass vom Boden aus der Eindruck einer stehenden zweidimensionalen Abbildung entsteht (Abb. 36).

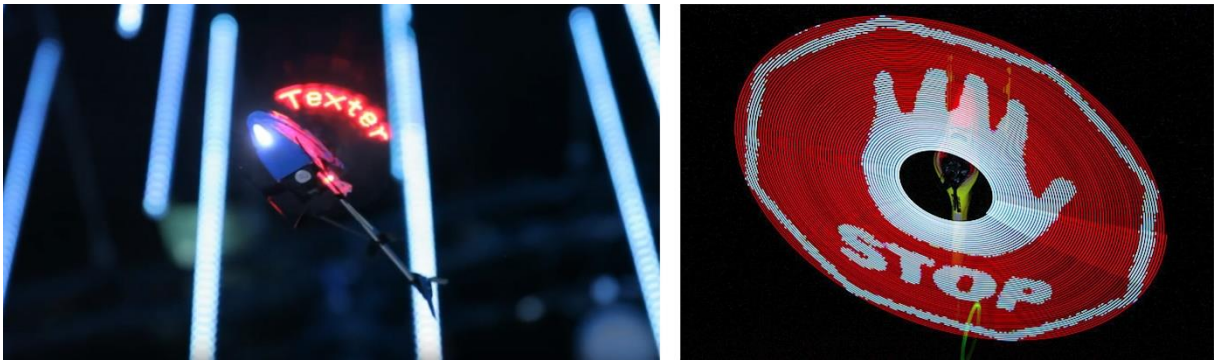


Abb. 36 Darstellung von Text und Grafiken mittels rotierender LED-Propeller, übernommen aus [297, 298]

Dieses Prinzip lässt sich auch auf Multirotorsysteme übertragen, es ist jedoch eine zusätzliche Synchronisation notwendig, weil die Rotoren mit unterschiedlichen Drehzahlen arbeiten.

Projektionen

Die Projektion von interaktiven Schaltflächen [217] oder Karten [299] auf vorhandene (Abb. 37) oder mitgeführte (in-situ) Projektionsflächen [300] stellen weitere Möglichkeiten für einen visuellen Rückkanal dar. Die großformatige Projektion erlaubt die Anwendung bekannter Interaktionsmuster aus dem Umgang mit berührungsempfindlichen Oberflächen wie Smartphones und Tablets. Aufgrund der geringen, erreichbaren Lichtstärken eignen sich derartige Ausgabegeräte jedoch nur für schwach beleuchtete Umgebungen.

© 2019 IEEE

Abb. 37 Auf den Boden projizierte Schaltflächen zur gestischen Interaktion mit einem schwebenden UAV, übernommen aus [217, S. 153]

Räumliche Displays

Räumliche Displays bezeichnen einen Schwarm aus räumlich koordinierten und zeitlich synchronisierten LED-bestückten UAVs, die von einer zentralen Einheit gesteuert, dreidimensionale Animationen und Grafiken in den Himmel zeichnen [301]. Abhängig von der Anzahl der dabei eingesetzten

Fluggeräte[46] können überaus großformatige Darstellungen erreicht werden. Mit steigender Anzahl an UAVs erhöht sich jedoch auch der Koordinationsaufwand und die Gefahr von Kollisionen. Räumliche Displays werden bevorzugt bei Großveranstaltungen zu Unterhaltungszwecken genutzt. Die Verwendung dieser Technologie in kommunikativer Absicht wurde bislang in der kollokierten HDI nicht untersucht. Tab. 7 zeigt alle vorgestellten Verfahren in der Übersicht.

Tab. 7 Literaturübersicht visueller Rückmeldeverfahren für dein Einsatz auf UAVs

Verfahren	Einsatz	Sichtbarkeit bei großen Distanzen	Kommunikations-inhalte	Vorteile	Nachteile
Flugmanöver [292]	Tag	Abhängig von Plattformgröße	Sehr eingeschränkt	Keine zusätzlichen Geräte	Abhängig von Art der Auftriebserzeugung
LED-Leiste [294]	Tag und Nacht	Mittel	Eingeschränkt	Geringes Gewicht	Kodierung muss bekannt sein
LED-Matrix [202]	Tag und Nacht	Mittel	Eingeschränkt	Gleichzeitige Darstellung mehrerer Inhalte Geringes Gewicht	Separate Stromversorgung Niedrige Auflösung
LED-Propeller [298]	Nacht	Abhängig von Rotorgröße	Beliebig	Geringes Gewicht	Synchronisation bei Multirotorsystemen notwendig
LCD-Display [296]	Tag und Nacht	Gering	Beliebig	Hohe Auflösung	Hohes Gewicht Separate Stromversorgung
Projektionen auf vorhandene Flächen [299]	Nacht	Gering	Beliebig	Hohe Auflösung	Leuchtstärke der Projektion nimmt mit Entfernung stark ab Darstellungsqualität abhängig von Projektionsfläche
Projektion (in-situ) [300]	Nacht	Gering	Beliebig	Hohe Auflösung	Zusätzliches Gewicht durch mitgeführte Projektionsfläche Eingeschränkter Sichtbereich
Räumliches Display [301]	Nacht	Hoch	Abhängig von Anzahl an UAVs	Großflächige Darstellung	Viele UAVs notwendig Hoher Koordinationsaufwand

3.5 Research Gaps und Forschungsfragen

Im Grundlagenkapitel dieser Arbeit wurden die vielfältigen Einsatz- und Kommunikationsmöglichkeiten der visuellen und insbesondere der handzeichenbasierten interpersonellen Verständigung in der Luft- und Raumfahrt aufgezeigt und die Signifikanz dieses alternativen Kommunikationsverfahrens bei Ausfall oder der Nichtverfügbarkeit funkbasierter Schnittstellen verdeutlicht. Das gestiegene Interesse an unbemannten Flugsystemen verlangt nach einer geeigneten UAV-Schnittstelle, die ähnliche Funktionalitäten bereitstellen kann, um Kompatibilität zu ermöglichen und die sichere und nahtlose Integration von UAVs in bestehende menschenzentrierte Verfahren und Abläufe zu erlauben.

Der Blick in den aktuellen Stand der Forschung hat verdeutlicht, dass im Bereich der cHDI verschiedenste Anstrengungen unternommen werden, um eine natürliche, direkte Mensch-UAV-Interaktion ohne den Einsatz zusätzlicher Kommunikationsgeräte zu ermöglichen. Die nonverbale Interaktion über eine visuelle Schnittstelle bietet dabei im Vergleich zu anderen Modalitäten die meisten Vorteile. Dennoch wird das Potential dieser Interaktionsform nicht vollständig ausgeschöpft, sodass die nonverbalen Kommunikationsfähigkeiten aktueller UAVs sich auf einem rudimentären Niveau befinden und auf die Erkennung singulärer Low- und High-Level-Befehle beschränken, die meist direkt auf Systemfunktionen abgebildet werden. Daher wird aktuell, wie die Benennung des Teilgebietes

[46] Der aktuelle Weltrekord aus 2021 liegt bei 5164 simultan fliegenden UAVs [302].

(cHD**I**) es vermuten lässt, lediglich der *Interaktionsaspekt* und nicht der *Kommunikationsaspekt* behandelt. Der wechselseitige, zielgerichtete Informationsaustausch stellt jedoch die Voraussetzung für eine Kommunikation dar. Das Ziel dieser Arbeit ist somit, einen Beitrag zum Ausbau dieser kommunikativen Fähigkeiten zu leisten.

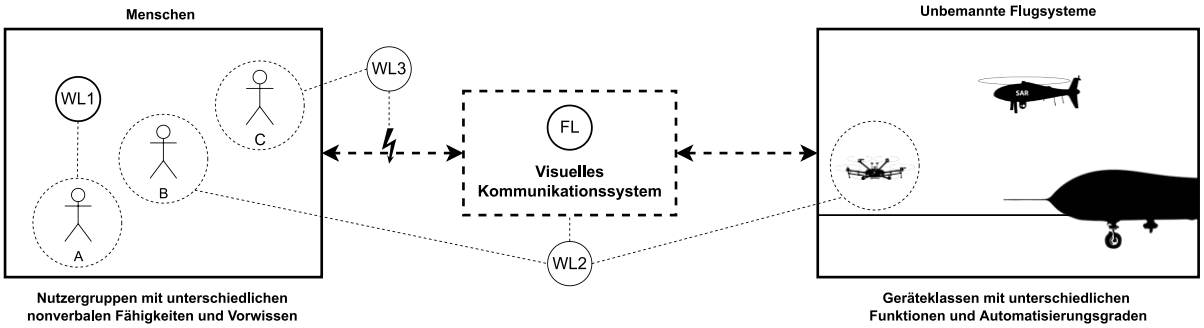


Abb. 38 Die identifizierten Forschungslücken im Überblick (FL = Fähigkeitslücke, WL = Wissenslücke)

Bei einer genaueren Betrachtung des gegenwärtigen Forschungsstandes zeichnet sich eine Art „Henne-Ei-Problem“ ab: Einerseits besteht der Wunsch nach einem tieferen Verständnis der kommunikativen Vorgänge während einer nonverbalen Mensch-UAV-Interaktion, um die notwendigen Systemfunktionen identifizieren zu können, andererseits fehlt es jedoch an entsprechenden technischen Möglichkeiten, die einen geeigneten Rahmen schaffen und grundlegende Funktionalitäten für die Evaluierung bereitstellen. Um das Problemfeld besser beschreiben zu können wird der identifizierte Forschungsbedarf dieser Arbeit in Fähigkeitslücken und Wissenslücken unterteilt (Abb. 38), welche von Hypothesen begleitet werden.

3.5.1 Fähigkeitslücke 1: Visuelles Kommunikationssystem zur Intentionserkennung

Für eine möglichst realitätsnahe Untersuchung der nonverbalen Mensch-UAV-Verständigung wird zunächst ein Konzept für ein technisches Experimentalsystem benötigt, welches erweiterte Funktionalitäten bereitstellen kann, die konventionelle Gestenerkennungssysteme nicht bieten. Hierzu gehört unter anderem die Fähigkeit, aus einer Menge von nonverbalen Eingaben, das Kommunikationsziel des Nutzers, d.h. dessen Intention zu bestimmen. Statt einer klassischen Gestenerkennung wird daher eine *Intentionserkennung* benötigt. Aus forschungsökonomischen Gesichtspunkten sollte der Entwurf für ein derartiges System so generisch wie möglich gehalten werden, um unterschiedliche Nutzungssituationen und Kommunikationsabsichten abbilden zu können. Das erste Forschungsziel ist daher die konzeptionelle Entwicklung und Implementierung eines geeigneten Systementwurfs.

Hypothese 1:

Das entwickelte und implementiere Systemkonzept ermöglicht die Untersuchung verschiedener Nutzungssituationen und Kommunikationsabsichten.

3.5.2 Fähigkeitslücke 2: Erkennung generischer Nutzereingaben

Um einen großen Nutzerkreis ansprechen zu können, sollte eine visuelle Kommunikationsschnittstelle in der Lage sein, nonverbale Eingaben unterschiedlicher Nutzergruppen erfassen und auswerten zu können. Eine besondere Nutzergruppe repräsentieren hierbei Menschen, die bisher noch keine Berührungspunkte mit einer nonverbalen Mensch-Maschine-Schnittstelle hatten. Zwar setzt diese Gruppe Gestik in der täglichen Kommunikation mit anderen Menschen ein, jedoch erfolgt dies primär in un-

terstützender Funktion zum Gesagten. Wenn nur noch der nonverbale Kommunikationsweg zur Verfügung steht, muss damit gerechnet werden, dass eine breite Palette von generischen Signalisierungsmustern zum Einsatz kommt. Im Systemkonzept wird daher auch eine Unterstützung für solche generische Nutzereingaben benötigt, um eine rudimentäre Form der Kommunikation auch mit Nutzern ohne Systemerfahrung zu erlauben.

Hypothese 2:

Die im Systemkonzept vorgesehenen Verfahren erlauben die Erfassung und Erkennung generischer Körpersignale, um auch eine rudimentäre Kommunikation mit Nutzern ohne Systemerfahrung zu ermöglichen.

3.5.3 Wissenslücke 1: Nonverbale Kommunikationssignale von Nutzern ohne Systemerfahrung im Kontext einer simulierten Notsituation

Um die Wirksamkeit dieser Verfahren überprüfen zu können, werden entsprechende Daten benötigt. Im Kontext einer simulierten Notsituation mit Nutzern ohne Systemerfahrung soll daher untersucht werden, welche nonverbalen Signale Menschen in derartigen Situationen tatsächlich zur Kommunikation verwenden und ob sich die angedachten Verfahren zur Erkennung dieser eignen. Insbesondere sollen hierbei die verwendeten Kommunikationsmittel zur Erreichung der folgenden rudimentären Kommunikationsziele untersucht werden:

- Mitteilung eines Interaktionswunsches (WL 1.1),
- Mitteilung von Zustimmung/Ablehnung (WL 1.2) sowie
- Mitteilung eines körperlichen Problems (WL 1.3).

3.5.4 Wissenslücke 2: Kollokierte UAV-Führung über eine visuelle Schnittstelle

Die Einweisung von Luftfahrzeugbesatzungen bei Roll- oder Schwebevorgängen durch visuelle Handzeichen stellt in der Luftfahrt ein Standardverfahren dar, sodass auch von zukünftigen unbemannten Systemen erwartet wird, dass sie dieses unterstützen. Die bisherigen Untersuchung solcher Interaktionsszenarien wurden ohne flugfähige Experimentalsysteme oder nur mit kleinen, ferngesteuerten oder teilautomatisierten Fluggeräten durchgeführt [165, S. 4]. Insofern werden in der Literatur realitätsnahe Experimente mit vollintegrierten Fluggeräten gefordert [212, S. 9]. Im Kontext einer nonverbalen Einweisung eines hochautomatisierten Multikopters aus der 15 kg Klasse soll die Mensch-UAV-Interaktion über eine visuelle Schnittstelle im Realflug untersucht und Antworten auf folgende Fragestellungen gefunden werden:

- Welches nonverbale Flugführungsverfahren eignet sich für die sichere Fernführung eines Multikoptersystems über eine visuelle Schnittstelle (WL 2.1)?
- Welches Raumnutzungsverhalten zeigen Menschen während der kollokierten Interaktion mit einem hochautomatisierten, schweren Multikopter (WL 2.2)?
- Welchen Einfluss hat dabei die Erfahrung des Nutzers im Umgang mit UAVs auf die Interaktionsdistanz (WL 2.3)?

3.5.5 Wissenslücke 3: Verständigungsprobleme bei nonverbalen Dialogen mit technischen Systemen

In der alltäglichen Interaktion weichen Menschen manchmal kurzzeitig auf eine nonverbale Kommunikationsform aus. Ein zustimmendes Nicken, eine Daumen-Hoch-Geste oder eine herbeiholende

Handbewegung stellen singuläre „Schlüsselwörter“ dar, die dem Kommunikationspartner in der jeweiligen Situation den entscheidenden Input geben. Wenn jedoch höherwertige Kommunikationsziele an ein technisches System übermittelt werden müssen, beispielsweise eine Auftragsbeschreibung, werden höhere Anforderungen an die manuellen und nonmanuellen Ausdrucksfähigkeiten und die Awareness über den aktuellen Dialogfortschritt der Nutzer gestellt. Bei dieser Form des *nonverbalen Dialogs* kann auf bekannte Bestätigungsindikatoren (wie eine verständnisvolle Mimik) nicht zurückgegriffen werden, weil sie in dieser Form vor allem bei fliegenden Systemen nicht verfügbar sind. Darüber hinaus wird ein geeignetes Vokabular benötigt. Hilfsmittel aus anderen Handzeichensystemen (z.B. Gebärdensprachen) wie Metaphern, Embleme und Synonyme können dabei helfen, den Umfang der möglichen Kommunikationsinhalte zu vergrößern und gleichzeitig das dafür notwendige Vokabular klein zu halten. Mit der Verwendung von mehrdeutigen Vokabeln erhöht sich jedoch die Gefahr von Missverständnissen. Die dritte Wissenslücke betrifft somit das Problemlösungsverhalten von Menschen bei der nonverbalen Kommunikation mit einem technischen System. Hierzu sollen folgende Detailfragen beantwortet werden:

- Welche Faktoren begünstigen Verständigungsprobleme bei einem nonverbalen Dialog mit einem technischen System (WL 3.1)?
- Welche nonverbalen Lösungsstrategien wenden Menschen an, wenn es Probleme bei der Verständigung über eine visuelle Schnittstelle gibt (WL 3.2)?
- Wo liegt die Frustrationsgrenze von Menschen bei einem nonverbalen Dialog bei anhaltenden Verständigungsproblemen (WL 3.3)?

Erkenntnisse hierüber können Hinweise darauf geben, wann ein technisches System mit weiteren Unterstützungsfunktionen eingreifen muss, um einen negativen Dialogverlauf zu verhindern.

3.5.6 Fähigkeitslücke 3: Dialogunterstützungsfunktionen zur Verbesserung der nonverbalen Mensch-UAV-Kommunikation

Um die Gefahr von Missverständnissen während der Kommunikation zu minimieren werden daher geeignete *Dialogunterstützungsfunktionen* benötigt, die einerseits dem Systemnutzer eine Rückmeldung über den gegenwärtigen Dialogfortschritt erlauben und andererseits das technische System in die Lage versetzen, zwischen relevanten und irrelevanten Körpersignalen im jeweiligen situativen und inhaltlichen Kontext unterscheiden zu können. Daher soll das Systemkonzept entsprechende Unterstützungsfunktionen enthalten, deren Nutzen in Mensch-Maschine-Experimenten überprüft werden muss.

Hypothese 3:

Die im Systemkonzept vorgeschlagenen Dialogunterstützungsfunktionen helfen Nutzern während eines nonverbalen Dialogs mit einem technischen System bei der Erreichung des Kommunikationsziels.

Das nachfolgende Kapitel 4 stellt das Konzept für ein geeignetes visuelles Mensch-UAV-Kommunikationssystem vor, welches die identifizierten Fähigkeitslücken adressieren kann. Die Umsetzung dieses Konzepts in ein prototypisches Experimentalsystem ist in Kapitel 5 dokumentiert, während dessen Funktionsnachweis und die experimentelle Beantwortung der Wissenslücken in Kapitel 6 erfolgt. Abb. 39 verdeutlicht graphisch die thematische Positionierung dieser Arbeit innerhalb der cHDI.

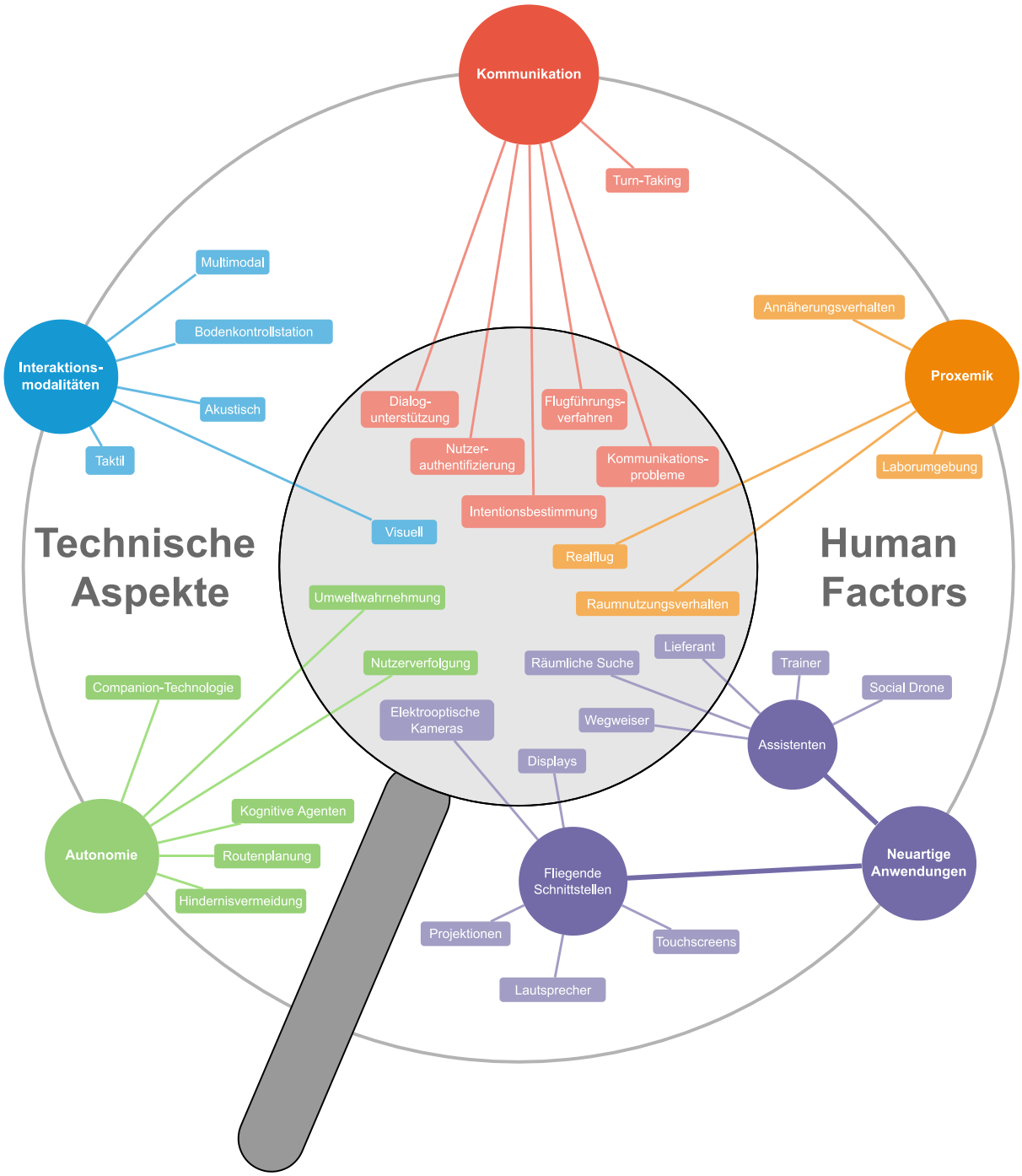


Abb. 39 Die thematische Positionierung dieser Arbeit im Kontext der *collocated Human-Drone Interaction*

4 Konzept für ein visuelles Mensch-UAV-Kommunikationssystem

Dieses Kapitel beschreibt das grundlegende Konzept für ein visuelles Mensch-UAV-Kommunikationssystem, welches einerseits die identifizierte Fähigkeitslücke 1 adressiert und zum anderen die Grundlage für die Umsetzung in ein prototypisches Experimentalsystem zur Untersuchung der nonverbalen Mensch-UAV-Interaktion darstellt. Nonverbal bedeutet hierbei, dass Informationen mithilfe von bewusst eingesetzten Körperbewegungen an ein UAV übermittelt werden und auf ein konventionelles, funkbasiertes Kommunikationsgerät verzichtet wird. Die Erfassung dieser Bewegungen soll vom UAV aus im für den Menschen sichtbaren Teil des elektromagnetischen Spektrums erfolgen. Um eine dialogähnliche Konversation zu ermöglichen, soll die Kommunikationsrichtung bidirektional sein, sodass der Informationsaustausch nicht nur vom Menschen zum UAV, sondern auch in umgekehrter Richtung möglich ist. Im Folgenden wird daher zunächst eine Eingrenzung vorgenommen, in welchen Flugphasen und unter welchen Annahmen das Systemkonzept mit dem Namen *VisCom*[47] Verwendung finden soll. Anschließend werden die für die Beschreibung des Konzepts genutzten Begrifflichkeiten definiert, gefolgt von den allgemeinen Anforderungen an eine visuelle Mensch-UAV-Schnittstelle. Daran anschließend wir das vorgeschlagene Prinzip des UAVs als „Dienstleister“ als Grundlage für eine Intentionserkennung näher erörtert und Modelle zur Beschreibung des Nutzers und seinem Verhalten als auch des Systems vorgestellt. Der letzte Teil dieses Kapitels geht auf die entwickelte Verkehrssprache zur Kommunikation zwischen Mensch und UAV näher ein und präsentiert ein Verfahren zur kontextsensitiven Signalübersetzung und Intentions- bzw. Diensterkennung.

4.1 Annahmen und Umgebungsbedingungen

Bei konventionellen UAS kommuniziert der Nutzer mit dem UAV mittels einer digitalen, drahtlosen Datenverbindung. Bei ausreichender Sende- und Empfangsleistung des Kommunikationssystems kann die Datenübermittlung auch unter BVLOS Bedingungen erfolgen, sodass beide Interaktionspartner örtlich getrennt miteinander kommunizieren können. Dies stellt insbesondere im militärischen Anwendungskontext den Normalfall dar. Für eine visuelle Kommunikation ist hingegen eine stabile, direkte Sichtverbindung zwingend erforderlich, welche beispielsweise durch eine schwebefähige Plattform und/oder ein kardanisch stabilisiertes Gimbalsystem sichergestellt werden kann. Für das VisCom-Konzept sollen daher einige Annahmen und Eingrenzungen gelten.

Das Missionsprofil eines UAV ist immer abhängig vom jeweiligen Einsatzzweck und gliedert sich in unterschiedliche operationelle Flugphasen, die an bestimmte Flughöhen gekoppelt sein können. Eine visuelle Kommunikation kann nur in solchen Flugphasen stattfinden, in denen zwischen Mensch und UAV eine hindernisfreie, direkte Sichtverbindung besteht und eine, von der Abbildungsleistung der jeweiligen optischen Systeme abhängige, Maximalentfernung nicht überschritten wird (*Interaktionsbereich*). Diese Bedingungen sind vorranging im An- und Abflug des UAVs anzutreffen, d.h. in bodennahen Flughöhen, können aber auch im Verlauf der Mission auftreten, wenn das UAV mit Dritten, d.h. nicht in die Kontrollschleife involvierten Nutzern, in Kontakt treten soll (beispielsweise bei der automatisierten Suche nach Vermissten).

[47] Akronym für **Vis**ual **Com**munication.

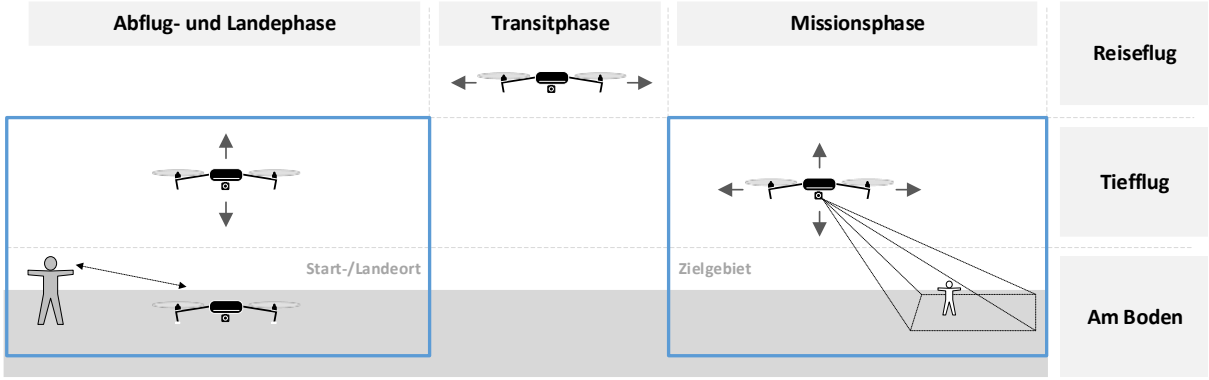


Abb. 40 Relevante Flugphasen für eine visuelle Interaktion (blaue Markierung)

Grundsätzlich können visuelle Signale auch mit zusätzlichen Hilfsmitteln (Flaggen, Leuchtstäbe, Taschenlampen, geschriebene Informationen auf Papier, Zeichnungen auf den Boden) übermittelt werden. In diesem Konzept wird jedoch der Fall behandelt, dass keine weiteren Gegenstände vorhanden sind und dem Nutzer nur der eigene Körper als Kommunikationsmittel zur Verfügung steht. Während der Interaktion befindet sich das UAV am Boden oder schwebt in dessen Nähe, wohingegen sich der Nutzer im Interaktionsbereich des UAVs aufhält und diesem frontal zugewandt steht. Der Nutzer verfolgt immer nur ein Kommunikationsziel gleichzeitig, kann dieses jedoch während der Interaktion ändern.

4.2 Definitionen

Abhängig vom Hintergrund des Lesers können die in dieser Arbeit verwendeten Begriffe mit unterschiedlichen Bedeutungen belegt sein. Daher sollen die nachfolgenden Definitionen eine gemeinsame Grundlage für die weiteren Kapitel bilden.

System

Mit System soll in dieser Arbeit nicht der Mensch-UAV-Verbund, sondern das technische Kommunikationssystem, bestehend aus elektrischen, mechanischen, optischen und informationsverarbeitenden Komponenten, bezeichnet werden, welches als Übersetzungsschnittstelle fungiert und eine nonverbale Interaktion von Mensch und UAV ermöglicht.

Nutzer

Eine Person, die mit dem UAV nonverbal interagieren möchte und über die notwendigen Zugriffsrechte verfügt.

Sichtkontakt

Ein Zustand, bei dem das System einen Menschen sensorisch erfasst und diesen als Nutzer identifiziert hat.

Blickkontakt

Ein Zustand, bei dem der Nutzer eine hindernisfreie Sichtverbindung zum System hergestellt hat und dessen Reaktionen beobachten kann.

Signal

Eine intentionale, sichtbare Änderung eines Körpermerkmals des Nutzers oder eine sichtbare Zustandsveränderung des UAV, welche eine kodierte Information enthält.

Beobachtung

Das Ergebnis der Klassifikation eines bestimmten Körpermerkmals des Nutzers.

Symbol

Ein dekodiertes Körpersignal, d.h. eine bedeutungstragende Information innerhalb eines definierten semantischen Kontextes.

Semantischer Kontext

Eine Menge von gültigen Symbolen während einer bestimmten Flug- und Dialogphase der Interaktion.

Fernführung

Ein kollaborativer Vorgang, bei dem das UAV vom Menschen bei der Durchführung einer Tätigkeit durch nonverbale Signale unterstützt wird.

Flugführungskommando

Ein nonverbal formulierter Flugführungswunsch des Nutzers, welcher vom UAV in eine Fluglageänderung umgesetzt wird.

Dienst

Eine Systemfunktion, bestehend aus informationsverarbeitenden und mechatronischen Komponenten, die vom UAV angeboten wird und vom Nutzer angefragt werden kann. Die Anzahl und Komplexität der verfügbaren Dienste orientiert sich am Automatisierungsgrad und der vorhandenen technischen Ausstattung des UAV.

Intention

Eine nonverbale Beschreibung eines UAV-Dienstes, welchen der Nutzer in Anspruch nehmen möchte.

Dialog

Basierend auf der Definition der ISO 9241-110 stellt ein Dialog in diesem Zusammenhang eine nonverbale „Interaktion zwischen einem Benutzer und einem interaktiven System in Form einer Folge von Handlungen des Benutzers (Eingaben) und Antworten des interaktiven Systems (Ausgaben), um ein Ziel zu erreichen“ [171, S. 6]. Das Ziel ist in diesem Fall die Übermittlung einer Intention in einem dialogähnlichen Vorgang.

Auftrag

Eine Aneinanderreihung mehrerer UAV-Dienste, welche sequentiell vom UAV ausgeführt werden.

Nonverbale Teleoperation

Die Bedienung der Nutzlast eines UAVs, beispielsweise eines befestigten Roboterarms, aus der Ferne mittels des eigenen Körpers. Dabei erfasst das UAV neben der Anordnung der Gelenke auch die Ausrichtung der Gliedmaßen des Nutzers und übersetzt diese in Steuerkommandos für die Motoren und Aktuatoren der Nutzlast.

4.3 Anforderungen an ein visuelles Mensch-UAV-Kommunikationssystem

Angelehnt an die allgemeinen Anforderungen an gestenbasierte Kommunikationssysteme [303, S. 62-64] werden in den nachfolgenden Abschnitten die besonderen Anforderungen an ein visuelles Mensch-UAV-Kommunikationssystem definiert.

4.3.1 Variabler Interaktionsbereich

Im Unterschied zu konventionellen, ortsfesten (Tastatur am Schreibtisch) oder mobilen, natürlichen Benutzerschnittstellen (Service-Roboter mit Spracheingabe), beschränkt sich die Interaktionsdistanz in der Mensch-UAV-Interaktion nicht auf wenige Meter, sondern kann darüber hinaus bis zu mehreren Hundert Metern reichen. Die minimale Interaktionsdistanz ist dabei primär durch die Größe des fliegenden Systems und dem einhergehenden sicheren Mindestabstand vorgegeben. Das Feldhandbuch des amerikanischen Heeres schreibt beispielsweise für die Einweisung eines Drehflüglers einen Sicherheitsabstand von mindestens 40 Metern vor [131, S. 46]. Die maximale Interaktionsdistanz hingegen ist durch die Sehleistung des Nutzers und der Leistungsfähigkeit der sensorischen Komponente des Systems beschränkt. Damit ist die Interaktionsdistanz mitunter der treibende Faktor beim Systemdesign und der Wahl geeigneter Komponenten. Spezialisierte Sensoren (z.B. thermische Sensoren oder Tiefensensoren) liefern zwar zusätzliche Daten im Vergleich zu einfachen elektrooptischen Bildsensoren (EO), erreichen jedoch typischerweise nicht dasselbe Auflösungsvermögen oder sind – je nach verwendeter Technologie – in ihrem Erfassungsbereich stark eingeschränkt. EO-Sensoren sind hingegen auf den meisten unbemannten Fluggeräten in mindestens einfacher Ausführung vorhanden und bieten dank variabler Brennweiten (Zoom) eine hohe Flexibilität bei gleichzeitig geringem Integrationsaufwand.

Das Systemkonzept muss somit in der Lage sein, unter realitätsnahen Bedingungen (unterschiedliche Witterung, Windbedingungen, wechselnde Beleuchtung) eine Interaktion mit einem Nutzer zu ermöglichen. Vibrationen im Antriebsstrang als auch die dreidimensionale Bewegung der fliegenden Plattform sind weitere Parameter, die bei traditionellen Benutzerschnittstellen nicht in dem gleichen Umfang zum Tragen kommen. Atmosphärische Effekte beeinträchtigen zusätzlich die Abbildungsleistung des optischen Systems, sodass neben einer automatischen Sensorparametrisierung (Belichtungs- und Blendenautomatik, Autofokus, automatischer Weißabgleich) auch eine Sensorstabilisierung und Sensornachführung in Längs- und Querrichtung benötigt wird, um den Nutzer im optimalen Sensorbereich halten und dessen Körpermerkmale mit der höchstmöglichen Auflösung abbilden zu können.

Orts- und Zeitauflösung des Sensors

Neben einer minimalen Ortsauflösung, welche durch die Interaktionsdistanz und den eingesetzten Merkmalextraktionsverfahren vorgegeben werden, wird auch eine ausreichende zeitliche Auflösung der Sensordaten benötigt. Aus empirischen Untersuchungen mit Gebärdenden ist bekannt, dass für eine visuelle Verständigung mittels Handzeichen über ein bildgebendes Medium die Untergrenze der Bildwiederholrate im Bereich von 10 bis 15 Hz vermutet werden kann und dass Menschen selbst bei einer enormen Reduktion der Auflösung auf nur 24x16 Pixeln noch 86 % der Kommunikationsinhalte verstehen können [304, S. 363].

4.3.2 Anpassungsfähigkeit an unterschiedliche Nutzergruppen

In Anlehnung an die Fähigkeitslücke 1 und im Hinblick auf den Einsatz des Kommunikationssystems in unterschiedlichen Nutzungskontexten (vgl. Mensch-Maschine-Experimente in Kapitel 6) und der

damit einhergehenden großen Bandbreite an potentiellen Nutzern, sollte das Systemkonzept eine Interaktion mit unterschiedlichen Nutzergruppen ermöglichen. Die konventionelle Klassifizierung eines Nutzers als Laie oder Experte bzw. nichttechnischer und technischer Nutzer scheint bei dieser Interaktionsform aus zweierlei Gründen nicht hinreichend genau zu sein. Zum einen beschreibt eine derartige binäre Klassifizierung nur zwei extreme Ausprägungen und erlaubt keine weitere graduelle Differenzierung. Zum anderen orientiert sich diese Kategorisierung primär am Systemverständnis und bezieht sich auf das Wissen des Nutzers über die Funktionsweise und Bedienung des technischen Systems. Durch das Fehlen eines entsprechenden Kommunikationsgerätes, welches üblicherweise für die Erzeugung der Kommunikationssignale zuständig ist, scheint es zweckdienlich zu sein, auch die nonverbale Kommunikationsfähigkeit des Nutzers bei dieser Interaktionsform in die Klassifizierung mit einfließen zu lassen. Daher werden in diesem Konzept, gemäß des Ansatzes des *User Modeling* [305], Nutzer mittels eines dreiteiligen Klassifizierungssystems in Abhängigkeit ihres nonverbalen Ausdrucksvermögens, dem geforderten Funktionsbedarf als auch anhand ihres Wissensstandes über die Funktionsweise des Systems in *Ungeschulte*, *Geschulte* und *Experten* eingeteilt.

Ungeschulte

Ungeschulte sind Personen, die bisher keine Berührungspunkte mit dem System hatten und daher auch kein Vorwissen über die Bedienung dessen besitzen. Ihr Interaktionsverhalten orientiert sich an bekannten Mustern aus der zwischenmenschlichen Erfahrung und beschränkt sich auf wenige Darstellungsformen, beispielsweise auf lautes Rufen, großräumiges Winken mit den Armen oder auf einfache Zeigegesten. Aufgrund des fehlenden Vorwissens über die unterstützten Funktionen des Systems sind auch die möglichen Kommunikationsinhalte sehr eingeschränkt. In gewissen Kontexten sind jedoch die wenigen Intentionen, wie die Erzeugung von Aufmerksamkeit *(„Ich brauche Hilfe!“)* oder die Mitteilung über ein physisches Problem (*„Ich habe ein Problem mit meinem Bein!“*) bereits ausreichend, um weitere Maßnahmen zum Wohle des Nutzers in Gang zu setzen. Die Verantwortlichkeit zum Führen und Aufrechterhalten des Dialogs mit dieser Nutzergruppe liegt dabei beim System.

Geschulte

Im Gegensatz zu Ungeschulten haben Geschulte bereits Erfahrung mit dieser Art der Mensch-UAV-Interaktion. Sie kennen die Grundzüge der Systembedienung und sind in der Lage, mit Hilfe eines kleinen, festgelegten Vokabulars einfache Anweisungen an das System zu übermitteln. Hierzu gehört beispielsweise die Fernführung des UAVs bei Start-, Lande- oder Absetzvorgängen. Zum Repertoire an gestischen Darstellungsformen gehören bei Geschulten statische und dynamische Bewegungen, die mit eindeutigen Bedeutungen belegt sind. Die Dialogführung übernimmt der Nutzer, wobei der Kommunikationskontext sich auf die kleinräumige Flugführung sowie die Aktivierung einfacher, mechatronischer Funktionen des UAVs (Scheinwerfer ein-/ausschalten, Transporthaken öffnen/schließen, etc.) beschränkt.

Experten

Den größten Einblick in den Funktionsumfang des visuellen Kommunikationssystems und die Fähigkeiten des verwendeten UAVs hat die Gruppe der Experten. Sie verfügt über das größte nonverbale Ausdrucksvermögen und nutzt eine definierte Syntax, um spezifische Systemparameter abzufragen oder Intentionen nonverbal zu formulieren. Das verfügbare Vokabular kann dabei je nach Kontext mit unterschiedlichen Bedeutungen belegt sein. Die Dialogführung obliegt hierbei beim Nutzer, das System verhält sich vorrangig passiv, gibt jedoch Rückmeldungen über verstandene Anweisungen und stellt Rückfragen bei unvollständigen Intentionsbeschreibungen.

Insgesamt wird also ein nutzeradaptives Systemkonzept benötigt, welches die Kommunikationsfähigkeiten und die gewünschten Kommunikationsinhalte der jeweiligen Nutzergruppe (Tab. 8) abbilden kann.

Tab. 8 Einteilung der Nutzergruppen nach nonverbalem Ausdrucksvermögen, Funktionsbedarf und Wissensstand über das Kommunikationssystem

	Ungeschulte	Geschulte	Experten
Vorwissen über Funktionsweise	kein	begrenzt	tiefergehend
Kommunikationsinhalte	eingeschränkt	anwendungsgebunden	umfangreich
Kommunikationsmittel	generisch	spezifisch	spezifisch
Mehrdeutiges Vokabular	nein	nein	ja
Steuerung des Dialogs	System	Nutzer	Nutzer

4.3.3 Inhärente Erkennung intentionaler Körpersignale

Bei Kommunikationssystemen werden üblicherweise spezielle Signale genutzt, um den Beginn und das Ende einer Übertragung zu markieren. Im Sinne einer natürlichen Interaktion sollte das System über einen Mechanismus verfügen, der feststellen kann, wann der Nutzer mit dem System nonverbal kommuniziert und wann nur ungerichtete Körpersignale ausgestrahlt werden oder die Aufmerksamkeit abseits des Systems liegt.

4.3.4 Zugangsregelung

Um den Zugang zum System regeln und damit zwischen autorisierten und nicht autorisierten Nutzern unterscheiden zu können, muss ein visuelles Kommunikationssystem über einen geeigneten Authentifizierungsmechanismus verfügen (vgl. Kapitel 3.1.3). Wenn Menschen miteinander sprechen möchten, dann teilen sie das in der Regel durch Ansprechen, Zurufen, eine entsprechende Geste oder einer Kombination aus allem mit. Kennen sich beide Personen, beginnt für gewöhnlich eine Konversation, andernfalls muss erst geklärt werden, in welcher Beziehung beide Gesprächspartner zueinanderstehen und welche Absicht durch das initiierte Gespräch verfolgt wird. Bei näherer Betrachtung lassen sich in diesem Verhalten zwei eigenständige Vorgänge identifizieren, die für die Umsetzung in ein technisches System getrennt modelliert werden müssen. Zum einen wird ein Verfahren benötigt, welches aus einer Menge an Personen, diejenige identifizieren kann, die einen Interaktionswunsch mit dem System beabsichtigt (*Nutzeridentifikation*). Zum anderen wird ein Mechanismus benötigt, mit dem die Identität und die damit verbundenen Befugnisse und Zugriffsrechte einer Person auf das System festgestellt werden können (*Nutzerauthentifizierung*). Beide Vorgänge können sowohl nacheinander als auch gleichzeitig ablaufen.

Nutzeridentifikation

Der Blickkontakt stellt in der zwischenmenschlichen Interaktion ein starkes Kommunikationssignal dar und wurde in der Vergangenheit auch zur Aktivierung interaktiver bodengebundener [306] als auch fliegender robotischer Systeme erfolgreich genutzt [196]. Probleme entstehen jedoch dann, wenn keine eindeutige Nutzeridentifikation möglich ist, weil die Entfernung für eine Bestimmung der Nutzerblickrichtung zu groß ist oder sich mehrere Personen im Interaktionsbereich des UAVs befinden. In solchen Fällen kann eine auffällige (Hand-) Bewegung oder ein für die jeweilige Situation atypisches Verhalten als Interaktionswunsch gedeutet werden.

Nutzerfernauthentifizierung

Neben der Erkennung eines Nutzungsinteresses, ist der Nachweis der entsprechenden Nutzungsbefugnisse ein weiterer zugriffsbezogener Aspekt der Interaktion. Bei der Nutzerauthentifikation muss sich eine Person authentisieren, d.h. sie muss Nachweise vorlegen, die beweisen, dass sie die vorgegebene Identität besitzt und über entsprechende Zugriffsrechte für die Systemdienste verfügt. Für eine solche visuelle Fernauthentifizierung kommen drei grundsätzliche Zugriffsregelungsverfahren in Frage, die sich am Wissen (*Authentifizierung mittels geheimer Information*), einer Eigenschaft (*Authentifizierung mittels biometrischen Merkmals*) oder dem Besitz orientieren (*Authentifizierung mittels eines definierten Gegenstandes*).

Da für die Authentifizierung nur der visuelle Kanal zur Verfügung steht, muss auch eine geheime Information (vgl. Parole) in Form von kodierten Körpersignalen ebenfalls über diesen Kanal übermittelt werden. Dies kann beispielsweise über eine bestimmte Körperpose oder eine spezielle Sequenz aus mehreren gestischen Darstellungen in kurzer zeitlicher Abfolge erfolgen. Zwar kann mit einem solchen Verfahren ein unberechtigter Zugriff von „Nicht-Eingeweihten“ verhindert werden, durch eine reine Beobachtung der Interaktion kann die geheime Information jedoch leicht abgefangen werden, sodass diese Methode ein sehr niedriges Sicherheitsniveau bietet.

Wesentlich sicherer in Bezug auf den Identitätsdiebstahl, ist die Fernauthentifizierung mittels sichtbarer, biometrischer Merkmale wie der Körperform [307], des Skeletts [308] oder des Gesichts [309]. Die Zuverlässigkeit dieser Authentifizierungsmethoden fällt und sinkt jedoch mit der Robustheit der verwendeten Verfahren gegenüber Umwelteinflüssen und den Anforderungen an das Sensorsignal. Für eine Authentifizierung mittels eines Skelettmodels wird beispielsweise eine tiefenfähige Sensorik benötigt, um Mehrdeutigkeiten auflösen zu können. Eine Gesichtserkennung kann hingegen auch mit einer zweidimensionalen Sensorik realisiert werden. Authentifizierungsmethoden, die biometrische Merkmale nutzen, haben jedoch den Nachteil, dass die Identitäten und damit auch die Rechte nicht übertragbar sind, sodass jeder autorisierte Nutzer sowie dessen Berechtigungen in einer Datenbank hinterlegt sein müssen.

Ein alternatives Verfahren, welches weniger Anforderungen an die Sensorik stellt und zudem nicht personengebunden ist, bildet die Authentifizierung mittels eines Gegenstandes mit definierten physikalischen Merkmalen. So wird beispielsweise auf dem Vorfeld von Flughäfen oder auf Flugzeugträgern vom Funktionspersonal spezielle Kleidung (Abb. 41 links) getragen, welche aufgrund ihrer markanten Farbgestaltung einerseits zu einer hohen Sichtbarkeit gegenüber anderen Verkehrsteilnehmer beiträgt, gleichzeitig aber auch den jeweiligen Aufgabenbereich klar erkennen lässt (z.B. Einweiser, Techniker oder Bauarbeiter). Da das Vorfeld zum Sicherheitsbereich eines Flughafens zählt, müssen die sich dort aufhaltenden Personen zudem einen weiteren Authentifizierungsgegenstand in Form eines gültigen Ausweises jederzeit sichtbar bei sich tragen. Für eine visuelle Validierung eines solchen Ausweises aus der Ferne reicht das begrenzte Auflösungsvermögen der UAV-Sensorik im Allgemeinen nicht aus, sodass alternative Merkmale gefunden werden müssen, wenn dieses Verfahren auf die Mensch-UAV-Interaktion übertragen werden soll.

Abb. 41 Warnschutzweste (links) und ArUco-Marker (rechts) als physikalische Authentifizierungsmittel

In der Logistik werden zur für die Kennzeichnung von Waren (z.B. Quick-Response-Codes), aber auch in der Robotik zur Orientierung und Lokalisierung (z.B. ArUco-Marker [310]), zweidimensionale symbolische Codes (Abb. 41 rechts) genutzt, die mit EO-Sensoren maschinell ausgelesen werden können und je nach eingesetztem Kodierungsverfahren auch über Fehlerkorrekturmechanismen verfügen, die das Auslesen der Daten auch bei teilweiser Verdeckung erlauben. Als Bestandteil eines tragbaren Ausweises oder eingebettet in eine Uniform mit digitalem Camouflage-Muster [311], kann mithilfe dieser Codes ein rudimentäres, visuelles Authentifizierungssystem für mehrere Nutzer mit unterschiedlichen Zugriffsberechtigungen realisiert werden.

4.3.5 Echtzeitfähigkeit des Systems

Ein wesentlicher Parameter zur Beurteilung der Leistungsfähigkeit eines technischen Systems sind seine Reaktionszeiten. Damit werden diejenigen Zeiten bezeichnet, die ein System benötigt, um auf bestimmte externe Reize in definierten Weisen zu reagieren [312, S. 255]. Im Falle eines nonverbalen Kommunikationssystems werden die Reize durch die sichtbaren Bewegungen des Nutzers ausgelöst, auf die das System innerhalb einer bestimmten Zeit reagiert. Wie bei den meisten interaktiven Systemen sollte auch hier ein möglichst schnelles Ansprechverhalten (geringe Latenz) angestrebt werden, um eine hohe Benutzerfreundlichkeit zu erreichen. Durch das bildgebende Datenerzeugungsverfahren und die elektronische Datenverarbeitung entstehen jedoch zwangsläufig Latenzen bei der:

- Belichtung des Bildsensors,
- Digitalisierung und dem Transport der Daten zur Recheneinheit,
- Detektion der relevanten Merkmale,
- Klassifizierung der Merkmale,
- Interpretation der Beobachtungen sowie der
- Generierung und Anzeige der Systemantwort.

Je nach Anwendungskontext werden unterschiedliche Anforderungen an die Reaktionszeiten gestellt, wobei zwischen zeitkritischen und zeitunkritischen Anwendungen unterschieden werden muss. Bei zeitkritischen Anwendungen ist „der Zeitpunkt, zu dem Ausgaben vom System erzeugt werden, bedeutend“ [313, S. 7]. Diese auch als *Echtzeitanwendungen* bekannten Anwendungen beinhalten im Allgemeinen Funktionen, deren Ausführung in einem festgelegten Zeitrahmen zu erfolgen hat, da andernfalls Gefahren für Mensch und Gerät entstehen können. Dazu zählen insbesondere Steuerungsaufgaben, die beispielsweise bei der Fernführung eines UAVs oder der Teleoperation auftreten können. Hohe Latenzen führen bei zeitkritischen Anwendungen zudem zur Reduzierung der Leistungsfähigkeit des Nutzers bei der Erreichung seines Ziels und wirken sich darüber hinaus auch negativ auf den Gemütszustand und damit auch auf das Benutzererlebnis aus. In Probandenversuchen, bei denen die Teilnehmer ein Fahrzeug aus einem Irrgarten per Fernsteuerung herausführen sollten, wurde der

Einfluss der Zeitverzögerung in Bezug auf den emotionalen Zustand und die Zufriedenheit hin untersucht [314, S. 446]. Dabei wurde festgestellt, dass eine zeitverzögerte Reaktion des Fahrzeugs von zwei bis drei Sekunden bereits Frustration und Wut bei den Probanden erzeugte und die Zufriedenheit während der Interaktion negativ beeinflusste.

Bei zeitunkritischen (sog. nebenläufigen) Anwendungen kann die Reaktionszeit zwar nicht komplett ignoriert werden, es werden jedoch andere Anforderungen an diese gestellt. In den Anfangszeiten der Mensch-Computer-Interaktion wurden in Probandenversuchen drei markante Schwellwerte für die tolerierbare Reaktionszeit bei der Interaktion mit einem Terminal ermittelt [315]:

- 0.1 Sekunde (bis zur Rückmeldung über die Eingabe),
- 1.0 Sekunde (bis zur Reaktion des Systems),
- 10 Sekunden (maximale Zeit, bis Aufmerksamkeit des Nutzers schwindet).

Bei noch längeren Wartezeiten treten Ablenkungstendenzen ein, sodass die „Freude an der Nutzung" (*Joy of Use*) sinkt. In der zwischenmenschlichen Interaktion hingegen werden Pausen von mehr als vier Sekunden als „peinliche Stille" wahrgenommen und implizieren einen Abbruch des Gesprächsfadens [315, S. 267]. Diese Werte können daher als Orientierungswerte für die maximal tolerierbaren Antwortzeiten des Systems herangezogen werden. Ein schnelles Ansprechverhalten alleine führt nicht zu einem effizienten Kommunikationssystem. Wenn die vom Sender übermittelten Daten vom Empfänger falsch dekodiert werden, müssen diese nochmals übertragen werden, wodurch sich die Übermittlungszeit verlängert. Im schlechtesten Fall muss dies mehrmals erfolgen. Die korrekte Detektion und korrekte Interpretation eines nonverbalen Signals kann daher unter dem Begriff *Erkennungsleistung* zusammengefasst werden. Darüber hinaus können weitere Metriken (Klassifikationsleistung, Genauigkeit, Fehlerrate) bestimmt werden, um die Leistungsfähigkeit des Erfassungssystems auf Merkmalebene zu bestimmen.

4.3.6 Visuelle Übermittlung der Systemantworten

In der Regel werden bei stationären und mobilen visuellen Kommunikationssystemen Lautsprecher und Displays für die Ausgabe der Systemantworten verwendet. Aufgrund der Geräuschkulisse der Auftriebserzeugung und der Gewichtsbegrenzung müssen für ein fliegendes System kompaktere und leichtere Lösungen als Rückmeldeverfahren gefunden werden. Bei der Nutzung eines rein visuellen Rückmeldeverfahrens muss auch die Sehleistung des Menschen miteinbezogen werden. Das Auflösungsvermögen eines normalsichtigen Menschen liegt bei optimalen Bedingungen[48] bei etwa einer Winkelminute (Visus von 1,0), d.h. auf einer Entfernung von fünf Metern können zwei Punkte aufgelöst werden, die rund 1,5 Millimeter[49] voneinander entfernt sind [317, S. 76]. Bei textuellen Rückmeldungen muss zusätzlich vorausgesetzt werden, dass der Nutzer zum einen überhaupt lesen kann und zum anderen auch die entsprechende Sprache beherrscht. Ist dies nicht der Fall, müssen andere Wege der Darstellung genutzt werden, beispielsweise über graphische Animationen oder allgemeinverständliche Symbole.

4.3.7 Ergonomisches Vokabular

Mensch-Maschine-Schnittstellen im Allgemeinen erlauben Nutzern die Interaktion mit technischen Systemen. Aus ergonomischer Sicht ist man bestrebt, neben einer möglichst hohen Benutzbarkeit auch

[48] Abbildung im Bereich des schärfsten Sehens (fovea centralis) unter photopischen Bedingungen (Tagsehen, Leuchtdichte 80-320 cd/m²) [316, S. 822].
[49] Entspricht der Breite der Lücke eines Landolt-Ringes.

eine intuitive Bedienung zu erreichen, um dem Nutzer den Einstieg in den Umgang mit diesen Systemen zu erleichtern. Intuitiv ist die Bedienung des Systems unter anderem dann, wenn sich das Gerät erwartungskonform verhält und die Interaktion auf bekannten mentalen Modellen fußt, sodass auch Nutzer ohne Vorwissen zügig den Umgang mit dem System erlernen können. Nach Nielsen beeinflussen fünf wesentliche Aspekte die Benutzbarkeit eines interaktiven Systems: *Erlernbarkeit*, *Effizienz in der Anwendung*, *Merkbarkeit*, *niedrige Fehlerrate* und ein *hohes Benutzererlebnis* [318, S. 25]. Einen ganzheitlichen Blick auf die Interaktion legt man beim Benutzererlebnis. Dabei versucht man, während und nach der Interaktion ein positives Nutzungsempfinden beim Nutzer zu erzeugen.

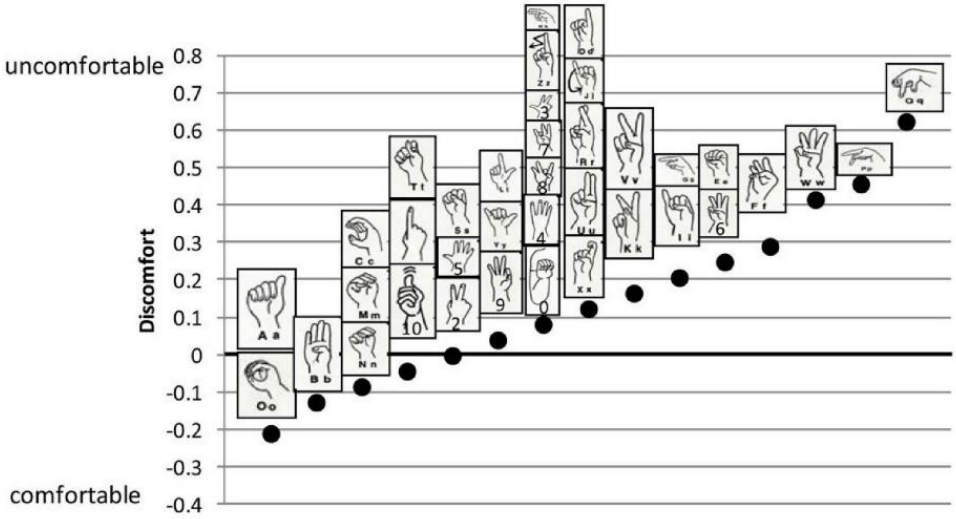


Abb. 42 Handzeichen sortiert nach Komfort, übernommen aus [319, S. 15]

Einen maßgeblichen Beitrag dazu kann ein ergonomisches Vokabular leisten. Erzeugen beispielsweise die gewählten Darstellungsformen während der Interaktion Spannungen und Verkrampfungen beim Nutzer, sinkt das Benutzererlebnis und damit auch die Akzeptanz dieser Eingabemethode. So wurden in einer Untersuchung mit professionellen Gebärdenübersetzern 37 Handzeichen bezüglich des Komforts bei der Ausführung untersucht und bewertet [319]. Handgesten, die stark von der Neutralstellung abwichen (d.h. Finger leicht gekrümmt und entspannt) als auch uneinheitliche Fingerstellungen erforderten, wurden als anstrengend und unkomfortabel bewertet (Abb. 42). Bei der Zusammenstellung des Vokabulars sollten zudem die psycho-physiologischen Faktoren des Menschen [320] beachtet werden und beispielsweise das Auftreten eines sogenannten *Gorilla-Arms* aufgrund von großräumigen, wiederkehrenden Armbewegungen vermieden werden [303, S. 63]. Die komfortabelste Position für die Hände beim Gestikulieren ist im unteren Brustbereich, nahe am Körper. Im Allgemeinen werden sanfte Bewegungen der Ellbogen oder Schultern als angenehm empfunden [319].

Ähnlich wie beim Morsecode, sollten Signale, die während der Interaktion häufig genutzt werden, durch kurze Darstellungen mit geringem Bewegungsanteil kodiert werden. Dadurch kann die Interaktionsdauer reduziert und die Effizienz der Kommunikation erhöht werden. Bei der Zuweisung einer Darstellung zu einem Signal muss immer auch der kulturelle Hintergrund der Nutzergruppe in Betracht gezogen werden. Nonverbale Äußerungen, die in einem Kulturkreis mit einer bestimmten Bedeutung belegt sind, müssen in einem anderen Kulturkreis nicht zwangsläufig mit derselben Bedeutung belegt sein. Die Intuitivität des Vokabulars kann durch einen hohen Anteil an ikonischen und metaphorischen Gesten erreicht werden [182, S. 176]. Diese haben in der Regel einen hohen Bezug zur Realität und sind daher oftmals leichter erlernbar als speziell erschaffene Gesten, die in der zwischenmenschlichen Interaktion selten oder überhaupt nicht vorkommen.

Kodierung von Signalen in nonverbale Darstellungen

Die Kodierung von Signalen in nonverbale Darstellungen kann als multikriterielles Optimierungsproblem (Abb. 43) mit den Parametern *Intuitivität*, *Erkennungsgenauigkeit*, *Komfort* und *Intuitivität* aufgefasst werden [320, S. 140]. Eine großräumige Bewegung bietet im Gegensatz zu einer kleinräumigen ein besseres „Signal-Rausch-Verhältnis" (hohe Detektierbarkeit), führt aber bei wiederholter Ausführung möglicherweise zu Ermüdungserscheinungen der betreffenden Körperpartien (niedrige Ergonomie). Gesten, die oft im Alltag Verwendung finden, werden als intuitiv angesehen (hohe Intuitivität), sind aber vielleicht mit mehreren situationsabhängigen Bedeutungen belegt (keine Eindeutigkeit).

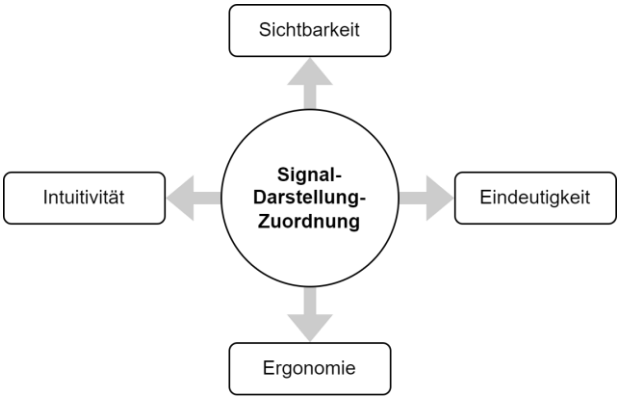


Abb. 43 Signal-Darstellung-Zuordnung als multikriterielles Optimierungsproblem

Größe und Ambiguität des Wortschatzes

Ein unbegrenztes Vokabular ist weder technisch realisierbar noch erstrebenswert für die Interaktion mit einem menschlichen Nutzer. Damit ein Mensch eine Sprache erlernen kann, muss die Anzahl verschiedener Grundformen begrenzt sein [106, S. 18-19]. Um dennoch einen großen Funktionsbereich abdecken zu können und gleichzeitig die Anzahl der Varianten gering zu halten, ist eine kontextabhängige Signal-Darstellung-Zuordnung unumgänglich. Für ein flexibles Vokabular muss diese Zuordnung auch konfigurierbar sein und die systematische Erzeugung neuer Signale und Darstellungen ermöglichen, um neue UAV-Funktionen mit einem angepassten Wortschatz beschreiben zu können.

4.3.8 Sicherheitsrelevante Aspekte

Abhängig vom Anwendungskontext, in dem das visuelle Mensch-UAV-Kommunikationssystem eingesetzt wird, müssen auch sicherheitsrelevante Aspekte beim Systemdesign beachtet werden. Im Hinblick auf die notwendige Erkennungsleistung und Fehlertoleranz der Signalverarbeitung können drei potentielle, risikobasierte Anwendungsszenarien[50] bestimmt werden, bei denen das UAV als *Unterhaltungsgerät*, *Werkzeug* oder *Transportvehikel* genutzt wird.

Das UAV als Unterhaltungsgerät

Kleine Multikoptersysteme mit einem Abfluggewicht von unter einem Kilogramm (vgl. „offene" UAS-Betriebskategorie C0 und C1) werden im Privatsegment vorrangig zur Unterhaltung oder Dokumentation von konservierungswürdigen Ereignissen (Urlaubserlebnisse, Familienfeiern, Hochzeiten) genutzt. Ein visuelles Kommunikationssystem kann bei dieser Geräteklasse den Komfort und den Joy of Use durch den Verzicht auf eine zusätzliche Fernsteuerung erhöhen. Die wenigen Nutzereingaben

[50] Angelehnt an die UAS-Betriebskategorien *offen*, *speziell* und *zulassungspflichtig* der Europäischen Union [40].

beziehen sich primär auf die Aktivierung unkritischer Bordfunktionen (Start/Stop der Photo-/Videoaufnahme) und die nonverbale Führung des UAVs in unmittelbarer Umgebung des Nutzers. Aufgrund der geringen kinetischen Aufprallenergie[51] stellen UAVs dieser Klasse eine verhältnismäßig geringe Gefahr für den Nutzer im Falle einer Fehlbedienung dar, sodass die Umsetzung einer gestischen Eingabe auch erst nach mehrmaliger Produktion toleriert werden kann. Dennoch muss das Kommunikationssystem zu einem gewissen Grad robust gegenüber unbeabsichtigten Nutzereingaben sein, da von einem ungeschulten Nutzerkreis ausgegangen werden muss.

Das UAV als Werkzeug

Im kommerziellen Anwendungskontext kommen UAVs der Größenordnung bis etwa 150 kg als luftgestützte Werkzeuge zum Einsatz, die neben Mess- und Aufzeichnungsinstrumenten auch leichte Waren zum Kunden ausliefern können. Ein visuelles Kommunikationssystem kann das UAV bei der lokalen Navigation und Kommunikation mit dem Warenempfänger unterstützen. In Anbetracht der wesentlich höheren Aufprallenergien (vgl. „spezielle“ UAS-Betriebskategorie), werden hierbei höhere Anforderungen an die Zuverlässigkeit des Fluggerätes und somit auch des Kommunikationssystems gestellt.

Das UAV als Transportvehikel

Ab einer gewissen Größenordnung des Fluggerätes nähern sich die Anforderungen an dessen Zuverlässigkeit an die der bemannten Luftfahrt (vgl. „zulassungspflichtige“ UAS-Betriebskategorie). Hierunter fallen unbemannte Fluggeräte, die nicht nur besonders schwere Lasten, sondern auch Menschen transportieren können (z.B. im Kontext von Urban Air Mobility). Aufgrund der Größe und Systemkomplexität, sind UAVs dieser Klasse in bestehende (Flughäfen) oder zukünftige Infrastrukturen (z.B. „Skyports“ für Flugtaxis [321]) integriert, bei denen auch visuelle Kommunikationsverfahren zum Einsatz kommen. Daher werden ähnlich hohe Anforderungen an die Zuverlässigkeit des Kommunikationssystems gestellt, welches als Voraussetzung für die sichere Integration von UAVs in ebensolche Infrastrukturen angesehen werden kann. Tab. 9 zeigt die drei Anwendungsszenarien in der Übersicht.

Tab. 9 Sicherheitsbezogene Anforderungen an ein visuelles Kommunikationssystem

Zweck des UAVs	Größenordnung	Nutzen eines visuellen Kommunikationssystems	Gefahrenpotential für Nutzer	Anforderungen an Zuverlässigkeit	Anforderungen an Robustheit
Unterhaltung	< 1 kg	Komfortgewinn	gering	Niedrig	hoch
Werkzeug	1 … 150 kg	Sicherheitsgewinn	hoch	Hoch	hoch
Transportvehikel	≥ 150 kg	Integration in bestehende Infrastrukturen und Verfahren	sehr hoch	sehr hoch	sehr hoch

4.3.9 Zusätzliche Anforderungen an ein Experimentalsystem

Für die Evaluierung der Mensch-UAV-Interaktion muss ein geeignetes Experimentalsystem nicht nur über eine Möglichkeit zur Aufzeichnung der Nutzereingaben und Systemausgaben verfügen, sondern erlaubt idealerweise auch eine Speicherung der Zwischenergebnisse, um Systementscheidungen im Nachgang nachvollziehen zu können. Um die Leistungsfähigkeit des Systems nicht negativ zu beeinflussen, sollte darüber hinaus der für die Speicherung der Daten notwendige zusätzliche Rechenlast so gering wie möglich gehalten werden. Auch müssen Absicherungsmaßnahmen in die Systemlogik

[51] Maximale kinetische Aufprallenergie von 80 J bei der UAS-Betriebskategorie C1.

miteinbezogen werden, um die Sicherheit der Probanden bei einer Erprobung im Realflug sicherzustellen.

4.4 Das UAV als „Dienstleister“

Unter der Annahme, dass sich ein hochautomatisiertes UAV kollisionsfrei in der Welt bewegen kann und damit die Notwendigkeit einer kontinuierlichen Steuerung durch einen menschlichen Fernpiloten entfällt, kann ein UAV als fliegendes Assistenzsystem angesehen werden, welches als zeitlich und lokal befristete Ressource zur Unterstützung oder Durchführung von Aufgaben genutzt werden kann und eine definierte Menge an mechatronischen Diensten anbietet (vgl. „UAV-as-a-Service“ [86, 322]). Die Art dieser Dienste kann dabei von einfachen Abfragen und Setzen von Systemparametern bis hin zu komplexen und hochautomatisierten Aufgaben wie dem vollautomatisierten Transport einer Nutzlast von A nach B reichen.

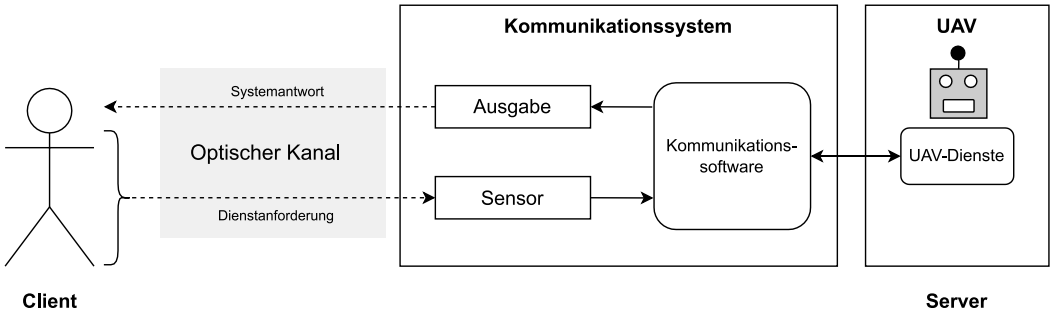


Abb. 44 Vereinfachte Darstellung der visuellen Mensch-UAV-Interaktion als Client/Server-Modell mithilfe eines zwischengeschalteten Kommunikationssystems

Um einen solchen Dienst in Anspruch zu nehmen, muss ein Nutzer dem UAV in Form einer Absichtserklärung mitteilen, welchen er aus einer Menge an verfügbaren Diensten anfordern möchte. Analog zu sprachbasierten Schnittstellen muss die Beschreibung dieser Intention sequentiell erfolgen, nur mit dem Unterschied, dass diese Beschreibung nicht verbal, sondern durch sichtbare Körpersignale erfolgt. Nutzer und UAV können in der Regel nicht direkt miteinander interagieren, sondern benötigen ein zwischengeschaltetes Kommunikationssystem, welches als Übersetzungsschnittstelle dient und die „Sprache“ des einen in die des jeweils anderen konvertieren kann. Für die Modellierung dieser Art von Mensch-Maschine-Interaktion existieren viele mögliche Herangehensweisen. Aufgrund der hohen Ähnlichkeit zum Client/Server-Prinzip (Abb. 44), welches dem Schema von Dienstanforderung und Diensterbringung folgt, können die Methoden und Modelle des *Protocol Engineerings* [323] angewendet werden. Diese bieten eine systematische Herangehensweise auf der Grundlage eines Kommunikationsmodels und nutzen Konzepte der Softwaretechnik für die Beschreibung der Interaktion, aufgeteilt in Dienste, Schichten und Protokolle.

4.4.1 Kommunikationsmodell

Die zugrunde liegende Kommunikationsform ähnelt einem klassischen Encoder/Decoder-Kommunikationsmodell (vgl. Shannon und Weaver [324]). Hierbei tauschen zwei Systeme, Sender und Empfänger alias Mensch und rechnergestütztes Kommunikationssystem, kodierte Nachrichten über einen gemeinsamen (optischen) Kanal aus. Der menschliche Körper bildet hierbei das senderseitige „Kodiergerät“, welches die symbolisch formulierte Intention des Senders in eine Sequenz aus nonverbalen Darstellungen überführt. Während sich die Intention, d.h. der Nachrichteninhalt, an den UAV-Fähigkeiten (UAV-Dienste) orientiert, definiert die gemeinsame Sprachgrundlage sowie der situative

und semantische Kontext das Kodierungsverfahren. Um den Nachrichteninhalt dekodieren zu können, muss der Empfänger die kodierten Körpersignale zunächst mit einer geeigneten Sensorik erfassen, relevante Körpermerkmale klassifizieren und diese Subsignale zu Beobachtungen zusammensetzen. Erst danach kann eine kontext- und situationssensitive Dekodierung der Beobachtungen mittels definierter Übersetzungsregeln erfolgen und die Intention des Senders bestimmt werden.

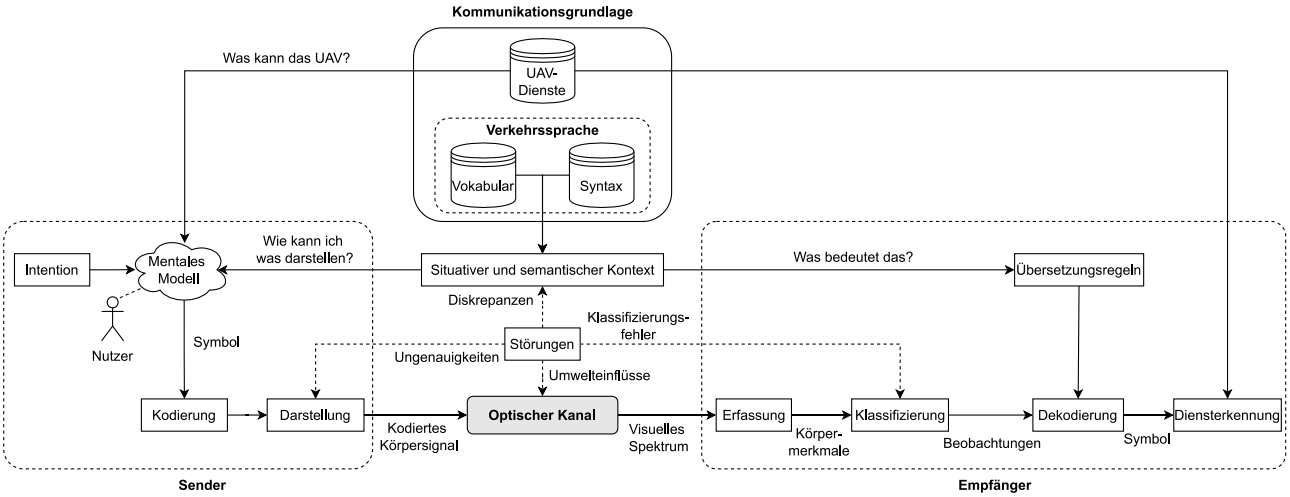


Abb. 45 Das VisCom-Kommunikationsmodell

In diesem Kommunikationsmodell (Abb. 45) können mehrere Störgrößen die Nachrichtenübermittlung negativ beeinträchtigen. Auf der physikalischen Ebene verschlechtern Umwelteinflüsse (z.B. Wetterverhältnisse, atmosphärische Abbildungsungenauigkeiten, Verdeckungen) die optische Signalübertragung und erhöhen dadurch den „Rauschanteil“ des Signals während der Übermittlung. Auf der Kodierungsebene kann eine fehlerhafte Kodierung (Darstellung) zu einer niedrigen „Signalgüte“ führen, sodass auf der Dekoderseite bei einem niedrigen Signal-Rausch-Verhältnis mit Klassifizierungsfehlern gerechnet werden muss. Auf der syntaktischen Ebene können zudem Diskrepanzen zwischen dem mentalen Modell des Nutzers über den angenommenen Systemzustand und dem tatsächlichen Systemzustand zu Synchronisierungsfehlern führen, die in Signalverwechslungen oder -verlusten resultieren können.

4.4.2 VisCom-Schichtenmodell

In der Netzwerktechnik wird zwischen verbindungslosen und verbindungsorientierten Diensten unterschieden. Verbindungslose Dienste zeichnen sich dadurch aus, dass für die Datenübertragung keine explizite Kommunikationsbeziehung zwischen den Teilnehmern benötigt wird. Die Daten werden somit ohne vorherige Kontaktaufnahme an den Empfänger geschickt und der Empfang auch nicht bestätigt [323, S. 10]. Ein typischer Vertreter dieser Dienstart ist das *User Datagram Protocol* (UDP), welches primär für die durchsatzstarke Übermittlung vieler kleiner Datenmengen, die beispielsweise bei Echtzeit-Sprachübertragungen anfallen, genutzt wird.

Verbindungsorientierte Dienste arbeiten hingegen nach dem Prinzip des klassischen, analogen Telefondienstes und bestehen aus den drei Phasen *Verbindungsaufbau*, *Datenübertragung* und *Verbindungsabbau*. Der Vorteil dieses Vorgehens liegt darin, dass der Empfänger dem Verbindungsaufbau erst zustimmen muss, bevor eine Datenübertragung stattfinden kann und der Verbindungsabbau ebenfalls durch ein entsprechendes Datenpaket angekündigt wird. Um den Zugriff auf das Kommunikationssystem und damit das UAV nur auf bestimmte Nutzer beschränken und zugleich Verbindungsabbrüche erkennen zu können, wird daher ein verbindungsorientiertes Protokoll benötigt. Ein Beispiel

für ein solches Protokoll ist das *Transmission Control Protocol*, welches auf dem *Internet Protocol* aufsetzt und daher oft als TCP/IP abgekürzt wird. TCP/IP ist eine Gruppe von Netzwerkprotokollen, die vorrangig zur Kommunikation im Internet eingesetzt wird und dessen zugrundeliegende Architektur sich dabei am OSI-Referenzmodell[52] der *International Organization for Standardization* (ISO) orientiert. Bei diesem Modell [325] wird der Kommunikationsvorgang zwischen zwei Systemen in sieben funktionale, hierarchisch aufeinander aufbauende Schichten unterteilt, die jeweils unterschiedliche Dienste zur Verfügung stellen (Tab. 10). So stellt beispielsweise die unterste Schicht 1 die physikalische Übertragung der einzelnen Bits zwischen zwei über ein Medium direkt miteinander verbundener Systeme sicher [326, S. 73]. TCP/IP fasst einige der Schichten zusammen und gruppiert diese in jeweils eine *Netzzugangs-*, *Internet-*, *Transport-* und *Anwendungsschicht*. Dieses vierschichtige Modell kann auf die Besonderheiten einer visuellen Kommunikation übertragen werden und wird im Folgenden als *VisCom-Schichtenmodell* bezeichnet.

Tab. 10 Schichtenmodelle von OSI, TCP/IP und VisCom im Vergleich

OSI-Schicht	OSI-Modell	TCP/IP-Schicht	TCP/IP-Modell	VisCom-Schicht	VisCom-Modell
7	Anwendungen (Application)	4	Anwendungen	6	Anwendungen
6	Darstellung (Presentation)			5	Dienste
5	Sitzung (Session)			4	Interaktion
4	Transport (Transport)	3	Transport	3	Körpersignale
3	Vermittlung (Network)	2	Internet	2	Adressierung
2	Sicherung (Data Link)	1	Netzzugang	1	Zugang zum Medium und Sicherung der Sichtverbindung
1	Bitübertragung (Physical Layer)				

VisCom Schicht 1: Zugang zum Medium und Sicherung der Sichtverbindung

Die erste Schicht des VisCom-Modells entspricht der Netzzugangsschicht des TCP/IP-Modells. Das Übertragungsmedium ist hierbei der visuelle Teil des elektromagnetischen Spektrums. Über eine bildgebende Sensorik wird die vom Nutzer reflektierte Strahlung erfasst und in eine digitale Form überführt. Gleichzeitig erhält der Nutzer über das biologische System „Auge“ Zugang zum selben Teil des Spektrums und kann die Licht- oder Bewegungssignale des UAVs erfassen. Das Nutzererfassungssystem steuert die Ausrichtung des Sensors und sorgt im Zusammenspiel mit dem FCS des UAVs für einen optimalen Sensoreinsatz, um ein möglichst hochwertiges Signal des Nutzers bereitzustellen sowie die Sichtverbindung aufrecht zu erhalten.

VisCom Schicht 2: Adressierung

Während die erste Schicht vorrangig aus der Perspektive des UAVs die Sichtverbindung sicherstellt, überwacht die Adressierungsschicht die Verbindung aus der Nutzersicht. Dazu muss sie in der Lage sein, zwischen denjenigen intentionalen, nonverbalen Signalen zu unterscheiden, die an das UAV gerichtet sind und denjenigen, die andere Personen oder Geräte erreichen sollen.

[52] Open Systems Interconnection, Modell zur flexiblen und weltweiten Kommunikation zwischen unterschiedlichen Computersystemen mittels standardisierter Protokolle.

VisCom Schicht 3: Körpersignale

In der dritten Schicht werden die relevanten Körpermerkmale definiert, über die Subsignale transportiert werden können. Dazu werden diesen Körpermerkmalen diskrete Beobachtungsklassen zugewiesen, die jeweils ein Subsignal repräsentieren. Diese klassifizierten Körpersignale werden den höheren Schichten als Beobachtungen bereitgestellt.

VisCom Schicht 4: Interaktion

Die vierte Schicht entspricht der Interaktion, sodass hier definiert ist, welches Kommunikationsverfahren (Sprache) für die Datenübertragung genutzt wird und nach welchen Protokollen eine Verbindung auf- oder abgebaut und die Datenübermittlung gesteuert wird.

VisCom Schicht 5: Dienste

Die fünfte Schicht entspricht der Darstellungsschicht des OSI-Modells und enthält die Formatdefinitionen für die standardisierte Beschreibung der angebotenen Dienste.

VisCom Schicht 6: Anwendungen

Die oberste, sechste Schicht repräsentiert die Anwendungsschicht des OSI-Modells. Hierzu gehört neben der Erkennung des angeforderten Dienstes auch die Steuerung der Datenübertragung durch ein Dialogsystem.

4.4.3 Kommunikationsprotokoll

Im ursprünglichen Sinne bezeichnen Protokolle Regeln für den Ablauf von zwischenmenschlichen Aktivitäten, beispielsweise die diplomatischen Gepflogenheiten für den angemessenen Empfang eines politischen Oberhauptes. Im Bereich der Informationstechnologie werden diese Regelsätze als Kommunikationsprotokolle bezeichnet und enthalten Konventionen über den Aufbau und den zeitlichen Ablauf der Übertragung von Daten von einem technischen System zum anderen. Um den Kommunikationsprozess bei der Mensch-UAV-Interaktion in Form eines Protokolls abbilden zu können, muss zunächst die Kommunikationsform charakterisiert, der Aufbau der zu übermittelnden Daten definiert als auch der zeitliche Ablauf der Datenübertragung festgelegt werden.

Charakterisierung der Kommunikationsform

Wie bereits dargestellt, orientiert sich die Kommunikation an einem Client/Server-Prinzip und erfolgt wechselweise in beide Richtungen (Halbduplex-Verfahren). Die Daten sind an keinem Taktsignal ausgerichtet und können daher zu beliebiger Zeit übertragen werden (asynchrone Datenübertragung). Eine Verbindung besteht, wenn zwischen Nutzer und UAV ein Blickkontakt hergestellt wurde und die Datenübermittlung durch ein entsprechendes Startsignal eingeleitet wird. Wendet der Nutzer seinen Blick während der Datenübermittlung vom UAV ab, wird dies vom Kommunikationssystem als Verbindungsunterbrechung wahrgenommen und die Verarbeitung der Körpersignale für die Dauer des Blickkontaktverlustes ausgesetzt. Das Über- und Unterschreiten eines Bewegungsschwellwertes definiert den Anfang bzw. das Ende eines Körpersignals. Irrtümlich übermittelte Signale können durch die Übermittlung eines Korrektursignals revidiert oder die Datenübermittlung gänzlich neu gestartet werden. Die Datenübertragung wird durch den Aufruf eines entsprechenden Verbindungsdienstes abgebaut.

Datenformate

Je nachdem, welche Prozessierungsebene betrachtet wird, muss das Kommunikationssystem unterschiedliche Arten von Daten verarbeiten (Abb. 46). Auf der untersten Sensorebene erhält das System digitalisierte Informationen über die Umwelt in Form einer Matrix aus Bildpunkten. Mit Verfahren aus der Bildverarbeitung werden aus dieser Menge von Pixeln zunächst diejenigen bestimmt, welche die relevanten Körpermerkmale des Nutzers abbilden. Darauf aufbauend, wird auf der nächsthöheren Signalverarbeitungsebene in den räumlichen und zeitlichen Veränderungen dieser Merkmale nach kodierten Subsignalen gesucht, welche als klassifizierte Beobachtungen an die oberste Verarbeitungsebene, der sprachlichen Ebene, zur Verfügung gestellt werden. Hier werden die Beobachtungen im Kontext der aktuellen Nutzungs- und Interaktionssituation gedeutet und in textuelle Symbole übersetzt, wodurch die Datenprozessierung vom Signalverarbeitungsbereich in den linguistischen Bereich wechselt. Auf dieser obersten Verarbeitungsebene findet die eigentliche symbolbasierte Informationsübertragung zwischen Mensch und UAV mittels einer gemeinsamen Verkehrssprache statt, welche im Kapitel 4.8 näher erläutert wird.

Abb. 46 Unterschiedliche Verarbeitungsebenen des Kommunikationssystems

Zeitlicher Ablauf der Datenübertragung

Sobald sich der Nutzer innerhalb der Interaktionsdistanz des UAVs befindet und von der Nutzererfassung detektiert wurde, gliedert sich die nachfolgende Kommunikation in fünf Phasen:

- Optimierung der Sichtverbindung,
- Authentifizierung des Nutzers,
- Begrüßung (Anfrage zum Verbindungsaufbau),
- Interaktion (Datenübermittlung) sowie
- Verabschiedung (Anfrage zum Verbindungsabbau).

Nach der systemseitigen Optimierung der Sichtverbindung, findet zunächst eine (optionale) Authentifizierung des Nutzers statt. Je nach verwendetem Verfahren kann dies implizit (vom System ausgehend) oder explizit (vom Nutzer ausgehend) erfolgen. Verfügt der Nutzer über die notwendigen Berechtigungen, erhält er eine entsprechende optische Signalisierung vom System, andernfalls unterbleiben die Systemausgaben. Um die eigentliche Interaktion (d.h. die Datenübertragung) zu beginnen, übermittelt der Nutzer ein Begrüßungssignal und aktiviert damit die Diensterkennung. In dieser Phase der Interaktion erwartet das Kommunikationssystem eine symbolbasierte Beschreibung des gewünschten UAV-Dienstes mittels kodierter Körpersignale. Bei Übertragungsfehlern, sei es bedingt durch den Fehleingaben des Nutzers oder Erkennungsfehlern des Systems, kann der Nutzer auf die

Dialogdienste zurückgreifen und so Fehler korrigieren oder die Datenübermittlung komplett zurücksetzen. Sobald der angefragte Dienst eindeutig vom Kommunikationssystem identifiziert wurde, erhält der Nutzer eine Zusammenfassung der Dienstanforderung und verlangt eine Rückbestätigung durch den Nutzer. Erst wenn diese positiv ausfällt, leitet das Kommunikationssystem diese Dienstanfrage an das UAV weiter. Je nach Anwendungsfall kann das Kommunikationssystem die rückbestätigten UAV-Dienstanforderungen auch zunächst akkumulieren und erst nach einer entsprechenden Aktivierungsaufforderung an das UAV weiterleiten. Die Interaktion und damit die Datenübermittlung wird durch ein entsprechendes Verabschiedungssignal des Nutzers beendet.

4.4.4 Interaktionsmodell des Nutzers

Während der Interaktion kann das Verhalten des Nutzers als endlicher Automat modelliert werden, welcher immer nur einen aktiven Zustand erlaubt und eine gewisse Erwartungshaltung des Nutzers abbildet. Abb. 47 zeigt die möglichen Zustände und Erwartungen an das System sowie die Übergangsbedingungen. Blau hinterlegt sind die vier möglichen Zustände, in Weiß diejenigen Aktionen, die vom System im jeweiligen Zustand erwartet werden.

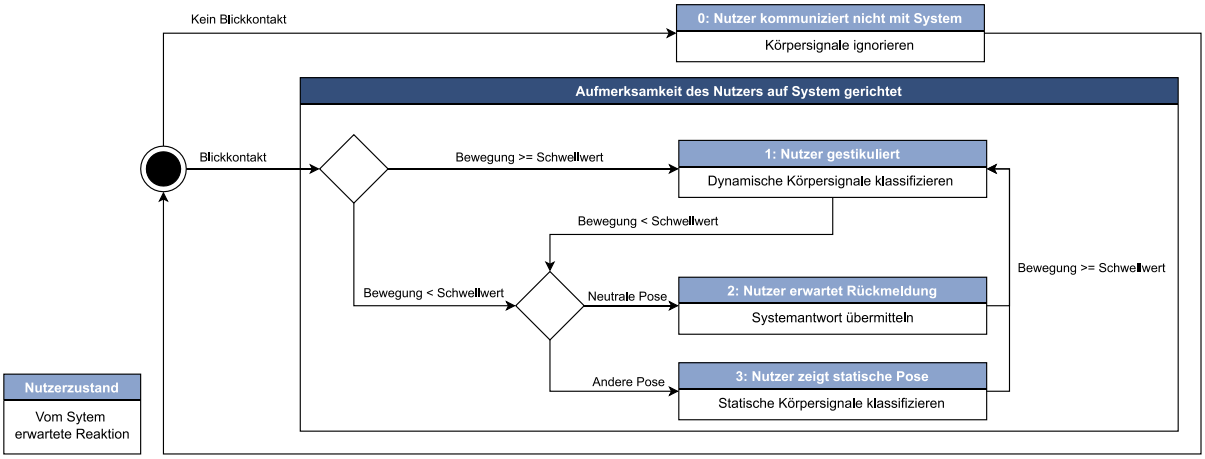


Abb. 47 Modellierung des Nutzerinteraktionsverhaltens als endlicher Automat

Die Adressierung eines technischen Systems mittels Ansehen ist ein geläufiges Verfahren in der HRI und wird auch für die Interaktion mit UAVs empfohlen [327, S. 3834]. Besteht daher kein Blickkontakt zum System, befindet es sich im *Zustand 0* und ignoriert alle Körpersignale. Wird der Blickkontakt hingegen geschlossen, bestimmt der Bewegungsanteil des Nutzers das weitere Systemverhalten. Übersteigt die Bewegung einen gewissen Schwellwert, wechselt das System in den *Zustand 1*, in dem es seinen Fokus auf die dynamischen, d.h. bewegungskodierten, Körpersignale richtet und diese klassifiziert. Geht der Nutzer anschließend in eine sogenannte *neutrale Pose* über, aktiviert das System den *Zustand 2*, welcher einen Richtungswechsel der Datenübermittlung und damit eine Systemantwort einleitet (z.B. über erkannte Eingaben oder noch fehlende Informationen). Dies ist auch der Zustand, in dem das System eine Hilfestellung im Falle von Missverständnissen anbieten kann, weil die Aufmerksamkeit des Nutzers hierbei beim System liegt und von einer geringen mentalen Belastung ausgegangen werden kann. Für den Fall, dass sich der Nutzer außerhalb der neutralen Pose befindet, der Bewegungsschwellwert aber nicht überschritten wurde, wird der *Systemzustand 3* aktiviert. Hierbei geht das System von einer statischen Signalübermittlung aus und legt damit den Fokus auf die Klassifizierung der statischen Körpersignale (z.B. Pose).

4.4.5 Modell zur Beschreibung des Systemverhaltens

Zum Erfolg einer nonverbalen Mensch-UAV-Interaktion trägt unter anderem ein erwartungskonformes Systemverhalten bei, welches dann entstehen kann, wenn zwischen der Vorstellung des Nutzers über die Funktionsweise des technischen Systems und dessen tatsächlichem Systemdesign eine möglichst hohe Ähnlichkeit und Passung erreicht wird [328, S. 93]. In Anlehnung an die interpersonelle Interaktion sollen daher folgende Erwartungshaltungen aus Nutzersicht erfüllt werden, um eine möglichst hohe Kongruenz zwischen dem mentalen Modell des Nutzers und dem implementierten Systemdesign zu ermöglichen. Das System:

- signalisiert mir, wenn es mich wahrnimmt,
- signalisiert mir, wenn es mir „zuhört“,
- reagiert nur auf meine Bewegungen, wenn ich es anschaue,
- gibt mir nach jeder Eingabe eine Rückmeldung über das erkannte Symbol und
- fasst den Inhalt der erkannten Intention zusammen und wartet auf meine Rückbestätigung, bevor es eine Intention speichert oder eine Aufgabe selbsttägig beginnt.

Für den Großteil der perzeptiven Signalisierungen eignen sich lichtbasierte Verfahren [292, S. 8]. Der letztgenannte Punkt orientiert sich am sogenannten *Read-Back-Verfahren* aus der Luftfahrt, bei dem durch ein wortwörtliches Wiederholen die Deckungsgleichheit der übermittelten und empfangenen Anweisungen überprüft und damit Übermittlungsfehler verhindert werden sollen.

4.4.6 Dialog- und UAV-bezogene Dienste

In diesem Konzept kann sowohl das UAV als auch das Kommunikationssystem eigene Dienste anbieten, sodass zwischen UAV-bezogenen Diensten (den eigentlichen UAV-Funktionen) und dialogbezogenen Diensten (d.h. der Verbindungssteuerung) unterschieden werden muss. Das Kommunikationssystem fungiert hierbei als eine gemeinsame Interaktionsplattform, welche die Anfragen eines angeschlossenen Interaktionspartners übersetzt und an den anderen Interaktionspartner weiterleitet.

In der Kommunikationstechnik wird die Nutzung eines Dienstes unter der Verwendung sogenannter „Dienstprimitiven“ (*service primitives*) beschrieben, welche Abstraktionen für die Beschreibung der Wechselbeziehung zwischen einem Dienstnutzer und einem Diensterbringer über definierte Schnittstellen, den Dienstzugangspunkten (*service access points*), darstellen [323, S. 6]. Von diesen Dienstprimitiven existieren vier grundlegende Funktionstypen: *request*, *indication*, *response* und *confirmation*. Dienstprimitive vom Typ *request* stellen Dienstanforderungen dar, die von einem Dienstnutzer A (dem Anfragenden) an einem Dienstzugangspunkt des Diensterbringers formuliert werden, während *indications* die Anzeige dieser Anforderungen am entsprechenden Dienstzugangspunkt eines anderen Dienstnutzers B (dem Angefragten) bilden. Der Angefragte antwortet auf die Dienstanfrage typischerweise mittels einer Dienstprimitiven vom Typ *response*, sodass beim Anfragenden eine entsprechende *confirmation* als Rückmeldung eintrifft. Mensch und UAV stellen somit in diesem Zusammenhang auf abstrakter Ebene jene zwei Dienstnutzer A und B dar, die die Kommunikationsdienste eines Diensterbringers (das visuelle Kommunikationssystem) über die Dienstzugangspunkte nutzen (Abb. 48).

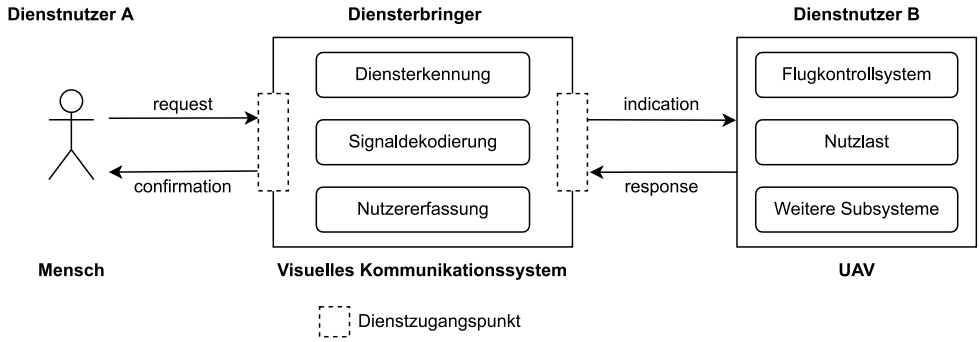


Abb. 48 Das Dienstmodell des VisCom-Konzepts

Für den Menschen stellen Sensoren und der visuelle Rückkanal die Dienstzugangspunkte des Kommunikationssystems dar, während das UAV – als technisches System – auf elektrische Dienstzugangspunkte in Form von digitalen Datenschnittstellen zurückgreifen kann. Die dialogbezogenen Dienste werden vom Kommunikationssystem direkt bearbeitet, wohingegen UAV-bezogene Dienstanforderungen zunächst vollständig vorliegen müssen, bevor sie an das UAV weitergeleitet werden können. Hierfür wird im Kommunikationssystem eine Nutzererfassung benötigt, die die relevanten Nutzersignale am Dienstzugangspunkt abgreifen, diese dekodieren (Signaldekodierung) und die angeforderte Dienstart bestimmen kann (Diensterkennung).

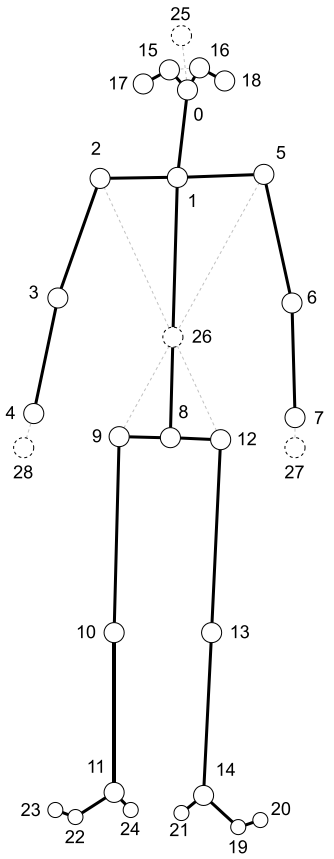
4.5 Visuelle Nutzer- und Subsignalmodellierung

Für die systemseitige Erfassung dieser Nutzersignale, wird neben einem anatomischen Modell mit visuellen Merkmalen auch ein Modell der verfügbaren Kommunikationsmittel benötigt.

4.5.1 Anatomisches Modell

Die Abbildung des menschlichen Körpers mittels eines digitalen Körpermodells kann sehr grob erfolgen (einfaches Strichmännchen), aber auch beliebig komplex umgesetzt werden (detailliertes Skelett inklusive Muskeln und Sehnen). Ein in der Literatur bewährtes Körpermodell stellt das in Abb. 49 dargestellte *BODY_25* Modell von OpenPose [251] dar. Es basiert auf der Notation des COCO Datensatzes [244] mit 18 Körpermerkmalen und wurde um sieben weitere Merkmale[53] erweitert.

[53] Kleiner Zeh, großer Zeh, Ferse (jeweils links und rechts) sowie Mitte der Hüfte



Index	Bezeichnung	Abkürzung
0	Nase	Na
1	Nacken	N
2	Schulter R	S_R
3	Ellenbogen R	E_R
4	Handgelenk R	Ha_R
5	Schulter L	S_L
6	Ellenbogen L	E_L
7	Handgelenk L	Ha_L
8	Hüfte (Mitte)	H_M
9	Hüftgelenk R	H_R
10	Knie R	K_R
11	Fußknöchel R	F_R
12	Hüftgelenk L	H_L
13	Knie L	K_L
14	Fußknöchel L	F_L
15	Auge R	A_R
16	Auge L	A_L
17	Ohr R	O_R
18	Ohr L	O_L
19	Großer Zeh L	GZ_L
20	Kleiner Zeh L	KZ_L
21	Ferse L	FE_L
22	Großer Zeh R	GZ_R
23	Kleiner Zeh R	KZ_R
24	Ferse R	FE_R
25	Schädeldach	SD
26	Torsomitte	TM
27	Handmitte L	HaM_L
28	Handmitte R	HaM_R

OpenPose BODY_25 Modell

VisCom Erweiterungen

Abb. 49 Das BODY_25 Körpermodell von OpenPose inklusive der VisCom-Erweiterungen

Um bestimmte Funktionen im VisCom-Konzept umsetzen zu können, muss das Skelettmodell um vier zusätzliche Körpermerkmale erweitert werden:

- das Schädeldach (zur genaueren Schätzung der Abbildungsgröße und für die Erzeugung des ROIs für die Gesichtsanalyse),
- die Torsomitte (zur Unterteilung des Rumpfes in Brust- und Bauchbereich) sowie
- die Mitte der linken und rechten Hand (zur genaueren Analyse der Körperteilbeziehungen bei Berührungsgesten).

Diese Erweiterungen können durch Interpolation (Torsomitte) oder Extrapolation (Schädeldach und Handmitten) mittels der umliegenden Körpermerkmale errechnet werden. In Bezug auf Gesichtsmerkmale (*Facial Landmarks*) bietet OpenPose auch ein zusätzliches Gesichtsmodell mit 70 weiteren Merkmalen, die zur Beschreibung des Mundbilds genutzt werden können. Obwohl in diesem Konzept kein akustischer Kanal vorgesehen ist, kann das Mundbild zur besseren Modellierung der Nutzerintention dienen. Öffnet und schließt sich der Mund beispielsweise innerhalb einer bestimmten Zeitspanne, kann angenommen werden, dass der Nutzer spricht oder zumindest Laute von sich gibt.

4.5.2 Nonverbale Kommunikationsmittel

Das VisCom-Konzept baut auf der Phonologie der DGS (vgl. Kapitel 2.2.6) auf, schließt jedoch aufgrund des zweidimensionalen Ansatzes und der großen Interaktionsentfernungen einige Phoneme aus. Hierzu zählt die Handstellung (Handflächenorientierung, Fingeransatzrichtung) als auch handinterne und -externe Bewegungen. Die Mimik des Nutzers wird, mit Ausnahme des Mundbildes, in diesem Konzept ebenfalls nicht betrachtet. Abb. 50 zeigt die im VisCom-Konzept unterstützten Phoneme, mit denen Informationen übertragen werden können.

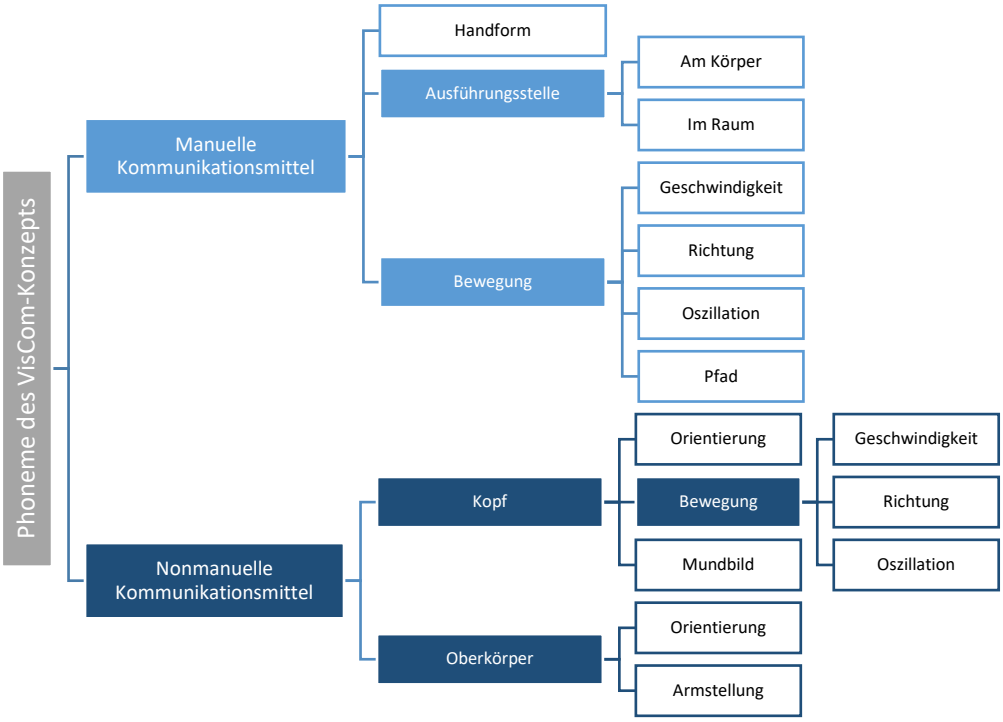


Abb. 50 Phoneme des VisCom-Konzepts (eingerahmt) gruppiert nach manuellen und nonmanuellen Kommunikationsmitteln

Für die Nutzererfassungskomponente des Kommunikationssystems wird daher für jedes Phonem ein eigener Merkmalbeschreiber $M_1 \dots M_n$ benötigt, der den jeweiligen Zustand klassifizieren kann (vgl. Implementierung in Kapitel 5.2).

Diese Klassifikationsergebnisse können als nonverbale Subsignale verstanden werden, welche, zu einem Beobachtungsvektor $\vec{B}$ zusammengefasst, das kodierte Körpersignal repräsentieren:

$$\vec{B} := \begin{bmatrix} M_1 \\ \vdots \\ M_n \end{bmatrix} \tag{1}$$

Die Aufgabe der nachgeschalteten Signalklassifizierungskomponente liegt in der Dekodierung dieses Beobachtungsvektors in ein textuelles Symbol.

4.6 Kontextsensitive Signaldekodierung mittels Motion-to-Symbol-Ansatz

In der neueren Literatur wird für die Klassifizierung nonverbaler Äußerungen vermehrt auf holistische Ansätze in Kombination mit ML-Verfahren gesetzt, bei denen alle erfassten Körpermerkmale gleichzeitig und gleichwertig in die Klassifikation einer Darstellung einfließen [329]. Dieses Vorgehen bietet den Vorteil, dass relevante Körpermerkmale nicht manuell modelliert werden müssen, sondern während der Lernphase des Klassifikationssystems in den Gewichtungsfaktoren des Modells abgebildet werden. Der große Nachteil hierbei ist jedoch, dass das Modell mit jeder neuen Darstellungsform aufs Neue trainiert werden muss und dadurch das Klassifikationsverfahren wenig flexibel ist. Daher wird in diesem Konzept ein kombinatorischer *Motion-to-Symbol*-Ansatz vorgeschlagen, bei dem die Klassifizierung nur auf der Merkmalebene stattfindet und die Dekodierung der Beobachtungen in tex-

tuelle Symbole anhand von kontextsensitiven Übersetzungsregeln erfolgt, die für jedes Nutzungsszenario (und damit auch für jedes Experiment dieser Arbeit) separat festgelegt werden können. Dieses Vorgehen ist in Abb. 51 abgebildet und bietet mehrere Vorteile.

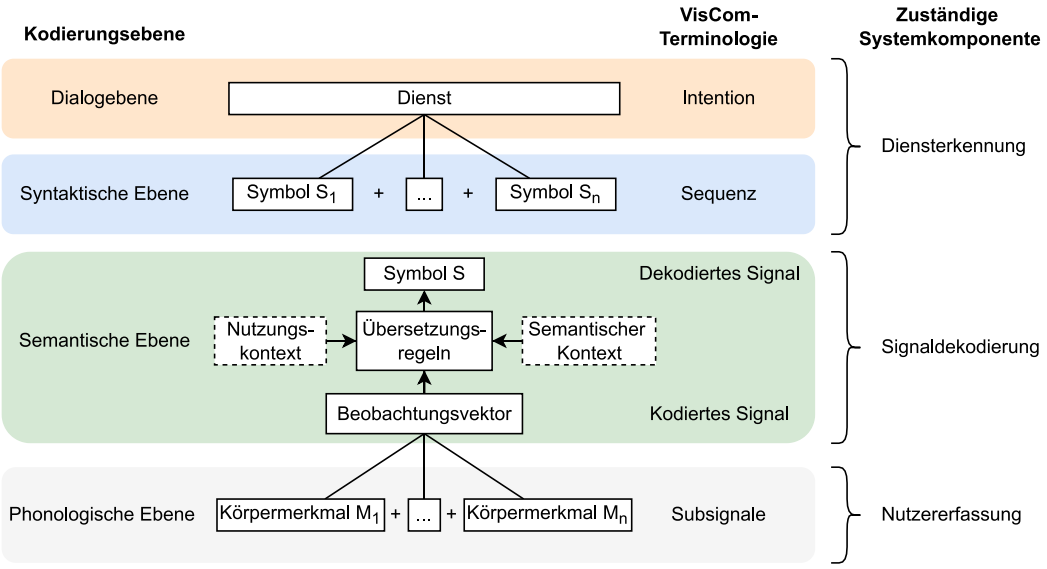


Abb. 51 Kodierungsebenen des VisCom-Konzepts

Erzeugung und Klassifizierung von Körpersignalen nach dem „Baukastenprinzip“

Der modulare Ansatz erlaubt es, unterschiedliche Klassifizierungsansätze miteinander zu kombinieren und deren Leistungsfähigkeit und Zuverlässigkeit in das Signaldesign miteinzubeziehen. Darüber hinaus wird auch die Erzeugung neuer Darstellungsformen möglich, indem bereits trainierte Beobachtungsmerkmale miteinander kombiniert werden, ohne das System insgesamt neu trainieren zu müssen. Zudem erlaubt das „Baukastenprinzip“ auch die Modellierung generischer (z.B. oszillierende Armbewegung in horizontaler Richtung) als auch spezifischer Nutzersignale (z.B. linke Arm ausgestreckt, rechte Hand offen auf Kopfhöhe, rechts vom Körper), womit die Fähigkeitslücke 2 adressiert wird.

Entkopplung der Beobachtung von der Bedeutung eines Signals

Durch den reinen beschreibenden Charakter der merkmalbasierten Klassifizierung, wird eine Entkopplung der Beobachtung (phonologische Ebene) von der Bedeutung der Signale (semantische Ebene) erreicht. Dies führt einerseits dazu, dass mehrere Darstellungen (Synonyme) ein und dasselbe Signal transportieren können. Rückmeldungen können beispielsweise sowohl mit den Händen (Daumen hoch/runter) als auch über Kopfbewegungen[54] transportiert werden. In Europa ist der „Nick-Schüttel-Code“ die verbreitetste Variante zur Bejahung oder Verneinung[55] [330], sodass dieser für die Experimente genutzt werden kann. Andererseits können auch mit einer Darstellung unterschiedliche Signale transportiert werden (Wiederverwendbarkeit). Möglich macht das eine regelbasierte Dekodierung der Beobachtungen (Abb. 52) mittels anpassbarer Übersetzungsregeln, in die neben dem Nutzungskontext (Anwendungsszenario, Interaktionsmodus, Flugzustand), auch der semantische Kontext (Dialoghistorie, verfügbare Dienste) miteinbezogen wird. Anstatt die dekodierten Signale nun direkt

[54] Die durchschnittlichen Winkel beim Kopfschütteln und Kopfnicken liegen in einem Bereich zwischen 11,6 und 22,1 Grad [313, S. 223].
[55] Daneben existieren noch der Roll-Werf-Code (Ja=Kopfwackeln von Schulter zu Schulter) und der Senk-Wurf-Code (Ja= Fallbewegung des Kopfes nach vorne, Nein=Fallbewegung des Kopfes in den Nacken)

Systemreaktionen zuzuweisen, werden sie zunächst in eine textbasierte Repräsentation (Symbol) überführt und können von der nachfolgenden Diensterkennung weiterverarbeitet werden.

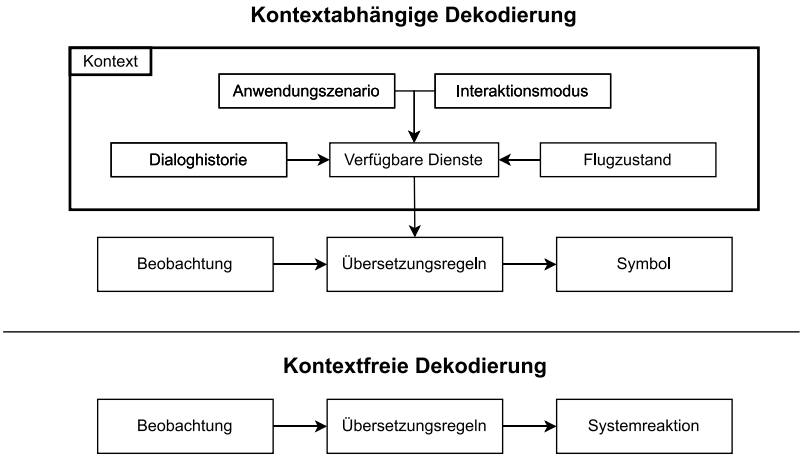


Abb. 52 Kontextabhängige Dekodierung von Beobachtungen in Symbole

Dialogunterstützungsfunktion zum Filtern von irrelevanten Nutzereingaben

Mithilfe der Übersetzungsregeln kann die Relevanz einzelner Subsignale im jeweiligen Nutzungskontext gesteuert und damit eine kontextsensitive Signaldekodierung erreicht werden, die es dem System ermöglicht, relevante von irrelevanten Nutzereingaben im jeweiligen Nutzungskontext zu unterscheiden. Diese Kontextsensitivität (KS) stellt somit die erste Dialogunterstützungsfunktion dar, mit der die Fähigkeitslücke 3 adressiert wird.

4.7 NLU-gestützte Diensterkennung

Die Erkennung des angeforderten Dienstes (und damit der Nutzerintention) anhand von textuellen Symbolen ähnelt der Problemstellung, mit der auch dialogbasierte Schnittstellen (*Conversational User Interfaces*) wie Chatbots[56] konfrontiert sind. Dabei kommunizieren Nutzer per gesprochener oder geschriebener Sprache mit einem virtuellen Gesprächspartner (Agenten), der Auskunft zu einem bestimmten Domänenwissen liefern kann, sodass der Eindruck eines Gesprächs entsteht. Bevor jedoch der virtuelle Agent die Anfrage bearbeiten kann, muss er die für den jeweiligen Anwendungszweck relevanten Nutzerabsicht zunächst identifizieren (*Intent Matching*) und die notwendigen Parameter aus der Nutzereingabe extrahieren (*Slot-Filling*). Dies kann sowohl regelbasiert (mittels regulärer Grammatik) als auch datengetrieben (mittels statistischer Verfahren) umgesetzt werden. Im Gegensatz zu regelbasierten Verfahren, können statistische Verfahren kontextabhängige Eingaben in natürlicher Sprache verarbeiten (*Natural Language Understanding*, NLU). Einen kombinierten Ansatz nutzt beispielsweise die SNIPS-Plattform [331] in der NLU-Komponente (*NLU-Engine*). Neben einem deterministischen Parser, der mit einem domänenspezifischen Datensatz (Korpus) trainiert werden kann, enthält die NLU-Engine zudem einen probabilistischen Intent Parser, der auch Syntaxvariationen erkennen kann, die nicht im Datensatz enthalten sind. In Abb. 53 ist dargestellt, wie eine solche NLU-Engine in Kombination mit einem Dialogmanager für die Diensterkennung genutzt wurde.

[56] Kunstwort aus „Chat“ und „Robot“

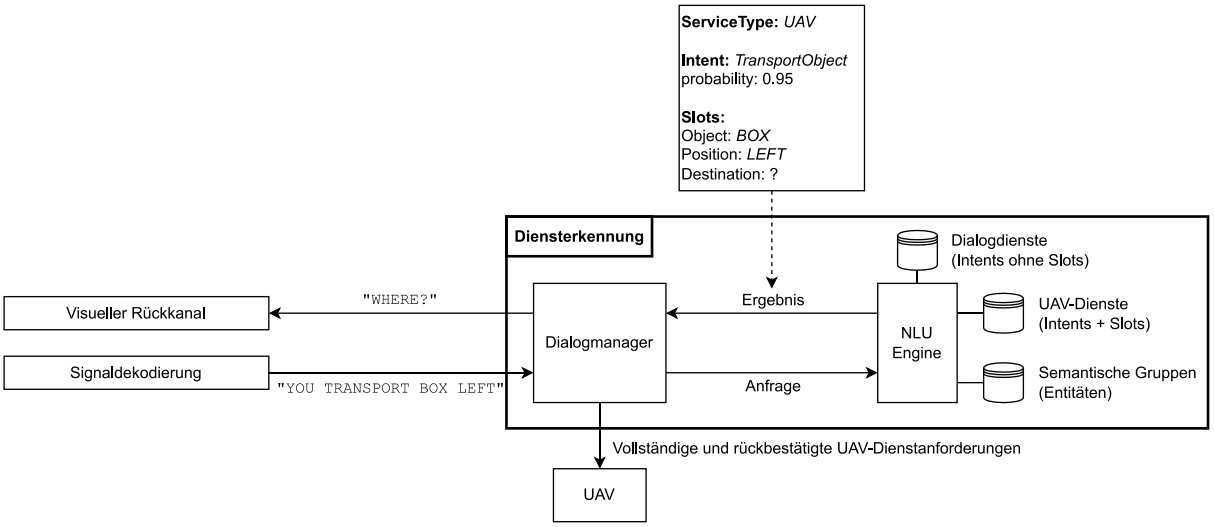


Abb. 53 NLU-gestützte Erkennung von UAV- und Dialogdiensten

Die UAV-Dienste können hierbei als Intents mit den notwendigen Parametern als Slots abgebildet werden. Jeder Slot besteht dabei aus dem Parametertyp (Entität) sowie dem eigentlichen Wert. Dialogdienste hingegen benötigen keine Slots und können durch ein einzelnes Schlüsselsymbol (z.B. „RESET“) abgebildet werden. In Kombination mit einem Dialogmanager und einer einfachen Dialogstruktur kann die Diensterkennung als ein domänenspezifisches, aufgabenbasiertes Dialogsystem verstanden werden, welches auf Nutzereingaben reagieren und damit den Verlauf der Interaktion beeinflussen kann. Hiermit wird also direkt auf die identifizierte Fähigkeitslücke 1 Bezug genommen.

4.8 Modellierung einer symbolbasierten Mensch-UAV-Verkehrssprache

In mehrsprachigen Regionen der Welt, in denen Sprecher mit unterschiedlichen Muttersprachen miteinander in Kontakt treten, wird oftmals auf eine gemeinsame Verkehrssprache als Verständigungsmittel zurückgegriffen. Wenn jedoch eine gemeinsame Sprachgrundlage fehlt, bedienen sich Menschen verschiedener Strategien, um die Komplexität ihrer eigenen Sprache zu reduzieren, um die Verständigung für das Gegenüber zu erleichtern. Eine dieser Sprachvariationen ist als *Xenolekt* oder „Foreigner Talk“ [332] bekannt. Zu den Hauptmerkmalen dieser Sprachveränderung zählen phonologische, morphologische, semantische und lexikalische Modifikationen, die sich in Auslassungen (von Artikeln, Endungen, Wörtern), häufigem Infinitivgebrauch und einer verkürzten Syntax mit wenigen Wörtern äußern [333, S. 226]. Darüber hinaus ist das verwendete Lexikon zumeist limitiert und die Themenauswahl eingeschränkt, sodass vorwiegend ein Frage-Antwort-Handlungsmuster mit vielen Verständigungssicherungen und -rückmeldungen genutzt wird [334]. Eine solche stark vereinfachte Struktur bietet die ideale Grundlage für eine nonverbale Mensch-UAV-Verkehrssprache, weil sie aufgrund der Nähe zu einer gesprochenen Sprache und der geringen grammatikalischen Komplexität einfach zu erlernen ist, gleichzeitig aber durch die kompakte Beschreibungsform eine effiziente und aufgabenzentrierte Kommunikation mit möglichst wenigen Nutzereingaben ermöglicht.

Verfahren zur Reduktion der grammatikalischen Komplexität

Bittet man Menschen, auf natürliche Weise mit einem kleinen Multikopter zu interagieren, ähneln die Verhaltensmuster der Interaktion mit Haustieren oder kleineren Menschengruppen, denen man kurze verbale und nonverbale Kommandos gibt [156, 173]. In einer vorausgegangenen explorativen Proban-

denstudie [219] wurden solche UAV-bezogenen Sprachkommandos gesammelt und deren grammatikalische Komplexität mittels eines *Natural Language Processing* (NLP) gestützten Verfahrens systematisch vereinfacht. In Abb. 54 sind die wesentlichen Verarbeitungsschritte dieses Verfahrens dargestellt.

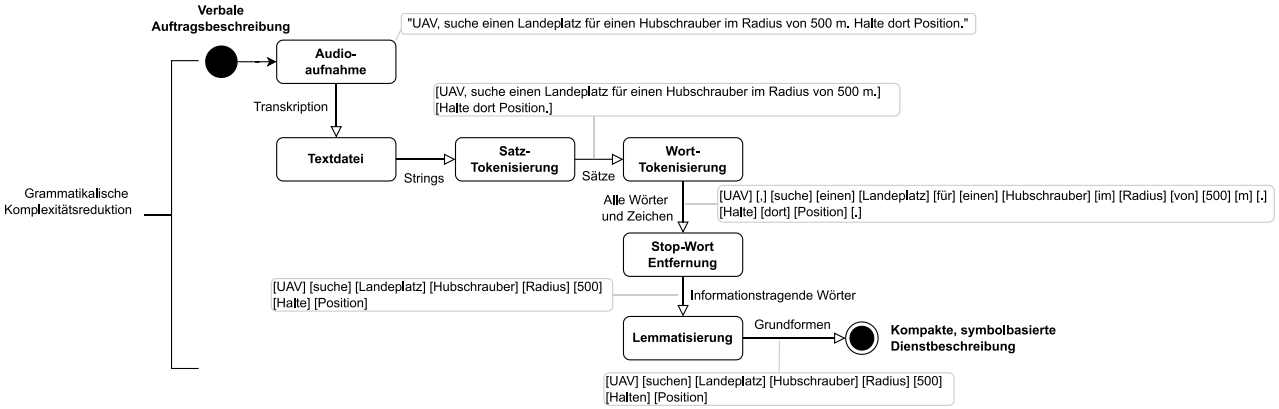


Abb. 54 NLP-basiertes Verfahren zur grammatikalischen Komplexitätsreduktion von verbal formulierten UAV-Auftragsbeschreibungen

Zunächst wurden die verbal formulierten Auftragsbeschreibungen transkribiert und in eine textuelle Form überführt, um anschließend rechnergestützt weiterverarbeitet werden zu können [335]. Nach einer Aufteilung der Beschreibungen in einzelne Sätze (Satz-Tokenisierung) und Worte (Wort-Tokenisierung), erfolgte im Anschluss die Entfernung nicht informationstragender Wörter (Stop-Wort-Entfernung) sowie die Lemmatisierung, d.h. die Überführung der Wörter in ihre Grundform (Lemma). Das Resultat dieses Standardverfahrens aus der Computerlinguistik kann als kompakte Beschreibung eines UAV-Dienstes verstanden werden. Nach diesem Verfahren wurden beispielsweise auch die zu übermittelnden Dienstbeschreibungen im Experiment C erzeugt (Kapitel 6.4).

Tab. 11 In einer Voruntersuchung ermittelte Dienstbeschreibungen nach der Komplexitätsreduktion

UAV-Aktion[57]	Anzahl Lemmata		Beispiele für Dienstbeschreibungen
	Mittelwert	Min	(Sätze durch Semikolon getrennt)
	SD	Max	
A1	2.90	2	*fliegen Haus*
	1.45	7	*fliegen Richtung 10 Uhr Haus; halten Position*
A2	4.15	2	*scannen Baum*
	2.62	12	*Drohne fliegen Richtung 10 Uhr 20 Meter Baum; scannen Baum; halten Position*
A3	3.15	2	*führen Haus*
	1.46	7	*UAV finden Weg Labyrinth; fliegen Weg Haus*
A4	3.45	2	*folgen Paket*
	1.88	10	*UAV fliegen rechts 2 Meter Gegenstand; greifen Gegenstand; nehmen; folgen*
A5	6.00	2	*suchen Hubschrauberlandeplatz*
	1.97	12	*UAV aufklären Umkreis 300 Meter; Gebiet bewerten; Gebiet gut bewerten entfernen Gebiet*
A6	6.90	4	*transportieren Gegenstand Haus; landen*
	2.91	15	*UAV fliegen rechts 2 Meter; nehmen Gegenstand; fliegen Haus; legen Gegenstand ab; kommen zurück; landen*

Für die Beschreibung der insgesamt sechs vorgestellten UAV-bezogenen Aktionen (Tab. 11, Details in Fußnote), nutzte ein Großteil der Probanden bei der Voruntersuchung ein bis zwei Sätze mit jeweils zwei bis drei Wörtern. Die dabei bevorzugte Syntax orientiert sich am Subjekt-Verb-Objekt Wortstellungstyp und kann wie folgt beschrieben werden:

```
ADRESSAT + AKTION + [PARAMETER 1 … PARAMETER N] + RÜCKBESTÄTIGUNG
```

Der mögliche Aufbau einer nonverbalen Eingabesequenz zur Beschreibung eines UAV-Dienstes besteht somit aus einem Adressaten (falls mehrere UAVs zur Verfügung stehen), gefolgt von einer Aktion und zwei bis drei dienstspezifischen Parametern. Nach dem Zurücklesen der Eingabe durch das System, schließt diese Sequenz mit einer positiven Rückbestätigung durch den Nutzer ab.

Semantische Parametergruppen

Betrachtet man die verwendeten Parameter im Detail, so können diese zu semantischen Parametergruppen zusammengefasst werden. Neben den bereits genannten Subjekten (Adressat), Handlungsanweisungen (Aktionen) und Rückmeldungen (positiv, negativ), können in den in Tab. 11 genannten Beispielen weitere solcher semantischen Gruppen identifiziert werden:

- Objektbezeichnungen (Gegenstand, Haus, Baum, Hubschrauberlandeplatz),
- Ortsangaben (Basis, Haus, Gebiet),
- Numerischen Angaben (300 m),
- Entfernungsangaben (Umkreis, Distanz) sowie
- Richtungsangaben (rechts).

Im linguistischen Kontext können diese Gruppen als Entitäten modelliert und damit zur Beschreibung der UAV-Dienste genutzt werden.

[57] A1: Verlegung des UAVs, A2: Dokumentation eines Objektes, A3: Evakuierung des Nutzers, A4: Anheben eines Objektes, A5: Suche eines Ortes, A6: Transport eines Objekts zu einem Ziel und Rückflug zum Nutzer.

5 Umsetzung des Konzepts in ein Experimentalsystem

Dieses Kapitel beschreibt die Umsetzung des vorgestellten VisCom-Konzepts in ein echtzeitfähiges Experimentalsystem, welches für die Evaluierung der nonverbalen Mensch-UAV-Interaktion genutzt wurde. Das dafür entwickelte Softwaresystem fußt auf den Prinzipen von Kapselung, Modularität und der Delegation von Aufgaben. Abb. 55 zeigt die funktionalen Komponenten des Experimentalsystems, die einem der vier Subsysteme zugeordnet werden können:

- Nutzererfassungssystem (VisCom-Schicht 1 und 2)
- Klassifizierungssystem (VisCom-Schicht 3)
- Dialogsystem (VisCom-Schicht 4, 5 und 6)
- Flugsteuerungssystem

Die nachfolgenden Kapitel gehen auf die Implementierung der jeweiligen Software-Komponenten (SWK) detaillierter ein. Für die Interprozesskommunikation wurde das in der Robotik etablierte Framework *Robot Operating System* (ROS) gewählt [336]. Einzelne Programmmodule (*Nodes*) kommunizieren dabei mittels IP-Paketen miteinander und können voneinander getrennt betrieben und damit separat entwickelt und getestet werden. Die netzwerkbasierte Kommunikation erlaubt nicht nur die Verteilung der Rechenlast auf mehrere Hardware-Komponenten (HWK), sondern ermöglicht auch den Fernzugriff auf das Datennetzwerk über drahtlose Verbindungen. Eine Möglichkeit zur Datenaufzeichnung gehört ebenso zum Funktionsumfang wie diverse Analyse- und Visualisierungskomponenten, die für eine Systementwicklung hilfreich sind.

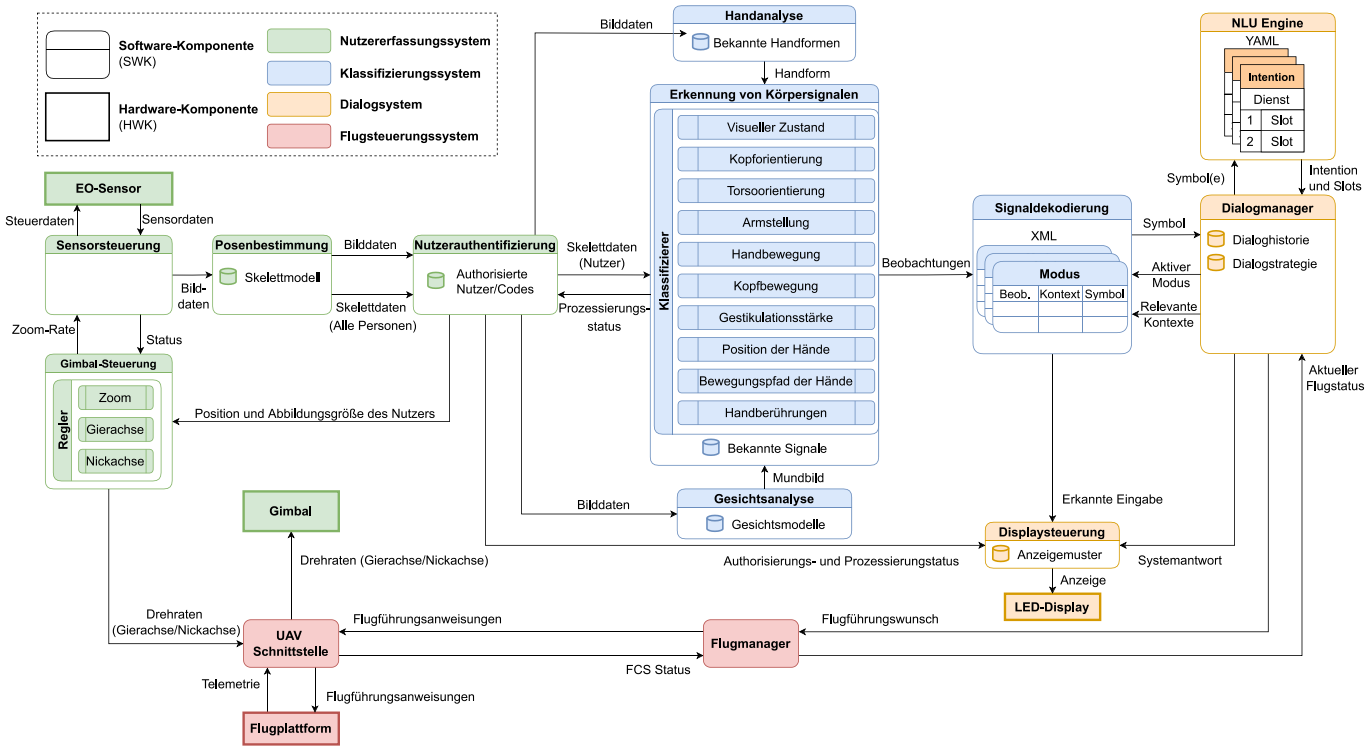


Abb. 55 Übersicht der funktionalen Hard- und Softwarekomponenten des Experimentalsystems

5.1 Nutzererfassungssystem

Das Nutzererfassungssystem bildet die perzeptive Schnittstelle zur Umwelt und stellt den anderen Subsystemen Informationen über den Nutzer in Form von Bild- und Bewegungsdaten zur Verfügung. Als Sensor wurde hierfür eine industrielle EO-Blockkamera[58] mit variabler Brennweite und einer integrierten Ethernet-Schnittstelle (*GigE Vision*) gewählt.

5.1.1 Sensorsteuerung

Die SWK *Sensorsteuerung* steuert die Parameter des EO-Sensors (Auflösung, Bildwiederholrate, Zoomrate) und wandelt dessen proprietäre Datenpakete in ROS-konforme Sensornachrichten um. Diese werden, zusammen mit den aktuellen Sensorparametern wie der aktuellen Vergrößerungsstufe (Zoomfaktor), allen anderen SWK im Netzwerk zur Verfügung gestellt.

5.1.2 Posenbestimmung

Die SWK *Posenbestimmung* abonniert Sensornachrichten und liefert Gelenkdaten für die im Bild erkannten Personen. Als Convolutional Pose Machine wurde die von die von den Autoren veröffentlichte Implementierung des OpenPose-Detektors in die ROS-Umgebung integriert [337]. Dieser liefert neben einem normierten X/Y-Koordinatenpaar auch einen Konfidenzwert für jedes Gelenk des Skelettmodells. Dieser Parameter wird in mehreren Algorithmen des Experimentalsystems genutzt, um beispielsweise die Sichtbarkeit von Körperteilen zu überprüfen, was jedoch nicht ganz unproblematisch ist. Wie bei allen Mustererkennungsverfahren, kann auch das hier verwendete CNN-basierte Posenbestimmungsverfahren durch optisch ähnliche Strukturen getäuscht werden. Ebenso kann eine bewusste Manipulation des Eingangssignals – bspw. durch die Platzierung eines präparierten Musters im Erfassungsbereich des Sensors (sog. *Adversarial Attacks*) – dazu führen, dass trotz eines falschen Klassifikationsergebnisses ein hoher Konfidenzwert ausgeben wird [338, 339] oder überhaupt keine Detektion stattfindet [340]. In diesem Kontext durchgeführte Voruntersuchungen, sowohl unter Laborbedingungen als auch in Feldversuchen, haben jedoch ergeben, dass OpenPose zumindest gegenüber natürlichen Bildstörungen (Hintergrundobjekten wie Büsche, Bäume, Gräser, etc.) ein robustes Verhalten aufweist, sofern bestimmte Parameter eingehalten werden. Hierzu zählt die gewählte Auflösung des Detektornetzwerkes als auch die Abbildungsgröße einer Person im Sensorbild.

Die Auflösung des Detektornetzwerkes (im Folgenden als Netzauflösung abgekürzt) stellt eine wichtige Stellschraube zur Parametrisierung des Detektors dar und hat einen wesentlichen Einfluss sowohl auf die Detektionsleistung als auch auf die Inferenzdauer und damit die Performanz des Detektors. Eine Verringerung der Netzauflösung führt zwar zu einer Verkürzung der Inferenzzeit, verschlechtert jedoch auch die Lokalisationsgenauigkeit und erfordert für eine korrekte Detektion eine größere Abbildung einer Person im Sensorbild.

[58] Sony FCB-EH6300, max. 30 B/s, bei einer Auflösung von 1920x1080 Pixeln und einer 20-fachen optischer Vergrößerung.

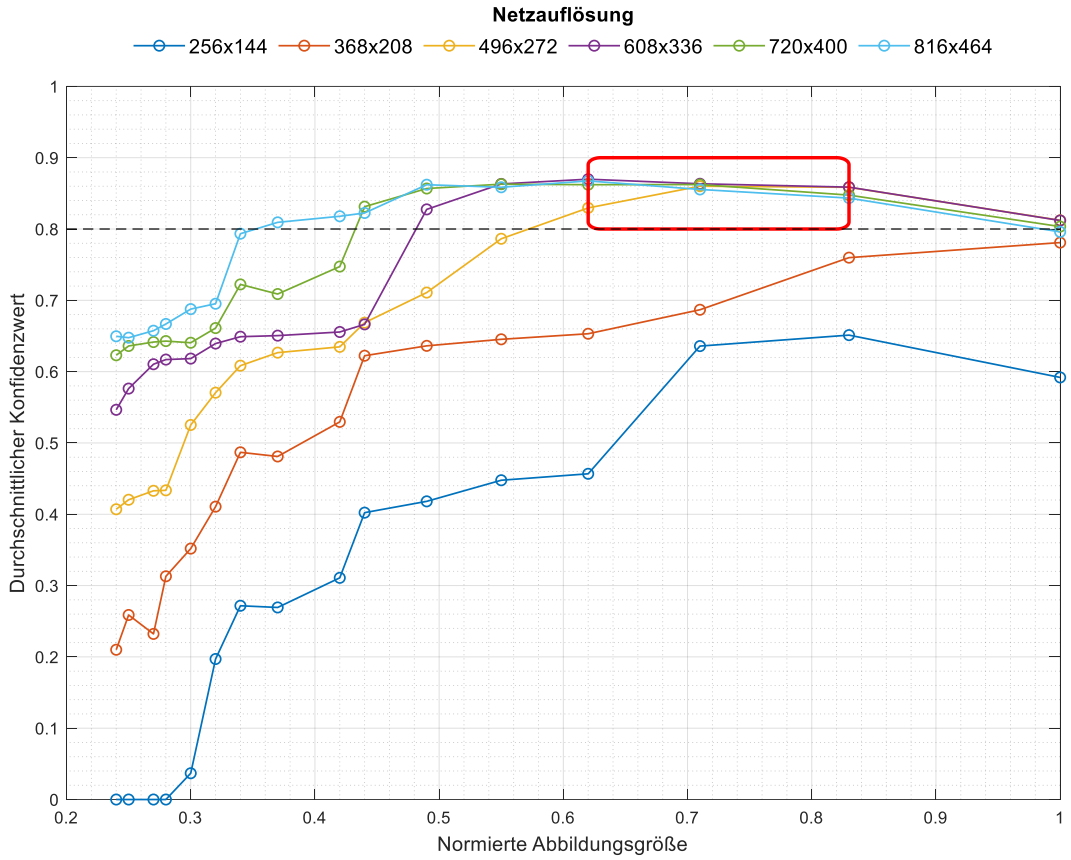


Abb. 56 Experimentell ermittelte Detektionsleistung von Openpose in Abhängigkeit von der Abbildungsgröße einer Person und der gewählten Auflösung des Detektornetzwerkes (**gestrichelte Linie** = gewähltes Mindestkonfidenzniveau, **rote Markierung** = optimaler Abbildungsbereich)

Um den optimalen Abbildungsbereich für ein möglichst hohes Konfidenzniveau bei gleichzeitig niedriger Netzauflösung zu bestimmen, wurde die Detektionsleistung von OpenPose in einem Vorversuch experimentell charakterisiert. Dazu wurden in einem Laboraufbau zunächst mehrere Aufnahmen einer Person in unterschiedlichen Abbildungsgrößen aufgezeichnet und anschließend in mehreren Durchläufen mit verschiedenen Netzauflösungen die Posenbestimmung durchgeführt. Aus den ermittelten Konfidenzwerten der erkannten Gelenkpunkte wurde im Anschluss der Mittelwert bestimmt. Abb. 56 zeigt die ermittelten Konfidenzverläufe von OpenPose in Abhängigkeit von der gewählten Netzauflösung und der Abbildungsgröße einer Person. Nimmt man ein Mindestkonfidenzniveau von 0,8 an[59], wird eine Netzauflösung von mindestens 496x272 Pixeln (gelbe Kurve) und eine Abbildungsgröße von mindestens 62 % der Bildhöhe benötigt. Interessanterweise sinkt die Konfidenz wieder ab einer Abbildungsgröße von etwa 83 % der Bildhöhe bei der Mehrzahl der getesteten Netzauflösungen, sodass eine Vollbilddarstellung einer Person aus der Sicht der Posenbestimmung in diesem Fall sogar nachteilig ist. Der optimale Abbildungsmaßstab einer Person liegt daher im Bereich von ca. 60 bis 80 % der Bildhöhe (rote Markierung).

Vorverarbeitung und Normalisierung

Aufgrund von unterschiedlichen Störeffekten im Signalweg (Abbildungsfehler im optischen System, Sensorrauschen, numerische Ungenauigkeiten), sind die von OpenPose gelieferten Positionsdaten mit einem unterschiedlich starken Rauschen behaftet. Um die hochfrequenten Signalanteile zu reduzieren,

[59] Empirisch ermittelter Wert, der eine für das Experimentalsystem ausreichende Lokalisationsgenauigkeit bei vertretbarem Signalrauschen zulässt.

wurde ein einfacher gleitender Mittelwertfilter mit einem Faktor von $F = 0.4$ gemäß folgender Vorschrift als Tiefpassfilter $f_{TPF}(x)$ implementiert:

$$f_{TPF}(x) = f(x_{-1}) \cdot F + x(1 - F) \tag{2}$$

Um invariant gegenüber der Positionierung des Nutzers im Bildausschnitt und dessen Abbildungsgröße zu sein [341], wurden die Gelenkdaten aus dem Bildkoordinatensystem in ein nutzerzentriertes Koordinatensystem mit dem Nacken als Ursprung überführt und die Entfernung von der Hüfte bis zum Nacken als normierte Referenzlänge $L_T = 1.0$ definiert.

5.1.3 Nutzerauthentifizierung

Die Aufgabe der SWK *Nutzerauthentifizierung* besteht darin, aus der Menge von sichtbaren Menschen im Erfassungsbereich der Sensorik, den autorisierten Nutzer zu identifizieren und dessen Position im Sensorbild allen anderen SWK zur Verfügung zu stellen. Für die Identifikation eines autorisierten Nutzers wurden zwei Verfahren implementiert, die eine physische Übertragung der Zugriffsrechte ermöglichen und keine weiteren Nutzerdaten benötigen (z.B. eine Datenbank mit Gesichtern), was für Untersuchungen mit Probanden im Hinblick auf den Datenschutz relevant ist. Zu den implementierten Verfahren zählt eine farbbasierte Warnwestenerkennung (WWE) sowie eine Code-basierte Ausweiserkennung (AE), welche kaskadiert betrieben oder auch komplett deaktiviert werden können. Beide Verfahren nutzen das aktuelle Sensorbild und die Gelenkdaten aller darin sichtbaren Personen als Datenbasis für die Klassifikation. Falls kein Verfahren ausgewählt wird, erhält die Person mit dem höchsten Konfidenzniveau (in der Regel die Person, die sich am Nächsten zum System befindet) die Zugriffserlaubnis auf das System. Andernfalls müssen alle in der Konfigurationsdatei festgelegten Verfahren eine positive Rückmeldung liefern, damit einer Person Zugriffsrechte gewährt werden. Ein gleichzeitiger Zugriff mehrerer autorisierter Nutzer auf das System ist im Systemkonzept nicht vorgesehen.

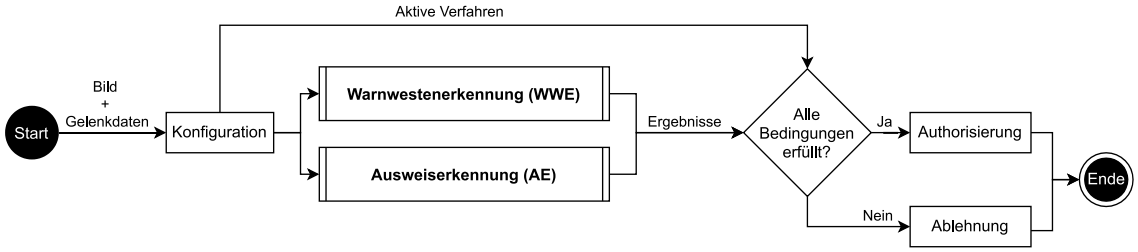


Abb. 57 Ablaufdiagramm zur Nutzerauthentifizierung

Warnwestenerkennung

Das Verfahren für die Authentifizierung mittels Warnweste nutzt vier Referenzpunkte[60] für die Erzeugung des ROI, welcher sich im Bereich des Oberkörpers befindet. Nach einer Farbraumtransformation in den HSV-Farbraum wird der Mittelwert des Farbtonkanals[61] berechnet und mit einem Sollwert[62] verglichen. Findet eine Übereinstimmung statt, liefert der Algorithmus eine positive Rückmeldung. Das Verfahren ist in Abb. 58 (links) dargestellt.

[60] S_L, S_R, H_L, H_R.
[61] Der Farbtonkanal des HSV-Farbmodells bietet eine höhere Robustheit gegenüber Helligkeitsschwankungen im Eingangssignal, beispielsweise bei wechselnder Bewölkung.
[62] H = 55 ± 15.

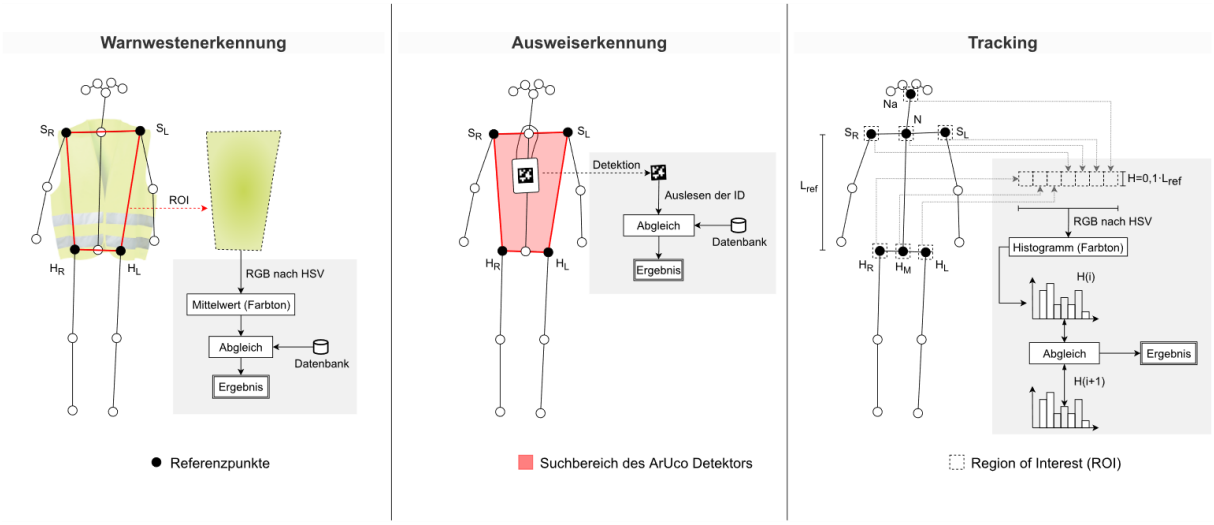


Abb. 58 Visualisierung der implementierten Authentifizierungsverfahren mittels Warnweste (links) und Ausweis (Mitte) sowie des farbbasierten Trackingverfahrens (rechts)

Ausweiserkennung

Für die Authentifizierung mittels eines Ausweises werden dieselben vier Referenzpunkte genutzt, um den Suchbereich für den Detektor für ArUco-Codes[63] zu definieren (Abb. 58 Mitte). Mit der Eingrenzung des Suchbereichs soll zum einen die Performanz gesteigert und zum anderen die Zuordnung eines präsentierten Ausweises zu einer Person sichergestellt werden. Die Ausweiserkennung hat sich in Labor- und Realflugversuchen als ausreichend robust gegenüber äußeren Störeinflüssen gezeigt und beanspruchte wenig Systemressourcen. Im Gegensatz zur farbbasierten Warnwestenerkennung können bei diesem Verfahren mehrere Codes mit unterschiedlichen Interaktionsrechten im System hinterlegt werden.

Tracking

Um die Rechenlast weiter zu senken und auf einen erneuten Aufruf der aktiven Authentifizierungsverfahren in jeder neuen Sensornachricht zu verzichten, wurde das Verfahren um eine Nutzerwiedererkennungsroutine (Tracking) erweitert. Dazu werden nach einer erfolgreichen Authentifizierung sieben Messpunkten[64] zu einem RGB-Bild zusammengefasst, in den HSV-Farbraum überführt und ein Histogramm H der Farbtonverteilung erstellt (Abb. 58 rechts). Dieses dient als Referenz H_{ref} und wird für die Wiedererkennung des Nutzers mit Hilfe der errechneten Histogramme H_i der nachfolgenden Sensorbildern genutzt. Das Verfahren meldet eine Übereinstimmung bei einer Korrelation $d(H_{ref}, H_i) \geq 0.9$, die sich wie folgt berechnet:

$$d(H_{ref}, H_i) = \frac{\sum_I \left(H_{ref}(I) - \overline{H_{ref}}\right)\left(H_i(I) - \overline{H_i}\right)}{\sqrt{\sum_I \left(H_{ref}(I) - \overline{H_{ref}}\right)^2 \sum_I \left(H_i(I) - \overline{H_i}\right)^2}} \tag{3}$$

Wobei für jedes Histogramm H_k gilt:

[63] In dieser Arbeit wurde die Implementierung von OpenCV [342] verwendet.
[64] $Na, N, S_L, S_R, H_L, H_M, H_R$.

$$\overline{H_k} = \frac{1}{N} \sum_J H_k(J) \tag{4}$$

N entspricht dabei der Gesamtzahl der Klassen (Bins) pro Histogramm. Mit diesem Tracking-Verfahren bleiben die Zugriffsrechte auch erhalten, wenn der Authentifizierungsgegenstand (z.B. der Ausweis) nach der Authentifizierung verdeckt oder weggelegt wird.

5.1.4 Gimbalsteuerung

Die SWK *Gimbalsteuerung* ist zuständig für die Aufrechterhaltung der Sichtverbindung zum identifizierten Nutzer sowie für dessen optimale Positionierung und Abbildung im Sensorbild. Dazu abonniert die SWK als Regelgrößen die aktuelle Position des Referenzpunktes P_{ref} (Nacken) als auch die Länge des Torsos L_T bezogen auf das normierte Bildkoordinatensystem.

Für jede der drei Regelgrößen (x, y, L_T) wurde ein eigener Proportionalregler implementiert, der die notwendigen Stellgrößen errechnet, um die Sollwerte zu erreichen. Als Zielvorgabe für die Positionierung wurde ein Punkt oberhalb der Sensorbildmitte gewählt ($x_{soll} = 0{,}5; y_{soll} = 0{,}4$), um die Sichtbarkeit von Handbewegungen oberhalb des Kopfes zu gewährleisten. Für P_L wurde ein Zielwert von 25 % der Bildhöhe (H) angesetzt, damit das Nutzermodell ca. 70 % der Bildhöhe einnimmt. Abb. 59 verdeutlicht die Zielvorgaben graphisch.

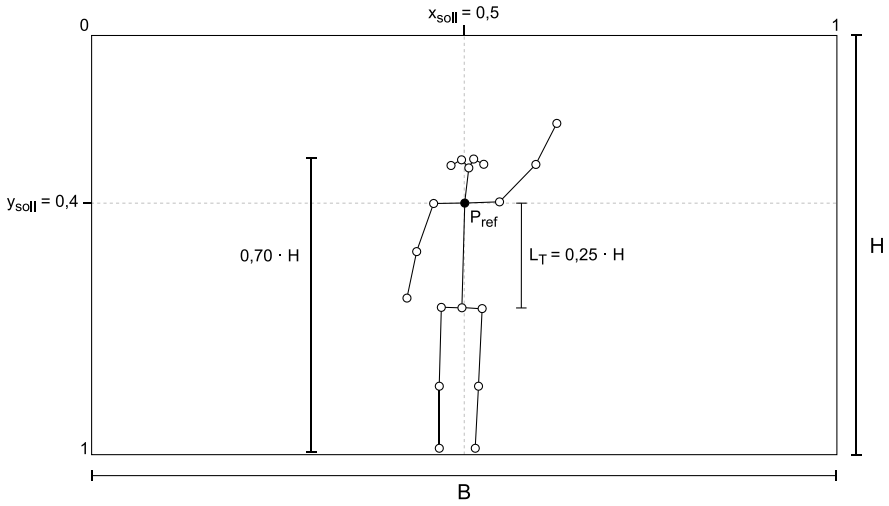


Abb. 59 Zielvorgaben für die Sensornachführung mit dem Nacken als Referenzpunkt für die X/Y-Positionierung und der Torsolänge als Referenzwert für die Abbildungsgröße

Abhängig von der aktuellen Vergrößerungsstufe Z errechnen die Positionsregler die Stellgrößen in Form von Drehratenänderungen für den Nick- und Gierwinkel des Gimbals. Um die Amplitude der Bewegungen bei zunehmender Vergrößerung zu reduzieren, wurde ein zusätzlicher Zoom-Korrekturfaktor $Z_k = 1/Z$ genutzt. Für die Vermeidung von Überschwingern und Oszillationen wurde ein Toleranzbereich T um die Sollwerte definiert. Für die Berechnung der Drehrate $r_G(x)$ um die Gierachse des Gimbalsystems wurde folgende Berechnungsvorschrift gewählt (Proportionalitätsfaktor $P_x = 12;\ T_x = 0.03$):

$$r_G(x) = \begin{cases} 0 & , \quad x_{soll} - T_x > x < x_{soll} + T_x \\ P_x \cdot (x_{soll} - x) \cdot Z_k, & sonst \end{cases} \tag{5}$$

Sowie für die Berechnung der Drehrate $r_N(y)$ um die Nickachse des Gimbalsystems ($P_y = 10$; $T_y = 0.03$):

$$r_N(y) = \begin{cases} 0 & , \quad y_{soll} - T_y > y < y_{soll} + T_y \\ P_y \cdot (y_{soll} - y) \cdot Z_k, & sonst \end{cases} \tag{6}$$

Die Regelung der Zoomrate r_Z des Sensors wurde auf ähnlichem Wege umgesetzt ($P_z = 2{,}0$; $T_z = 0{,}02$):

$$r_Z(L_T) = \begin{cases} 0 & , \quad L_{T_{soll}} - T_z > L_T < L_{T_{soll}} + T_z \\ P_z \cdot (L_{T_{soll}} - L_T), & sonst \end{cases} \tag{7}$$

Wiederherstellung der Sichtverbindung bei Sichtverlust zum Nutzer

Für den Fall eines Sichtverlustes zum Nutzer, wurde die Zoomregelung um eine Suchfunktion erweitert, bei der Regler nach einer festgelegten Wartezeit w ohne Positionsdaten vom Nutzer in einen Suchmodus übergeht und in definierten Zeitabständen t die optische Vergrößerung stufenweise um den Faktor s reduziert, um das Sichtfeld zu erweitern. Abb. 60 zeigt den Verlauf der optischen Vergrößerung vor, während und nach dem Sichtverlust. Als Parameter wurden $w = 1{,}5\ s$, $t = 0{,}5\ s$ und $s = 0{,}8$ gewählt.

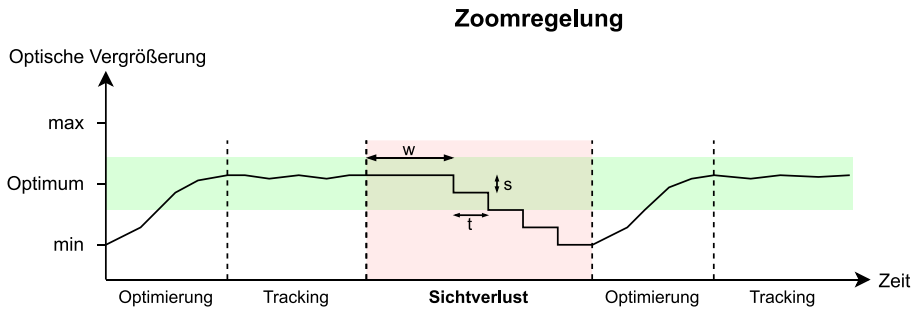


Abb. 60 Verlauf der Zoomregelung zur Optimierung der Abbildungsgröße und Aufrechterhaltung der Sichtverbindung

5.2 Klassifizierungssystem

Zur Abgrenzung von konventionellen Gestenerkennungssystemen und aufgrund der Tatsache, dass in diesem Konzept auch, neben der manuellen Komponente, auch andere Artikulatoren in kommunikativer Absicht genutzt werden können, wird diese Systemkomponente als *Klassifizierungssystem* bezeichnet. Die Aufgabe des Klassifizierungssystems ist die Generierung von Beobachtungen aus den Daten des Nutzererfassungssystems (Bild- und Skelettdaten des autorisierten Nutzers). Dafür greift es auf die in Abb. 61 abgebildeten Module zurück, um die relevanten Körpersignale (vgl. Kapitel 4.5.2) zu klassifizieren.

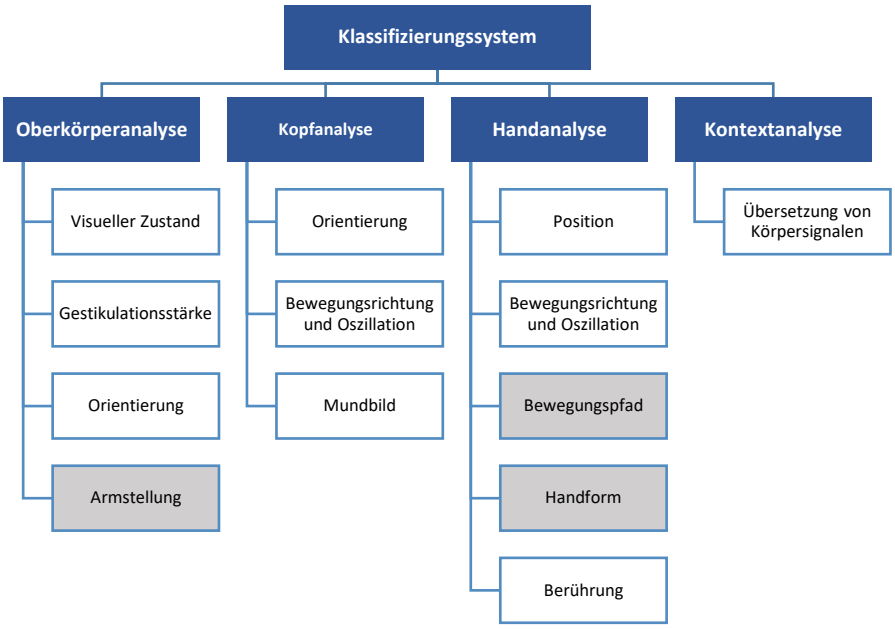


Abb. 61 Klassifikationsmodule des Klassifizierungssystems. Die grau hinterlegten Module nutzen lernbasierte ML-Verfahren, alle anderen Felder basieren auf lernfreien geometrischen oder signaltechnischen Klassifikationsverfahren.

Die jeweiligen implementierten Klassifizierungsmethoden werden in den folgenden Abschnitten im Detail vorgestellt.

5.2.1 Visueller Nutzerzustand

Für eine grobe Vorklassifizierung wurden drei Zustände definiert, die das visuelle Verhalten des Nutzers auf einer stark abstrahierten Ebene beschreiben. Der Nutzer kann sich dabei entweder in der *neutralen Pose* befinden, *gestikulieren* oder eine *statische Pose* einnehmen. Die Winkel beider Oberarme

$$W_{OA_L} = \sphericalangle(\overrightarrow{S_L N}, \overrightarrow{S_L E_L}) \tag{8}$$

$$W_{OA_R} = \sphericalangle(\overrightarrow{S_R N}, \overrightarrow{S_R E_R}) \tag{9}$$

und beider Unterarme

$$W_{UA_L} = \sphericalangle(\overrightarrow{E_L S_L}, \overrightarrow{E_L Ha_L}) \tag{10}$$

$$W_{UA_R} = \sphericalangle(\overrightarrow{E_R S_R}, \overrightarrow{E_R Ha_R}) \tag{11}$$

bestimmen zunächst, ob eine neutrale Pose eingenommen wurde (Winkeltoleranz $T = 15°$):

$$Z = \begin{cases} Neutrale\ Pose, & 90° - T \geq W_{OA_{L,R}} \leq 90° + T, 180° - T \geq W_{UA_{L,R}} \leq 180° + T \\ Gestikulation\ oder\ Pose, & sonst \end{cases} \tag{12}$$

Wird die neutrale Pose verlassen, kann anhand der zurückgelegten, mittleren Distanz ausgewählter Körpermerkmale $M = \{E_L, E_R, Ha_L, Ha_R\}$ über einen Beobachtungszeitraum $B = 10$ und einem Bewegungsschwellwert $T_G = 0{,}1 \cdot L_T$ geprüft werden, ob eine Bewegung stattfindet. Überschreitet die mittlere Distanz den Schwellwert, kann eine Gestikulation angenommen werden. Wird der Schwellwert jedoch unterschritten und befindet sich der Nutzer außerhalb der neutralen Pose, hat dieser wahrscheinlich eine (statische) Pose eingenommen. In Abb. 62 sind die verwendeten Körpermerkmale graphisch dargestellt.

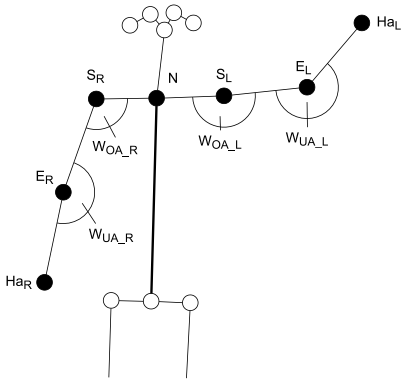


Abb. 62 Verwendete Merkmale zur groben Vorklassifizierung des Nutzerzustandes

5.2.2 Orientierung des Kopfes und Torsos

Mithilfe eines Mindestkonfidenzwertes K_{min} kann die Orientierung O eines Körperteils X, welches aus mindestens zwei symmetrisch angeordneten Merkmalen M_l, M_r besteht, bestimmt werden, wenn die entsprechenden Konfidenzwerte $K(M_l)$ und $K(M_r)$ bekannt sind:

$$O(X) = \begin{cases} frontal, & K(M_l) \geq K_{min}, K(M_r) \geq K_{min} \\ links & K(M_l) < K_{min}, K(M_r) \geq K_{min} \\ rechts, & K(M_l) \geq K_{min}, K(M_r) < K_{min} \\ unbekannt, & sonst \end{cases} \tag{13}$$

Diese *Konfidenzmethode* eignet sich, wenn kein anatomisches Modell für das jeweilige Körperteil hinterlegt ist, sodass der Konfidenzwert rasch sinkt, sobald das entsprechende Merkmal verdeckt wird. Im Experimentalsystem wurde dieses Verfahren genutzt, um die Orientierung des Kopfes anhand der Konfidenzwerte der Ohren ($K_{min} = 0{,}1$) zu bestimmen. Abb. 63 zeigt die Verläufe beider Konfidenzkurven, welche eine eindeutige Klassifizierung ermöglichen.

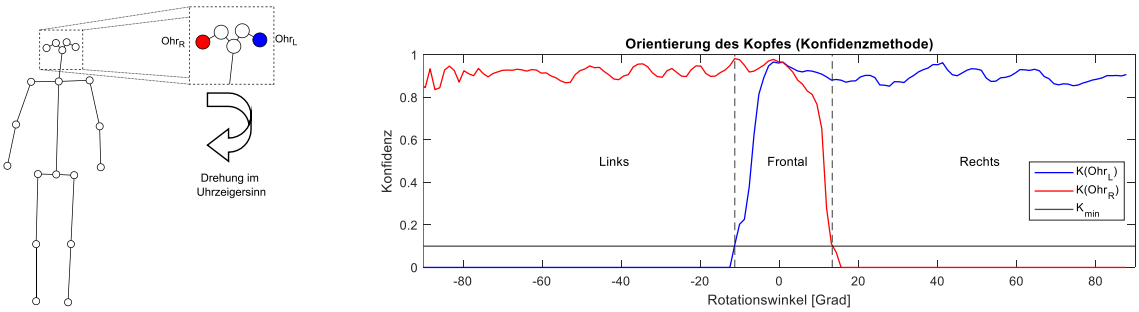


Abb. 63 Bestimmung der Orientierung des Kopfes mittels Konfidenzmethode

Für die Bestimmung der Orientierung des Torsos (bspw. anhand der Schultergelenke) eignet sich dieses Verfahren jedoch nicht. Im Modell von OpenPose ist für den Torso ein anatomisches Modell hinterlegt, sodass die Position von Körpermerkmalen auf Grundlage der angrenzenden Merkmale auch bei teilweiser oder kompletter Verdeckung geschätzt werden kann. Im oberen Diagramm von Abb. 64 sind die Verläufe der Konfidenzwerte des rechten und linken Schultergelenks abgebildet, die keine eindeutige Klassifizierung bei einer Rotation des Torsos erlauben.

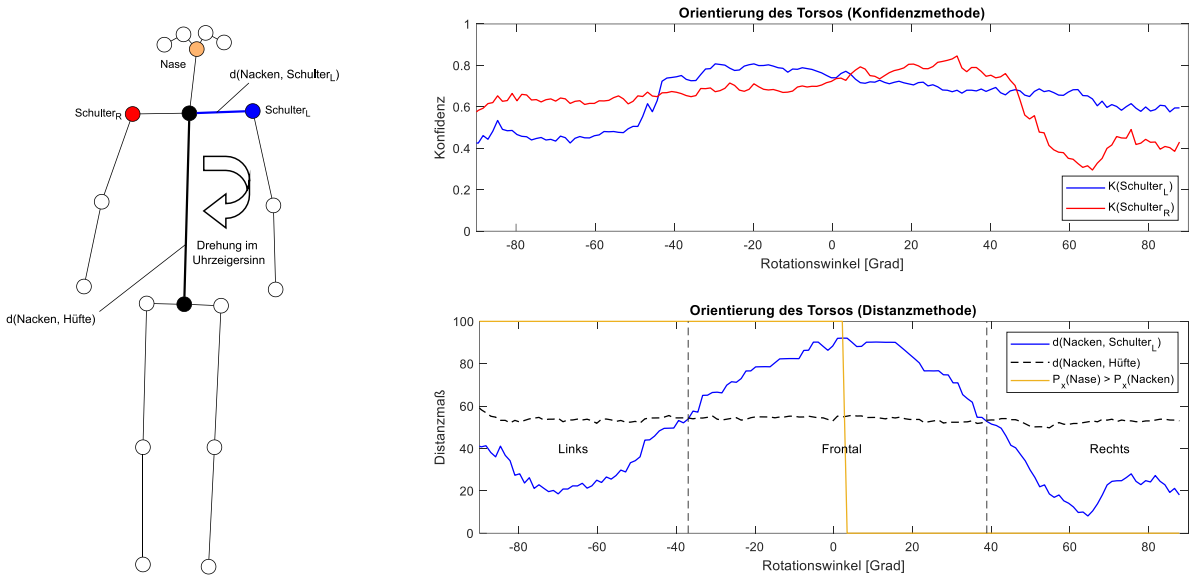


Abb. 64 Bestimmung der Orientierung des Torsos mittels Konfidenzmethode (oben) und Distanzmethode (unten)

Daher fand ein alternativer Ansatz Verwendung, der Längenverhältnisse als Grundlage für die Klassifizierung nutzt und hier als *Distanzmethode* bezeichnet wird. Bei Rotationen um die Hochachse verändert sich bei zweidimensionalen Abbildungen die Länge von vertikalen Strecken nicht, die Länge von horizontalen Strecken hingegen schon. Daher kann die Länge des Torsos vom Nacken bis zur Hüfte als dynamischer Referenzwert $L_T = d(Nacken, Hüfte)$ genutzt werden, um eine frontale Orientierung in Abhängigkeit von der euklidischen Distanz des Nackens zur linken Schulter $L_{Schulter} = d(Nacken, Schulter_L)$ zu erkennen (blaue Kurve in Abb. 64 unten). Um zusätzlich auch unterscheiden zu können, ob der Torso nach links oder rechts ausgerichtet ist, können die X-Koordinaten von Nase und Nacken als weitere Entscheidungsgrundlage dienen (gelbe Kurve in Abb. 64 unten), sodass sich folgende Rechenvorschrift ergibt:

$$O(Torso) = \begin{cases} frontal, & L_{Schulter} \geq L_T \\ links & L_{Schulter} < L_T,\ P_x(Nase) > P_x(Nacken) \\ rechts, & L_{Schulter} < L_T,\ P_x(Nase) < P_x(Nacken) \\ unbekannt, & sonst \end{cases} \tag{14}$$

5.2.3 Stellung der Arme

Für die Erkennung von bekannten Armstellungen wurde ein lernbasiertes Verfahren auf Basis zweier Multiklassen-SVM (eine für jede Körperseite) gewählt und ein Datensatz mit 15 Klassen (Tab. 12) basierend auf den Bewegungsdaten von elf Personen generiert.

Tab. 12 Der erzeugte Datensatz für statische Armstellungen

Klasse	Bezeichnung	Trainingsdaten	Validierungsdaten
1	Arm gekreuzt über Kopf	1.323	263
2	Arm gekreuzt unter Hüfte	1.382	275
3	Arm gekreuzt vor Brust	1.206	240
4	Arm hängend	1.191	237
5	Arm zeigt nach unten	1.275	254
6	Arm zeigt L-förmig nach außen	1.317	262
7	Arm zeigt L-förmig nach oben	1.364	271
8	Arm zeigt nach außen	1.218	242
9	Arm zeigt auf Brust	1.398	278
10	Arm zeigt auf Ohr	1.377	274
11	Arm zeigt auf Gesicht	1.337	266
12	Arm zeigt auf Knie	1.136	226
13	Arm zeigt auf Hals	1.290	257
14	Arm zeigt auf UAV	1.461	291
15	Arm zeigt nach oben	1.274	253
	Summe	19.549	3.889

Für das Training der SVM-Klassifikatoren[65] wurde der Datensatz im Verhältnis von etwa 80 zu 20 in einen Trainings- und Validierungsdatensatz aufgeteilt. Als Eingangssignale wurden jeweils die normierten X/Y-Koordinaten der Körpermerkmale $M_L = \{Ha_L, E_L\}$ sowie $M_R = \{Ha_R, E_R\}$ im nutzerzentrierten Koordinatensystem gewählt. Um die Anzahl an Fehldetektionen während der Transition von einer Armstellung zu einer anderen zu reduzieren, wurde der Klassifikator um einen Ausgangsfilter erweitert, der das Klassifikationsresultat erst nach einer Beobachtungszeit von $N = 10$ aufeinanderfolgenden, gleichartigen Klassifikationen weiterleitet. Andernfalls wird die Armstellung als *unbekannt* klassifiziert. Tab. 13 stellt die erreichte Klassifikationsleistung am Beispiel des Klassifikators für den rechten Arm dar.

Tab. 13 Klassifikationsleistung des Merkmals Armstellung (rechter Arm)

	Precision	Recall	Accuracy	F1-Score
Trainingsdatensatz	0,9946	0,9900	0,9900	0,9922
Validierungsdatensatz	0,9811	0,9636	0,9637	0,9722

Eine detaillierte Übersicht aller Klassenbeziehungen in Form von Konfusionsmatrizen befindet sich im Anhang A2.

[65] In dieser Arbeit wurde die SVM-Implementierung von Chang und Lin [343] genutzt. Modell: C-SVC, Kernel: RBF, Anzahl Support Vektoren: 2826.

5.2.4 Bewegungsrichtung und Oszillation des Kopfes und der Hände

Die Positionsänderungen des Kopfes und der Hände werden vom Klassifikationssystem ebenfalls analysiert und einer der in Tab. 14 dargestellten Bewegungsarten zugeordnet.

Tab. 14 Implementierte, generische Bewegungsarten

Bewegungsart	Ausprägung
Zirkulär	Im Uhrzeigersinn (clockwise, CW)
	Entgegen dem Uhrzeigersinn (counterclockwise, CCW)
Linear	Nach oben
	Nach unten
	Nach links
	Nach rechts
	Keine Bewegung
Oszillierend	Horizontal
	Vertikal
Unbekannt	-

Für die Klassifizierung der Handbewegungen[66] über die Merkmale $M = \{Ha_L, Ha_R\}$ wurde ein zweistufiger, geometrischer Ansatz implementiert. Im ersten Schritt (Abb. 65 oben) wird geprüft, ob die in einem Ringspeicher R mit der Kapazität C enthaltenen Koordinaten eine kreisförmige (zirkuläre) Bewegung über die Zeit darstellen. Dafür wird aus den Koordinaten zunächst durch Mittelwertbildung der Mittelpunkt (MP) errechnet und der euklidische Abstand zu jedem Datenpunkt bestimmt. Der minimal und maximal erlaubte Abstand vom MP wurde definiert als $T_{min} = 0{,}1 \cdot L_T$ sowie $T_{max} = 5 \cdot T_{min}$ mit einer maximalen Schwankungsbreite von $T_B = 0{,}6 \cdot (T_{max} - T_{min})$. Liegen alle Punkte innerhalb dieses Distanzbandes, wird im nächsten Schritt die Rotationsrichtung bestimmt, andernfalls handelt es sich bei dem Satz an Koordinaten um keine zirkuläre Bewegung. Die Drehrichtung wird bestimmt, indem die Anzahl der zunehmenden Winkel $N_{größer}$ mit der Anzahl der abnehmenden Winkel $N_{kleiner}$ verglichen wird. Gibt es mehr zunehmende Winkel als abnehmende, handelt es sich um eine Drehbewegung im Uhrzeigersinn, andernfalls entsprechen die Daten einer Drehbewegung entgegen dem Uhrzeigersinn.

[66] Parameter für Handbewegungen: $T_D = 0{,}025 \cdot L_T$; $K_B = 0{,}6$; $K_O = 0{,}3$

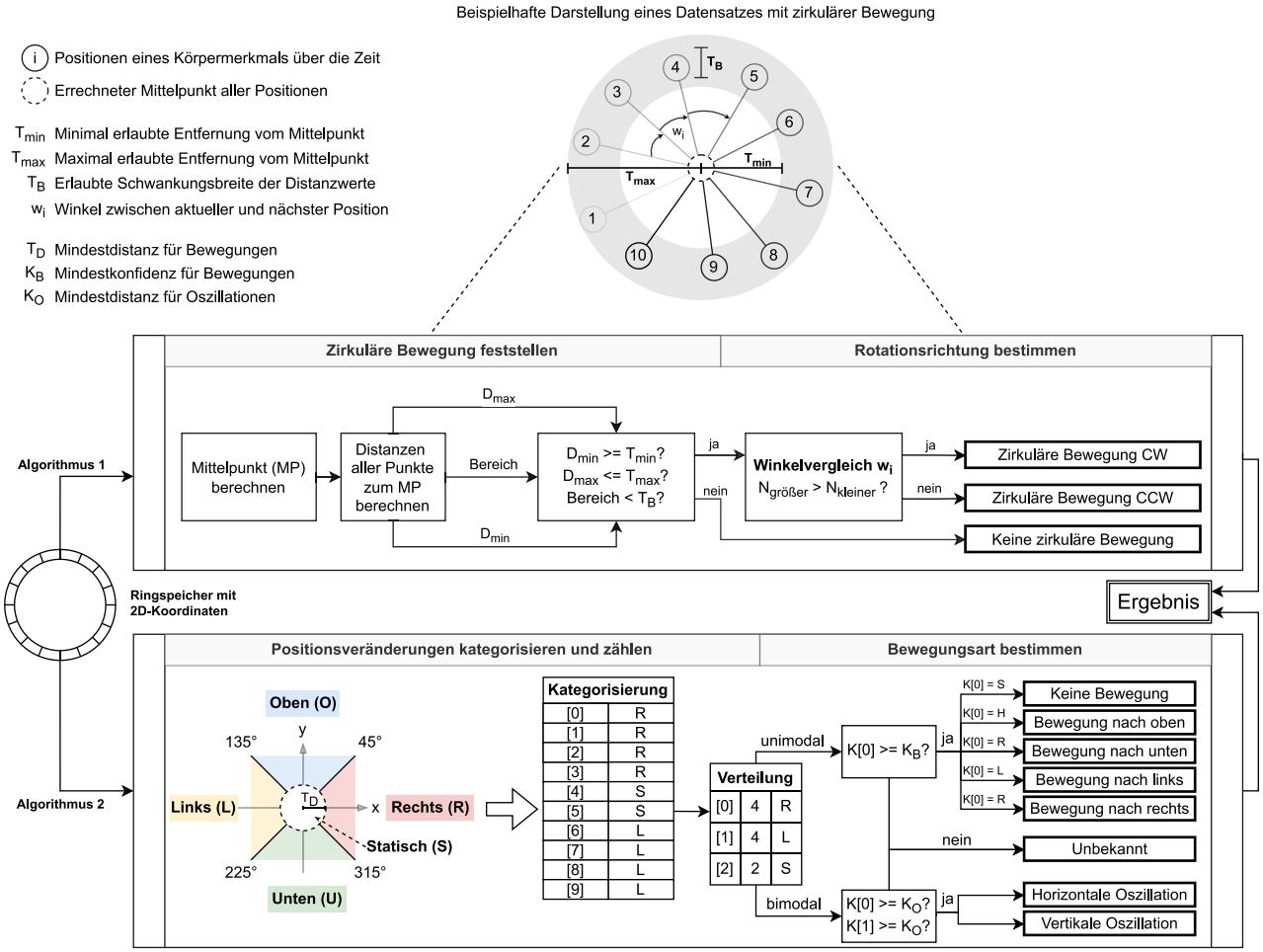


Abb. 65 Zweistufiges, geometrisches Verfahren zur Klassifikation von Kopf- und Handbewegungen anhand von 2D-Positionsdaten

Liegt keine Drehbewegung vor, prüft der zweite Algorithmus (Abb. 65 unten), ob es sich bei dem Datensatz um eine lineare oder oszillierende Bewegung handelt. Dafür wird die Positionsänderung jeder Koordinate K_i im Bezug zur vorherigen anhand des aufgespannten Winkels in fünf Kategorien eingeteilt (hoch, runter, links, rechts, statisch). Ein Datenpunkt wird dabei als statisch gekennzeichnet, wenn die Lageveränderung zum vorherigen Datenpunkt eine Mindestentfernung T_D nicht überschreitet. Aus den nun kategorisierten Koordinaten wird im nächsten Schritt eine Häufigkeitsverteilung erzeugt und in absteigender Reihenfolge sortiert. Ist die Verteilung unimodal, d.h. die Kategorie des ersten Elements tritt häufiger auf als der Mindestkonfidenzwert für Bewegungen K_B, sodass die Bedingung

$$\frac{K[0]}{C} \geq K_B \tag{15}$$

erfüllt wird, handelt es sich um eine lineare Bewegungsart. Im Falle einer bimodalen Verteilung, d.h. es existieren zwei dominierende Kategorien (z.B. $K[0] = L, K[1] = R$), muss der Mindestkonfidenzwert für Oszillationen K_O von den beiden ersten Elementen übertroffen werden und die Bedingungen

$$\frac{K[0]}{C} \geq K_O \tag{16}$$

sowie

$$\frac{K[1]}{C} \geq K_O \tag{17}$$

erfüllt sein, damit eine oszillierende Bewegung (in diesem Fall eine horizontale Oszillation) erkannt wird. Wird eine der Bedingungen (15), (16) oder (17) nicht erfüllt, wird die Bewegung als unbekannt klassifiziert. Für die Klassifizierung der Kopfbewegungen[67] wurde nur der zweite Algorithmus genutzt, da kreisförmige Kopfbewegungen nicht zum Systemkonzept gehören. Als Referenzpunkt diente hierbei das Körpermerkmal Nase (Na).

5.2.5 Gestikulationsstärke

Um die Gestikulationsstärke G zum Zeitpunkt i beschreiben zu können, wurden die Merkmale aus (8), (9), (10) und (11) zu einem Beobachtungsvektor zusammengefasst

$$\vec{W}(i) = \begin{pmatrix} W_{OA_L} \\ W_{OA_R} \\ W_{UA_L} \\ W_{UA_R} \end{pmatrix} \tag{18}$$

und über die Zeit in Form eines Ringspeichers R mit der Kapazität C (= Beobachtungszeitraum) analysiert. Als Bewegungsmaß $B(i)$ wurde die Summe aus den mittleren Manhattan-Distanzen der aktuellen und vorangegangenen Beobachtung berechnet und daraus wiederrum der Mittelwert gebildet:

$$B_i = \frac{1}{C}\sum_{i=1}^{C}\frac{1}{4}\left|\vec{W}(i) - \vec{W}(i-1)\right| \tag{19}$$

Als gestikulationsverstärkender Faktor wurde zusätzlich das Mundbild herangezogen, welcher das Bewegungsmaß um einen festen Betrag $B_i = B_i + 5$ erhöht, sobald der Mund während einer Beobachtung als offen klassifiziert wird. Anhand zweier Schwellwerte $T_{mittel} = 0{,}5 \cdot T_{hoch}$ und $T_{hoch} = 6{,}75$ kann so eine Klassifizierung der Gestikulationsstärke erfolgen:

$$G(i) = \begin{cases} hoch, & B_i \geq T_{hoch} \\ mittel & T_{mittel} \leq B_i < T_{hoch} \\ niedrig, & sonst \end{cases} \tag{20}$$

5.2.6 Mundbild

Über das Mundbild können Rückschlüsse darüber gezogen werden, ob der Nutzer schweigt, spricht oder lautstark ruft. Wang et al. entwickelten eine Methode zur Erkennung von Blinkvorgängen [344], bei der sie Längenverhältnisse ausgewählter Augenmerkmale für die Berechnung eines *Eye Aspect Ratio* (EAR) nutzen, um den Öffnungszustand eines Auges zu beschreiben (Abb. 66).

[67] Parameter für Kopfbewegungen: $T_D = 0{,}012 \cdot L_T$; $K_B = 0{,}6$; $K_O = 0{,}3$.

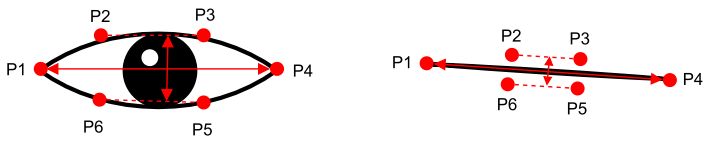


Abb. 66 Erkennung von Blinkvorgängen anhand von Längenverhältnissen am Auge (links offen, rechts geschlossen)

Dieses Verfahren lässt sich auf den Mund übertragen und wurde in dieser Arbeit zur Berechnung der im Folgenden als *Mouth Aspect Ratio* (MAR) bezeichneten Metrik verwendet (Abb. 67).

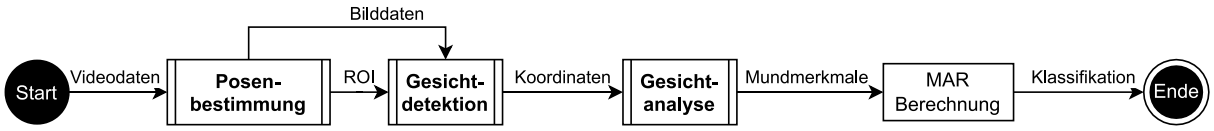


Abb. 67 Algorithmus zur Klassifikation des Mundbildes

Dazu wurde aus den Daten der Posenbestimmung die Position des Gesichtes geschätzt und mittels eines zusätzlichen Gesichtsdetektors[68] genauer lokalisiert. Im darauffolgenden Schritt wurden mit der Methode von [346] und der Implementierung von [345] 68 Gesichtsmerkmale (*Facial Landmarks*) im ROI lokalisiert, von denen sich sechs für die Berechnung des MAR eignen. Diese sind in Abb. 68 abgebildet und beziehen sich auf die Merkmale auf der Oberlippe (OL_1, OL_2), Unterlippe (UL_1, UL_2) sowie die zwei äußersten Punkte der Lippen (L_L, L_R).

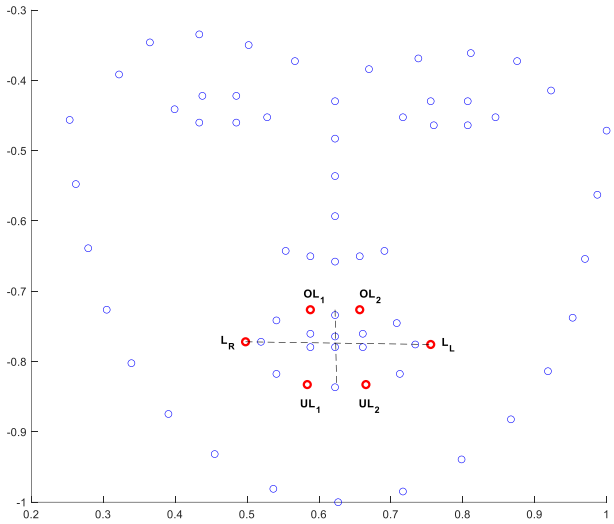


Abb. 68 Verwendete Gesichtsmerkmale für die Klassifizierung des Mundbildes anhand des MAR

Das MAR berechnet sich demnach:

$$MAR = \frac{d(OL_1, UL_1) + d(OL_2, UL_2)}{2d(L_L, L_R)} \tag{21}$$

Anhand eines Schwellwerts $T_{offen} = 0{,}15$ und einer Mindestanzahl an gleichartigen Beobachtungen N innerhalb eines Beobachtungszeitraums B kann das MAR und damit das Mundbild klassifiziert werden:

[68] In dieser Arbeit wurde die Implementierung von King [345] verwendet.

$$M(i) = \begin{cases} offen\ (1), & MAR(i) \geq T_{offen}, N = B \\ geschlossen\ (0), & sonst \end{cases} \tag{22}$$

Abb. 69 zeigt den Verlauf des MAR über die Zeit (blau Linie) sowie das klassifizierte Mundbild (orange Linie). Mit einer geeigneten Parametrisierung des Schwellwertes und der Beobachtungszeit kann mit dieser Methode auch eine visuelle Erkennung von lautem Rufen realisiert werden.

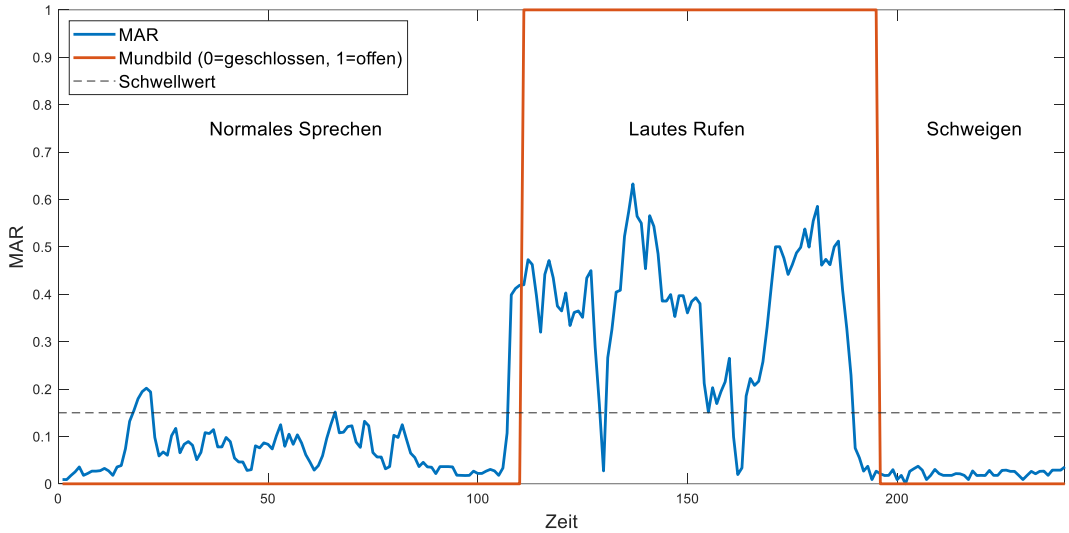


Abb. 69 Visuelle Erkennung von lautem Rufen anhand des MAR und einem Schwellwert $T_{offen} = 0{,}15$ sowie einer Beobachtungszeit $B = 10$

5.2.7 Position der Hände

Um die Position der Hände in Raum, in Relation zum Körper des Nutzers, bei statischen Darstellungen klassifizieren zu können, wurden mithilfe der Körpermerkmale $K = \{SD, S_R, S_L, TM, H_R, H_L\}$ zehn Gestenräume um und vor dem Körper (Abb. 70) definiert. Befindet sich eine der beiden Hände (genauer: eine der Handmitten HaM_R, HaM_L) länger als die gewählte Mindestverweildauer $T_{Min} = 10$ in einem der Gestenräume, wird der entsprechende Gestenraum aktiviert. Andernfalls wird die Position der Hand als *unbekannt* klassifiziert. Dadurch sollen Fehlklassifikationen, beispielsweise während einer bewegungsreichen Gestikulation, verhindert werden.

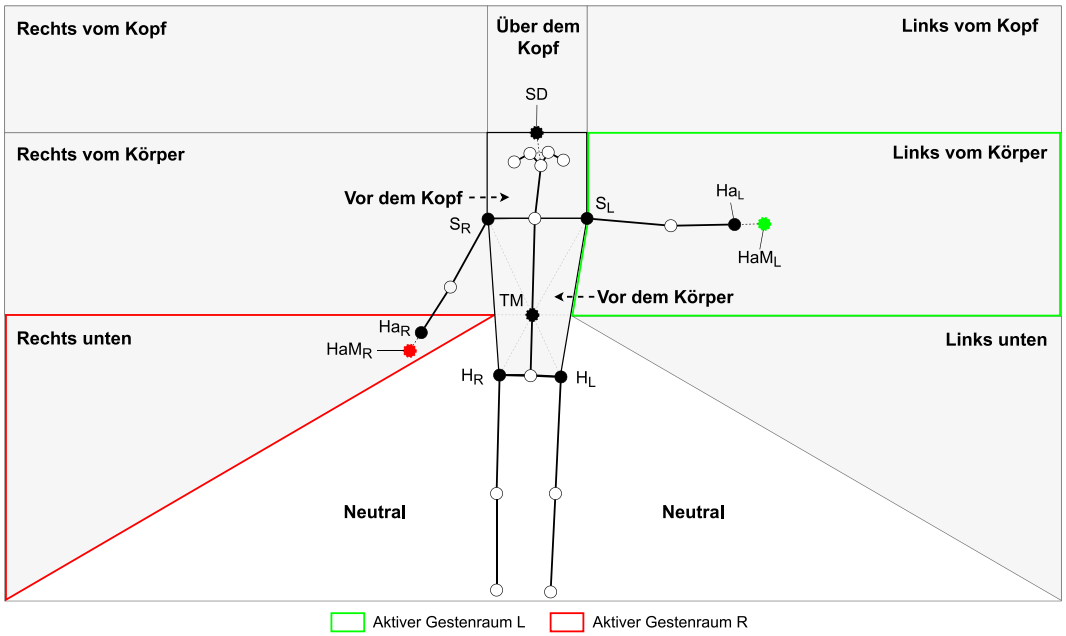


Abb. 70 Klassifizierung der Handpositionen anhand von zehn zweidimensionalen Gestenräumen

5.2.8 Bewegungspfad der Hände

Zur Beschreibung von abstrakten Konzepten kann der Nutzer geometrische Symbole mit den Händen in die Luft zeichnen, welche vom Kommunikationssystem erfasst werden. Die Erkennung nicht-periodischer Signalmuster aus unsegmentierten Datenströmen stellt besondere Ansprüche an das Klassifikationsverfahren, weil neben der Ausführungsgeschwindigkeit auch die Ausführungsgenauigkeit von Nutzer zu Nutzer variieren kann. Ein für diesen Fall geeignetes Verfahren stellt *Dynamic Time Warping* (DTW) dar. DTW hat zudem den Vorteil, dass es im Vergleich zu anderen Klassifikatoren auch bei relativ kleinen Datensätzen besser generalisiert. Für das Training des DTW-Klassifikators[69] wurde aus den Bewegungsdaten von elf Personen ein Datensatz mit sechs Bewegungsmustern erzeugt (Tab. 15) und die Daten im Verhältnis von etwa 80 zu 20 für Training und Validierung aufgeteilt.

Tab. 15 Der erzeugte Datensatz für Erkennung von beidarmigen Zeichengesten

Klasse	Bezeichnung	Trainingsdaten	Validierungsdaten
1	Arme zeichnen Brücke	27	6
2	Arme zeichnen Kreis	27	6
3	Arme zeichnen Unendlich-Zeichen	27	6
4	Arme zeichnen Rechteck	27	6
5	Arme zeichnen Dreieck	27	6
6	Arme zeichnen Z	27	6
	Summe	162	36

Als Eingangssignale wurden jeweils die normierten X/Y-Koordinaten der Körpermerkmale $M = \{Ha_L, E_L, Ha_R, E_R\}$ im nutzerzentrierten Koordinatensystem gewählt. Tab. 16 stellt die erreichte Klassifikationsleistung dar.

[69] In dieser Arbeit wurde die DTW-Implementierung von Gillian [347] genutzt.

Tab. 16 Klassifikationsleistung für das Merkmal „Bewegungspfad“

	Precision	Recall	Accuracy	F1-Score
Trainingsdatensatz	0,9235	0,9012	0,9012	0,9054
Validierungsdatensatz	0,9167	0,8333	0,8333	0,8434

Eine detaillierte Übersicht in Form von Konfusionsmatrizen befindet sich im Anhang A3.

5.2.9 Berührung

Als weiteres beobachtbares Merkmal wurde für beide Hände das jeweils nächstgelegene Körperteil bestimmt, um Körperberührungen detektieren zu können, wenn diese für eine Mindestverweildauer von $T = 8$ eingenommen werden. Da die Posenbestimmung lediglich die Positionen einzelner Körpermerkmale bereitstellt und keine Gliedmaßen an sich, mussten diese zunächst modelliert werden. Abb. 71 zeigt in Grau hinterlegt die Streckenabschnitte, die die jeweiligen Gliedmaßen repräsentieren. Zusätzlich wurden radiale Aktivierungszonen um den Kopf, die Ohren, Knie und Fußknöchel modelliert, um auch Annäherungen an diese Körperbereiche erkennen zu können. Die Radien orientieren sich an der Länge des Torsos ($r_{Ohren,Knie} = 0{,}1 \cdot L_T$ sowie $r_{Kopf,Fußknöchel} = 0{,}15 \cdot L_T$).

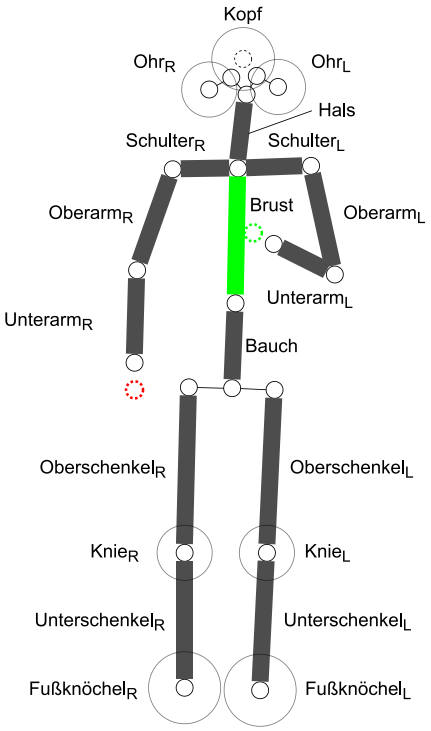


Abb. 71 Axiale und radiale Berührungsbereiche, grün hervorgehoben ist der von der linken Hand aktivierte Körperteil

Einzelne Körperpartien können auch aus der Menge an validen Gliedmaßen herausgenommen werden (z.B. die Oberschenkel), um den Anteil an Fehldetektionen zu reduzieren. Darüber hinaus findet die Klassifikation nur während des Nutzerzustandes „Pose“ statt. Andernfalls wird das Berührungsziel als *unbekannt* klassifiziert.

5.2.10 Handform

OpenPose verfügt neben einem Detektionsmodul für Körpergelenke auch über ein einen Handdetektor für Fingergelenke. Die Bestimmung der Fingerpositionen mittels dieses Hand-Modells hat sich jedoch aufgrund der hohen Hardware-Anforderungen als nicht echtzeitfähig auf den verwendeten

Missionsrechnern erwiesen, sodass ein alternativer, formbasierter Ansatz gewählt wurde. Anstatt die Gelenkpunkte jedes Fingers im Sensorbild zu bestimmen und die Handform, beispielsweise mit einer SVM, zu klassifizieren, kann auch ein Objektdetektor direkt auf den Bilddaten angewendet werden, um bekannte Handformen zu erkennen und damit eine Klassifizierung durchzuführen. Ein performanter Detektor für mehrere Klassen von Objekten wurde mit der YOLO-Architektur [235] erzeugt. Wie bei allen überwachten Verfahren des Maschinellen Lernens, ist die manuelle Generierung der für das Training benötigten annotierten Daten ein ressourcen- und zeitintensives Verfahren.

Automatisierung der Annotation

Durch eine Automatisierung konnte der Vorgang der Annotation jedoch beschleunigt werden. Der Handdetektor von OpenPose ist zwar nicht für Echtzeitanwendungen auf eingebetteten Systemen geeignet, verfügt jedoch über eine hohe Lokalisationsgenauigkeit, die in einem Offline-Annotationssystem genutzt wurde. Abb. 72 beschreibt das umgesetzte Verfahren als Blockschaltbild.

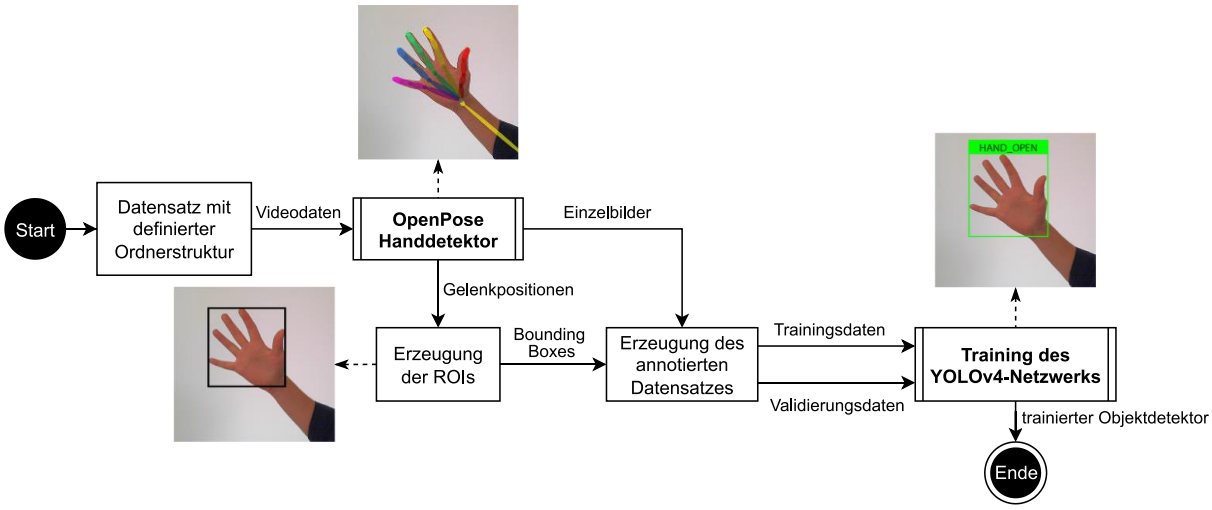


Abb. 72 Verfahren zur automatisierten Annotation von Bilddaten mit Handformen

Zunächst wurde ein Datensatz mit 16 Handformen in unterschiedlichen Umgebungsbedingungen (Labor- und Feldumgebung) von 17 Personen aufgezeichnet (Tab. 17) und die Videodaten in einer definierten Ordnerstruktur organisiert. Mithilfe des OpenPose Handdetektors[70] wurde in jedem Einzelbild der Videos die Position der Fingergelenke bestimmt und das Ergebnis als Bild-Gelenk-Datenpaar abgespeichert. Anschließend wurde aus den Extrempunkten der Gelenkdaten Begrenzungsrahmen (*Bounding-Boxes*) generiert, die für das Training des Faltungsnetzwerkes benötigt werden und zum annotierten Datensatz hinzugefügt, welcher anschließend in einen Trainings- und Validierungsdatensatz im Verhältnis von ca. 75 zu 25 aufgeteilt wurde.

[70] Konfiguration für max. Genauigkeit: $net_resolution = 1312x736,\ scale_number = 4,\ scale_gap = 0{,}25$.

Tab. 17 Übersicht des erzeugten Datensatzes für Handformen und die Detektionsleistung des implementierten Objektdetektors, im Anhang A4 befinden sich zusätzliche Abbildungen zu den jeweiligen Handformen

Klasse	Handform	Trainingsdaten	Validierungsdaten	AP in %
1	Geschlossen	2.840	1.060	98,26
2	Faust	1.470	414	99,61
3	Flach, frontal, vertikal	464	132	88,74
4	Flach, seitlich, horizontal	1.508	256	99,99
5	Flach, seitlich, vertikal	1.430	388	97,81
6	Geöffnet	3.078	1.036	99,45
7	Zeigefinger	2.016	570	99,99
8	Zeigefinger, Mittelfinger, Ringfinger, kleiner Finger	3.104	690	99,08
9	C	2.594	606	100,00
10	O	2.752	498	99,99
11	V	3.146	654	99,94
12	Daumen runter	1.802	498	99,96
13	Daumen, kleiner Finger	4.992	920	99,98
14	Daumen, Zeigefinger	2.592	668	99,89
15	Daumen, Zeigefinger, Mittelfinger	2.624	740	99,80
16	Daumen hoch	3.570	766	99,40
	Gesamt	39.982	9.896	**mAP@0.5 =** 98,87

Objektdetektoren liefern als Ergebnis im Allgemeinen einen Vektor mit den erkannten Objekten, die jeweils einen Klassenindex K, einen Konfidenzwerten C sowie die Bildkoordinaten der Bounding-Box B enthalten. Bei der Bewertung der Leistungsfähigkeit von Objektdetektoren spielt nicht nur eine korrekte Klassifizierung eine Rolle, sondern es wird auch eine möglichst genaue Lokalisierung der Objekte im Bildkoordinatensystem angestrebt. Die „mittlere Genauigkeit" (*Average Precision*, AP) wird daher in der Literatur oft als Leistungsmetrik genutzt, für dessen Berechnung zunächst festgelegt werden muss, ab wann eine Lokalisierung als korrekt bezeichnet werden kann. Für diesen Zweck bedient man sich der *Intersection over Union* (IoU), welche auf dem Jaccard-Index [348] beruht und ein Ähnlichkeitsmaß $J = [0,1]$ für zwei Mengen darstellt. Im Bereich der Objekterkennung [349] ist die IoU definiert als der Quotient aus der überlappenden Fläche (*Area of Overlap*) der wahren Objektposition B_W und der detektierten Objektposition B_D sowie der vereinten Fläche (*Area of Union*) von B_W und B_D:

$$J(B_W, B_D) = \frac{Fläche(B_D \cap B_W)}{Fläche(B_D \cup B_W)} = \frac{Area\ of\ Overlap}{Area\ of\ Union} = IoU \tag{23}$$

Abb. 73 veranschaulicht die beiden Vergleichsflächen.

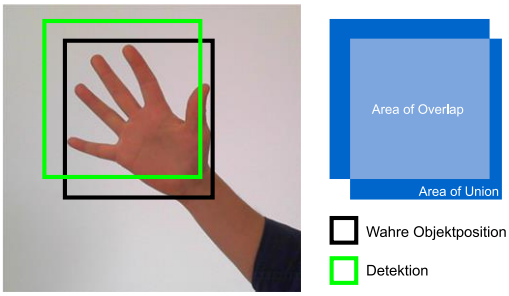


Abb. 73 Veranschaulichung der *Area of Overlap* und *Area of Union* zur Berechnung der *Intersection over Union*

Unter Zuhilfenahme eines Schwellwertes T kann mittels der Bedingung $IoU \geq T$ eine Detektion als korrekt (True Positive, TP) definiert werden, andernfalls handelt es sich um eine Falsch-Positive De-

tektion (False-Positive, FP). Basierend auf diesen beiden Parametern und der Anzahl an Falsch-Negativen Detektionen (False-Negative, FN), kann die Genauigkeit (Precision) und die Trefferquote (Recall) errechnet werden:

$$Genauigkeit = \frac{TP}{TP + FP} = \frac{TP}{Alle\ Detektionen} \tag{24}$$

$$Trefferquote = \frac{TP}{TP + FN} = \frac{TP}{Alle\ wahren\ Positionen} \tag{25}$$

Trägt man beide Werte in einem Diagramm auf, ist die Fläche unter der Precision-Recall-Kurve $P(r)$ definiert als AP:

$$AP = \int_0^1 P(r)dr$$

Um schließlich die mAP bilden zu können, wird für jede Klasse i aus der Gesamtzahl der Klassen N der AP errechnet und daraus der Durchschnittswert gebildet:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{26}$$

Der implementierte Handform-Detektor[71] auf Basis der YOLOv4-Architektur mit 162 Schichten erreichte bei einem Schwellwert von $T = 0{,}5$ nach rund 32.000 Iterationen eine mAP von 98,87 % auf dem Validierungsdatensatz.

5.2.11 Signaldekodierung

Die SWK *Signaldekodierung* abonniert alle klassifizierten Beobachtungen und übersetzt diese mittels einer Lookup-Tabelle in textuelle Symbole[72], die vom Dialogsystem verarbeitet werden. Als Parameter dienen der aktive Übersetzungsmodus (*mode*) sowie die gegenwärtig relevanten semantischen Kontexte (*contexts*). Eine Übersicht aller im Experimentalsystem verfügbarer semantischer Kontexte ist in Tab. 18 dargestellt.

[71] Für das Training des Objektdetektors wurde das Framework von Bochkovskiy et al. [350] verwendet.
[72] Die textuellen Symbole orientieren sich an Wörtern aus dem Englischen, bilden die Grammatik der Sprache jedoch nicht nach.

Tab. 18 Übersicht der im Experimentalsystem verfügbaren semantischen Kontexte

Semantischer Kontext	Beschreibung	Symbole
Actions	Handlungsanweisungen	HELP, CALL, TAKE_OFF, LAND, FLY, SCAN, FIND, LIFT, TRANSPORT, FOLLOW, ASK
Dialog	Steuerung des Dialogverlaufs	HELLO, SAY_AGAIN, IGNORE_LAST, RESET, BYE
Feedback	Rückmeldungen	YES, NO
HDirection	Horizontale Richtungsangaben	UP, DOWN, LEFT, RIGHT, FRONT, BACK, CW, CCW, NORTH, SOUTH, WEST, EAST
Interrupt	Unterbrechungen	STOP
Number	Numerische Daten	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Object	Objekttypen	BOX, TREE, HOUSE
Parameter	Parametertypen	TIME, HEIGHT, BATTERY, HUMAN
Problem	Problembehaftete Körperteile	HEAD, SHOULDER, BELLY, ARM, KNEE, LEG
Spot	Orte	HOME, AREA, SECRET, HERE
State	Betriebszustände	ON, OFF
Subject	Subjekte	ME, YOU, WE
Unit	Referenzmaß für numerische Daten	RADIUS, DISTANCE
User	Nutzeridentifikation	USER
Vehicle	Luftfahrzeugtypen	DRONE, HELI, PLANE
VDirection	Vertikale Richtungsangaben	UP, DOWN

Über eine Übersetzungsdatei können die erwarteten Körpersignale mittels der verfügbaren Merkmalbeschreiber und Mindestkonfidenzniveaus (*low/medium/high*) definiert werden. In Abb. 74 ist bespielhaft die formale Beschreibung der „Fernglas-Geste“ in XML-Notation abgebildet, die vom System im Übersetzungsmodus *BRIEFING* und dem semantischen Kontext *Action* als textuelles Symbol FIND[73] dekodiert wird. Die dabei relevanten Körpermerkmale sind die Handform (*hand_shape_o*), der aktive Gestenraum (*front_of_head*) und die Bewegungsrichtung (*steady*) der linken und rechten Hand.

```
<gesturedata name="viscom">
   <translation mode="BRIEFING">
     <context name="Action">
       <symbol name="FIND" need_all="true">
         <observation name="hand_shape_o" side="L" type="hand" confidence="high"/>
         <observation name="hand_shape_o" side="R" type="hand" confidence="high"/>
         <observation name="front_of_head" side="L" type="space" confidence="high"/>
         <observation name="front_of_head" side="R" type="space" confidence="high"/>
         <observation name="steady" side="L" type="dynamic" confidence="high"/>
         <observation name="steady" side="R" type="dynamic" confidence="high"/>
       </symbol>
     </context>
   </translation>
</gesturedata>
```

Abb. 74 Beispielhafte Beschreibung der „Fernglas-Geste“, welche im Übersetzungsmodus *BRIEFING* und dem semantischen Kontext *Action* als textuelles Symbol FIND dekodiert wird

5.3 Dialogsystem

Für die Steuerung des Dialogs wurde ein aufgabenbasiertes Dialogsystem implementiert, welches aus den drei SWK *Dialogsteuerung*, *Natural-Language-Processing* und *Displaysteuerung* besteht.

[73] Um die Lesbarkeit zu erleichtern, werden die vom VisCom-System erfassten und dekodierten Signale (= textuelle Symbole) in Versalschrift dargestellt.

5.3.1 Dialogmanager

Die SWK *Dialogmanager* bildet das zentrale Element des Dialogsystems, welches die Dialogstrategie sowie den bisherigen Dialogverlauf enthält und über das Klassifizierungssystem Nutzereingaben in Form von textuellen Symbolen entgegennimmt. Über die SWK *Natural Language Processing* kann der Dialogmanager auf die verfügbaren UAV Dienste zugreifen und mittels der SWK *Displaysteuerung* Systemantworten über den visuellen Rückkanal an den Nutzer übermitteln.

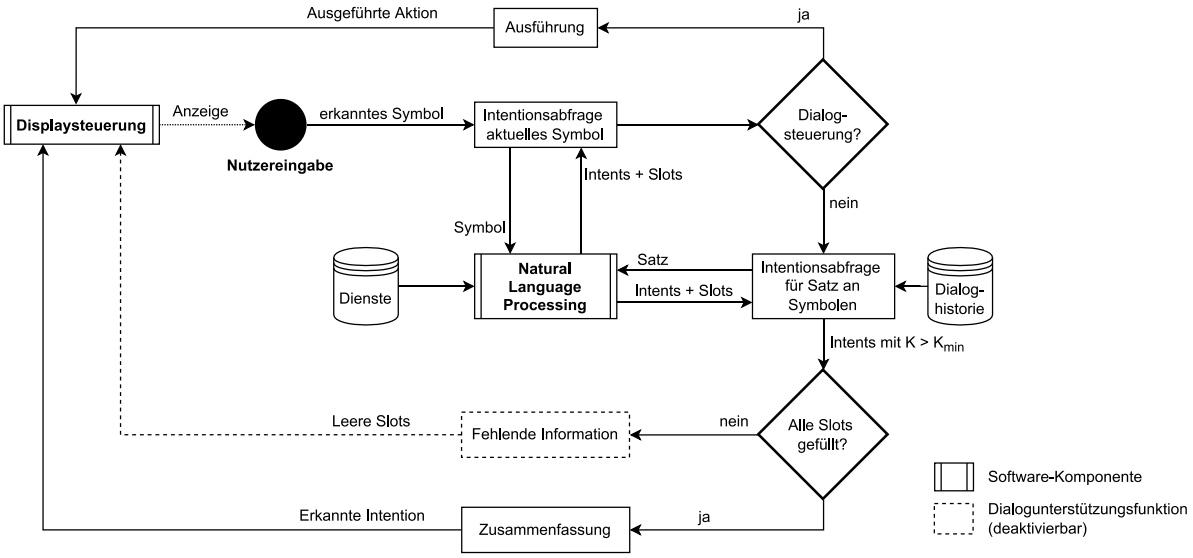


Abb. 75 Im Dialogsystem implementierte Dialogstrategie

Die implementierte Dialogstrategie zeigt Abb. 75. Jede Nutzereingabe wird zunächst auf Symbolebene überprüft, ob es sich dabei um ein Element zur Dialogsteuerung handelt. Ist dies der Fall, wird die entsprechende Aktion direkt ausgeführt. Hierzu gehören beispielsweise solche Dienstanforderungen, mit denen der Nutzer den Dialog beginnen, beenden, zurücksetzen oder eigene Eingaben korrigieren kann. Eine vollständige Übersicht der implementierten Dialogsteuerungselemente befindet sich in Kapitel 5.5.

Handelt es sich nicht um ein Element der Dialogsteuerung, werden mittels der Dialoghistorie mögliche Intentionen mit der Konfidenz $K > K_{min}$ auf Satzebene identifiziert. Die Parametertypen der noch unbesetzten Slots definieren die relevanten Kontexte für die zukünftigen Dialogverlauf und schränken damit den Suchraum für das System ein. Diese leeren Slots können optional an den Nutzer kommuniziert werden und bieten dem Nutzer einen Einblick in die Erwartungen des Systems. Diese Optionsanzeige stellt somit eine der zwei Dialogunterstützungsfunktionen dar und adressiert die Fähigkeitslücke 3.

Wird ein vollständiger Satz an Slots und damit eine Intention hoher Konfidenz erkannt, erstellt der Dialogmanager eine Zusammenfassung und fordert den Nutzer auf, diese zu bestätigen oder abzulehnen. Ist der Nutzer einverstanden, wird die validierte Dienstanforderung einer Liste hinzugefügt, die zu einem späteren Zeitpunkt sequentiell ausgeführt werden kann. Bei einer Ablehnung wird die Intention verworfen und die dazugehörige Dialoghistorie gelöscht.

5.3.2 Natural Language Processing

Die SWK *NLP* identifiziert aus einer Eingangsmenge an Symbolen E eine Menge an möglichen Intentionen I mit der Menge an Parametern (Slots) S und dem Konfidenzwert K. Dazu wurden die vom

Experimentalsystem angebotenen Dienste als Intent/Slot-Paare mit den dazugehörigen Parametertypen (*entities*) abgebildet und eine *Natural Language Understanding Engine*[74] (NLU) mit den generierten Datensätzen trainiert. In Abb. 76 ist die Beschreibung des Intents *LiftObjectDirection* abgebildet, welches das UAV anweist, ein in der Nähe des Nutzers befindliches Objekt aufzunehmen.

```
type: intent
name: LiftObjectDirection
action: LIFT
confirm: true
slots:
  - name: obj
    entity: object
    question: what
  - name: hdir
    entity: hdirection
    question: which one
utterances:
  - lift [obj](box) [hdir](left)
  - lift [hdir](left) [obj](box)
```

Abb. 76 Auszug aus dem Datensatz zur Beschreibung der Intention *LiftObjectDirection* mittels YAML-Notation, Angaben in eckigen Klammern stellen Platzhalter der entsprechenden Entitäten dar, während Angaben in runden Klammern ein konkretes Beispiel repräsentieren

Dazu wird neben der Nennung des aufzunehmenden Objekts auch eine horizontale Richtungsangabe relativ zur Nutzerposition benötigt. Die zu den Entitäten *object* und *hdirection* gehörenden und im Experimentalsystem hinterlegten Symbole sind in Abb. 77 dargestellt.

```
type: entity
name: object
values:
  - box
  - [house, building]
  - tree
---
type: entity
name: hdirection
values:
  - left
  - right
  - front
  - back
  - north
  - south
  - east
  - west
  - cw
  - ccw
```

Abb. 77 Auszug aus dem Datensatz zur Beschreibung der Entitäten *object* und *hdirection* mittels YAML-Notation, Angaben in eckigen Klammern stellen Synonyme dar

5.3.3 Displaysteuerung für den visuellen Rückkanal

Die Aufgabe der SWK *Displaysteuerung* ist die Visualisierung der Systemantworten für den Nutzer über den visuellen Rückkanal. Hierfür wurde eine lichtbasierte Anzeige in Form einer mehrfarbigen 8x32 LED-Matrix gewählt, die auch bei Umgebungsbedingungen mit hoher Beleuchtungsstärke eine ausreichende Lesbarkeit erlaubt und durch die mechanische Kopplung an das Gimbalsystem auch zum Nutzer hin ausgerichtet werden kann. Die Ansteuerung der Matrix erfolgte über den digitalen Port des Microcontrollers vom Typ *Arduino Mega* mittels der *FastLED* Bibliothek [351]. Für die Anbindung

[74] In dieser Arbeit wurde die Implementierung von Coucke et al. [331] genutzt.

an die ROS-Infrastruktur wurde zusätzlich die Bibliothek *rosserial_arduino*[75] auf dem Microcontroller integriert, die über die USB-Schnittstelle eine serielle Kommunikation mit dem ROS-Master auf einem der Missionsrechner ermöglicht. Um den begrenzten Anzeigeraum effektiv nutzen zu können, wurde ein mehrfarbiges Anzeigekonzept (Abb. 78) mit mehreren Signalisierungskomponenten umgesetzt, sodass neben Nutzereingaben und Systemausgaben auch der Zustand der Nutzerverfolgung (gelbe Ecken) sowie das Ergebnis der Blickkontakterkennung (grüne Ecken) signalisiert werden kann. Die Systemausgaben erfolgen in englischer Sprache und werden dem Nutzer in statischer (stehender Text bei weniger als sechs Zeichen) oder dynamischer Darstellung (horizontaler Lauftext) präsentiert.

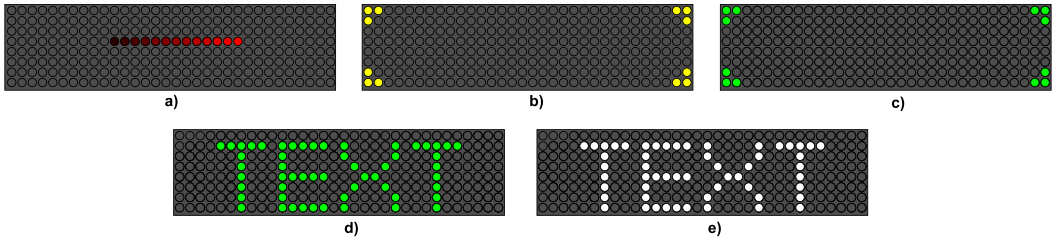


Abb. 78 Implementiertes Anzeigekonzept für den visuellen Rückkanal: **a)** Autorisierten Nutzer wird gesucht, **b)** Nutzerverfolgung aktiv, **c)** Blickkontakt erkannt, d.h. Körpersignale werden analysiert, **d)** Erkannte Nutzereingabe, **e)** Systemausgabe

5.4 Flugsteuerungssystem

Das Flugsteuerungssystem stellt das Bindeglied zwischen dem Dialogsystem und der Flugplattform dar und besteht aus der SWK *Flugmanager* und dem UAV API.

5.4.1 Flugmanager

Die Aufgabe des Flugmanagers ist die sichere Umsetzung der flugbezogenen Dienste und stellt eine Art Sicherheitsnetz für die Flugplattform dar. Dazu abonniert er sowohl die Flugführungswünsche vom Dialogsystem als auch alle relevanten Systemparameter der Flugplattform über die UAV-Schnittstelle. Hierzu zählt beispielsweise der Betriebszustand von Sensoren und Antriebseinheit, der aktive Flugmodus, die gegenwärtige Flughöhe als auch der Ladezustand der Energieversorgung (*State of Charge*, SoC). Stellt der Flugmanager eine systemgefährdende Dienstanfrage fest (z.B. die Abschaltung des Antriebs im Flug), verweigert er die Ausführung und gibt eine entsprechende Begründung an das Dialogsystem zurück, welche über den visuellen Rückkanal an den Nutzer weitergeleitet wird. Darüber hinaus wurden im Flugmanager zwei unterschiedliche Steuerungsverfahren implementiert, die in Kapitel 6.3.1 näher beschrieben werden.

5.4.2 UAV-Schnittstelle

Die UAV-Schnittstelle ist eine vom Hersteller der Flugplattform bereitgestellte Programmierschnittstelle[76], die eine Kommunikation des Experimentalsystems mit dem *Flight Control System* (FCS) der Flugplattform erlaubt. Über diese Schnittstelle werden dem Flugmanager ROS-konforme Telemetrienachrichten über die Flugplattform zur Verfügung gestellt. Gleichzeitig werden hierüber Flugführungsanweisungen als auch Gimbal-Drehraten über den CAN-Bus der Flugplattform an die entsprechenden Steuergeräte weitergeleitet.

[75] http://wiki.ros.org/rosserial_arduino
[76] https://github.com/dji-sdk/Onboard-SDK

5.5 Dienste und Systemkonfigurationen

Das Experimentalsystem verfügt über drei Konfigurationen (*Interview-Mode, Control-Mode, Briefing-Mode*), mit denen unterschiedliche Mensch-UAV-Interaktionsszenarien abgebildet werden können. Jede Konfiguration ist für eins der drei Mensch-Maschine-Experimente (Kapitel 6) optimiert und zeigt ein eigenes Dialogverhalten mit einem eigenständigen Satz an Vokabularen und Diensten. Die implementierten Dienste sind in Tab. 19 dargestellt und gliedern sich in die sieben funktionalen Gruppen:

- Notsignale,
- Dialogsteuerung,
- Rückmeldung,
- Systemeinblick,
- Führung,
- Verlegung und Transport sowie
- Aufklärung und Unterstützung.

Sobald das System eine Nutzerintention mit hinreichend hoher Konfidenz detektiert und ein entsprechender Dienst verfügbar ist, wird dieser sofort ausgeführt. Dienste, die sicherheitskritische oder hochautomatisierte Aktionen einleiten (z.B. Abflug, Landung), benötigen eine zusätzliche positive Rückbestätigung durch den Nutzer vor der Ausführung. Die Verfügbarkeit eines Dienstes kann dabei an andere Systemparameter oder Umgebungsbedingungen gekoppelt sein. So ist beispielsweise aus der funktionalen Gruppe *Führung* nur der Dienst *Starte Antrieb* verfügbar, solange sich das UAV am Boden befindet und die Propeller stillstehen. In diesem Fall würde das System die Ausführung aller anderen Dienste dieser Gruppe ablehnen.

Interview Mode

Dieser Modus ist für die Interaktion mit Ungeschulten in Notsituationen angepasst und verzichtet auf die Nutzerauthentifizierung. Um Interaktionswünsche zu erkennen, reagiert das System auf unspezifische Körpersignale, die eine hohe Gestikulationsstärke (bspw. großformatige Armbewegungen) aufweisen und versucht anschließend mit einem geführten Dialog die Nutzerintention zu bestimmen. Der Umfang an verfügbaren Diensten ist stark reduziert und beschränkt sich auf die Erkennung von Hilfsgesuchen, körperlichen Problemen und den Wunsch, einen Notruf abzusetzen. Relevante Körpersignale wurden möglichst generisch beschrieben und sind im Anhang A5 dargestellt. Vom Experimentalsystem werden in diesem Modus Embleme (z.B. Daumen hoch/runter, Telefon-Geste), Armbewegungen als auch Nick- und Schüttelbewegungen des Kopfes sowie Berührungsgesten erkannt.

Control Mode

Der Control Mode ist optimiert für Interaktion mit geschulten Nutzern während geführter Start-/Lande und Positionierungsvorgängen. Das Vokabular ist im Anhang A6 abgebildet und besteht aus High-Level Flugführungskommandos für die dreidimensionale Positionierung des UAVs in der Luft (Translation und Rotation), aber auch für die Aktivierung von autonomen Funktionen (Antrieb, Start, Landung). Die Körpersignale sind angelehnt an die offiziellen Einweiserzeichen aus der Luftfahrt. Die Interaktion wird durch ein Begrüßungssignal des autorisierten Nutzers (`HELLO`) initiiert.

Briefing Mode

Der Briefing Mode verfügt über die größte Anzahl an verfügbaren Diensten und das umfangreichste Vokabular (siehe Anhang A7) inklusive Synonymen, Metaphern und mehrdeutigen Körpersignalen. Dieser Modus richtet sich an die Nutzergruppe der Experten mit entsprechendem Wissen über das

Vokabular und die Syntax, um Auftragsbeschreibungen nonverbal zu einem UAV zu übermitteln. Die Interaktion wird hier ebenfalls durch das Begrüßungssignal eines autorisierten Nutzers eingeleitet.

Tab. 19 Übersicht der implementierten Dienste und Betriebsmodi im Experimentalsystem, Elemente in Klammern stellen Platzhalter für die entsprechende semantische Gruppe von Symbolen dar

Dienst/Intention	Syntax	Bestätigung notwendig	System-konfiguration*		
			I	C	B
Notsignale					
Benötige Hilfe	HELP	✓	•	-	-
Körperliches Problem	[BODYPART]	✓	•	-	-
Setze Notruf ab	CALL	✓	•	-	•
Dialogsteuerung					
Beginne Dialog	HELLO	×	-	•	•
Ignoriere letzte Eingabe	IGNORE_LAST	×	-	•	•
Wiederhole letzte Ausgabe	SAY_AGAIN	×	•	•	•
Beginne Dialog von vorne	RESET	×	-	•	•
Beende Dialog	BYE	✓	-	•	•
Rückmeldung					
Rückmeldung	[FEEDBACK]	×	•	•	•
Systemeinblick					
Zeige Parameter	YOU ASK [PARAMETER]	×	-	•	•
Führung					
Starte/Stoppe Antrieb	YOU [STATE]	✓	-	•	-
Hebe ab	YOU FLY	✓	-	•	-
Lande an aktueller Position	YOU LAND	✓	-	•	-
Versetze/drehe UAV horizontal	[HDIRECTION]	×	-	•	-
Ändere Flughöhe	[VDIRECTION]	×	-	•	-
Gehe in Schwebeflug über	STOP	×	-	•	-
Verlegung und Transport					
Verlege zu Ort	YOU FLY [SPOT]	✓	-	-	•
Lande bei Ort	YOU LAND [SPOT]	✓	-	-	•
Nehme Nutzlast auf	YOU LIFT [OBJECT] [HDIRECTION]	✓	-	-	•
Transportiere Nutzlast zu Ort	YOU TRANSPORT [OBJECT] [HDIRECTION] [SPOT]	✓	-	-	•
Aufklärung und Unterstützung					
Suche Landeplatz für Vehikel im Umkreis	YOU FIND AREA [VEHICLE] [NUMBER]	✓	-	-	•
Scanne Objekt	YOU SCAN [OBJECT] [HDIRECTION]	✓	-	-	•
Dokumentiere Umgebung	YOU SCAN RADIUS [NUMBER]	✓	-	-	•
Zeige Weg zu Ort	YOU SHOW [SPOT]	✓	-	-	•
Folge Person	YOU FOLLOW [SUBJECT]	✓	-	-	•

* I = Interview Mode, C = Control Mode, B = Briefing Mode

5.6 Realisierung und Integration des Experimentalsystems in eine fliegende Plattform

Das mechanische „Grundgerüst“ des Experimentalsystems stellt ein portables, drei Achsen stabilisierendes Gimbalsystem[77] dar, in welches die Hardwarekomponenten integriert wurden (Abb. 79). Die Rechenleistung wird dabei von zwei baugleichen Missionsrechnern[78] mit dedizierten GPU-Einheiten bereitgestellt. Der Mikrocontroller steuert das LED-Display, welches den visuellen Rückkanal darstellt. Obwohl das Systemkonzept nur einen EO-Sensor vorsieht, wurde im finalen Experimentalsystem ein zusätzlicher, baugleicher Sensor verbaut, um die Mensch-UAV-Interaktion zu Dokumentationszwecken mit einem größeren Sichtfeld aufzeichnen zu können. Darüber hinaus verfügt das Experimentalsystem über eine WLAN-Schnittstelle für die Überwachung der Softwarekomponenten aus der Ferne.

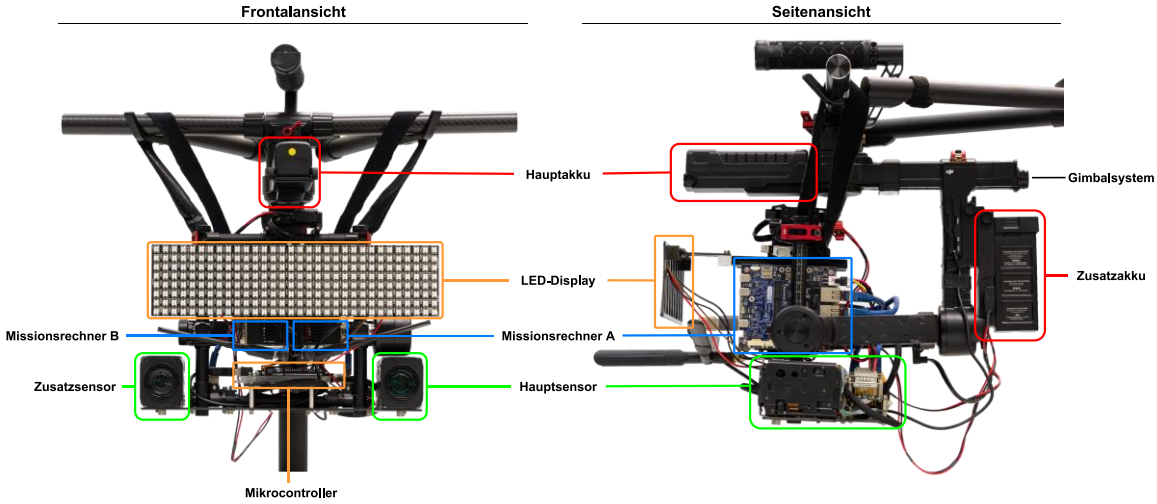


Abb. 79 Das realisierte Experimentalsystem in der Frontal- und Seitenansicht

Für eine möglichst realitätsnahe Mensch-UAV-Interaktion wurde das Experimentalsystem in eine geeignete, schwebefähige Flugplattform[79] größerer Bauart integriert. Diese musste neben den mechanischen und flugleistungsbezogenen Anforderungen[80] auch über eine ROS-kompatible Programmierschnittstelle (*Application Programming Interface*, API) zum FCS verfügen. Darüber hinaus wurde bei der Auswahl der Flugplattform auf eine hohe Zuverlässigkeit geachtet. Hierzu zählen unter anderem eine redundante Auslegung der Energieversorgung (parallel geschaltete Flugakkus) sowie der Sensorik (mehrere GNSS-Empfänger, Barometer und IMUs), um einen sicheren Flugbetrieb bei Realflugexperimenten mit Probanden sicherstellen zu können.

[77] Gimbalsystem vom Typ DJI RONIN-MX.
[78] Embedded System vom Typ Nvidia Jetson Xavier AGX mit Connect Tech Rogue (AGX101) als Carrier Board.
[79] Hexakopter vom Typ DJI Matrice 600 Pro, maximales Abfluggewicht 15,5 kg.
[80] Nutzlastkapazität von 5 kg bei einer Mindestflugzeit von 15 Minuten für Realflugexperimente.

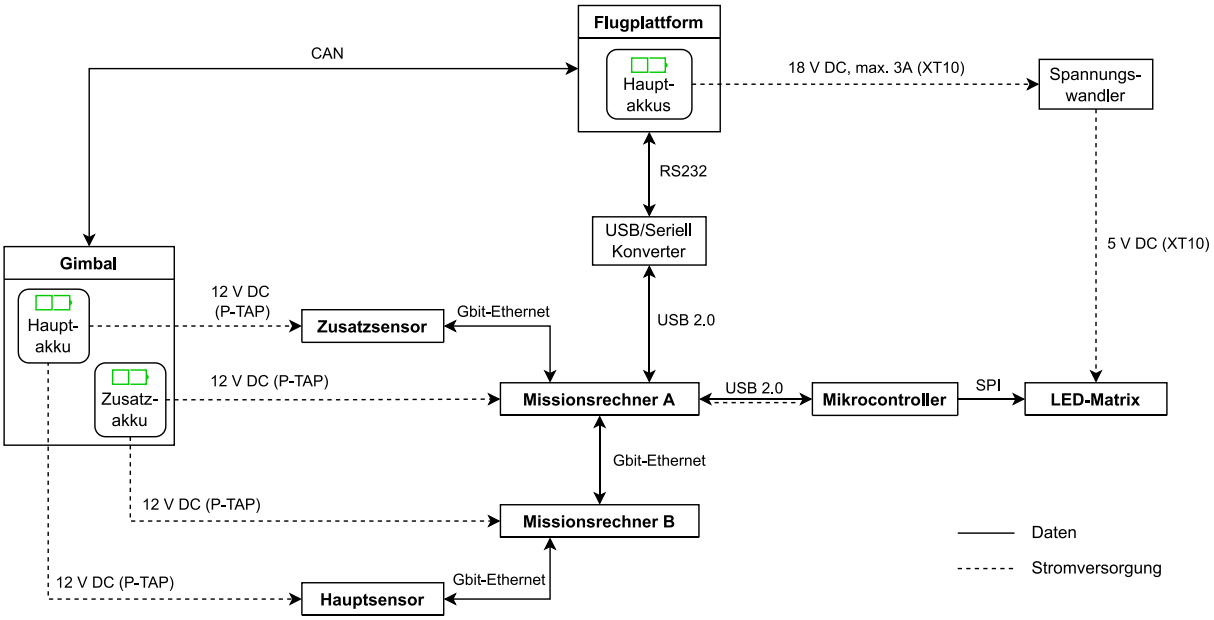


Abb. 80 Elektrische Verbindungsübersicht aller Hardware-Komponenten des Experimentalsystems

Die Energieversorgung der kritischen Komponenten des Experimentalsystems (Missionsrechner und Sensoren) wurde von der Energieversorgung der Flugplattform entkoppelt (Abb. 80). Lediglich die Versorgung der LED-Matrix erfolgte über den Nutzlastversorgungsanschluss der Flugplattform (18 V DC), der mittels eines Spannungswandlers auf 5 V DC reduziert wurde. Das vollintegrierte Experimentalsystem im Flug zeigt Abb. 81.

Abb. 81 Das vollintegrierte Experimentalsystem im Flug

6 Experimentelle Evaluierung

Dieses Kapitel beschreibt zunächst die objektive Bewertung der Klassifikationsleistung des umgesetzten Experimentalsystems in einem Mensch-Maschine-Experiment unter statischen Laborbedingungen. Im Rahmen dieser Arbeit wurden drei weitere Mensch-Maschine-Experimente durchgeführt, wobei jede Untersuchung einen anderen Aspekt der nonverbalen Mensch-UAV-Interaktion beleuchtete, um Antworten auf die identifizierten Wissenslücken zu finden. Im Experiment A wurde eine verhaltensbezogene Laborstudie durchgeführt, um relevante Signalisierungsmuster von ungeschulten Nutzern in einer simulierten Notlage zu identifizieren. Dabei wurde auch die Eignung der implementierten Verfahren zur Erkennung von generischen Nutzereingaben überprüft. Im zweiten Experiment B wurden zwei Flugführungsverfahren mittels subjektiver und objektiver Messmethoden verglichen und das Raumnutzungsverhalten von geschulten Nutzern bei der nonverbalen Einweisung eines hochautomatisierten Multikopters in einem Feldversuch untersucht. Im dritten Experiment C lag der thematische Schwerpunkt auf dem nonverbalen Problemlösungsverhalten von Nutzern mit Expertenwissen bei der Übermittlung von Dienstbeschreibungen in einem dialogähnlichen Vorgang mittels der visuellen Schnittstelle des Experimentalsystems. In einer Laborstudie wurden mit verhaltensbezogenen Analysen diejenigen Faktoren identifiziert, die Verständigungsprobleme während einer nonverbalen Interaktion begünstigen und wie Nutzer darauf reagieren als auch wo die Frustrationsgrenzen bei anhaltenden Verständigungsproblemen liegen. Zudem wurde hierbei auch der Nutzen der zwei implementierten Dialogunterstützungsfunktionen in Form einer Ablationsstudie untersucht.

6.1 Bewertung der Klassifikationsleistung des Experimentalsystems

Für die Charakterisierung des implementierten Experimentalsystems wurde dessen Klassifikationsleistung während der angeleiteten Trainingsphase der Probanden (N=22) im Experiment C (siehe Kapitel 6.4) in einer statischen Laborumgebung evaluiert. Die dabei produzierten und vom Experimentalsystem in Echtzeit erkannten Nutzereingaben wurden per Videoaufzeichnungen dokumentiert und im Nachgang ausgewertet. Als Konfiguration wurde der Briefing Mode gewählt, weil dieser das komplexeste Vokabular aufweist und über die meisten Dienste verfügt. Aufgrund des kontextsensitiven Ansatzes mit mehrfacher Klassenzugehörigkeit desselben Körpersignals, kann die Klassifikationsleistung jedoch nur innerhalb eines semantischen Kontextes bewertet werden. In Tab. 20 sind die betrachteten Kontexte und die ermittelten Klassifikationsleistungen als Übersicht dargestellt.

Tab. 20 Im Probandenversuch ermittelte Klassifikationsleistung des Experimentalsystems (Briefing Mode)

Semantischer Kontext	Anzahl Klassen	Anzahl Eingaben	Anzahl Gesamtfehler	Anzahl Nicht-Detektionen	Precision	Recall	Accuracy	F1-Score
Action	13	788	117	111	0,9639	0,8214	0,8515	0,8726
Number	12	609	67	56	0,9486	0,8403	0,8900	0,8801
Feedback	8	361	17	17	0,9901	0,9269	0,9529	0,9535
Spot	6	367	68	56	0,9064	0,8228	0,8147	0,8334
Object	6	359	71	68	0,9317	0,7367	0,8022	0,8010
Parameter	5	340	39	36	0,9493	0,8571	0,8879	0,8934
Vehicle	5	323	70	70	0,9307	0,7843	0,7833	0,8183
Dialog	4	227	4	4	0,9912	0,9804	0,9824	0,9853
Unit	3	207	23	23	0,9309	0,8734	0,8889	0,8911
Subject	3	200	25	25	0,9263	0,8484	0,8750	0,8725
HDirection	3	133	0	0	1,0000	1,0000	1,0000	1,0000
Gesamt	**68**	**3.914**	**501**	**466**	**0,9517**	**0,8629**	**0,8844**	**0,8910**

Bei den 3.914 Eingaben mit insgesamt 68 Klassen lag die Fehlerquote des Experimentalsystems bei etwa 12,8 %. Davon entfielen ca. 93 % der Fehler auf Nicht-Detektionen, d.h. das System hat keine Reaktion auf eine Nutzereingabe gezeigt. In nur 7 % der Fälle traten Klassenverwechslungen auf. Die Konfusionsmatrix in Abb. 82 gibt einen detaillierteren Einblick in die Klassenverhältnisse im semantischen Kontext *Action*. Die Konfusionsmatrizen aller anderen Kontexte befinden sich im Anhang A8.

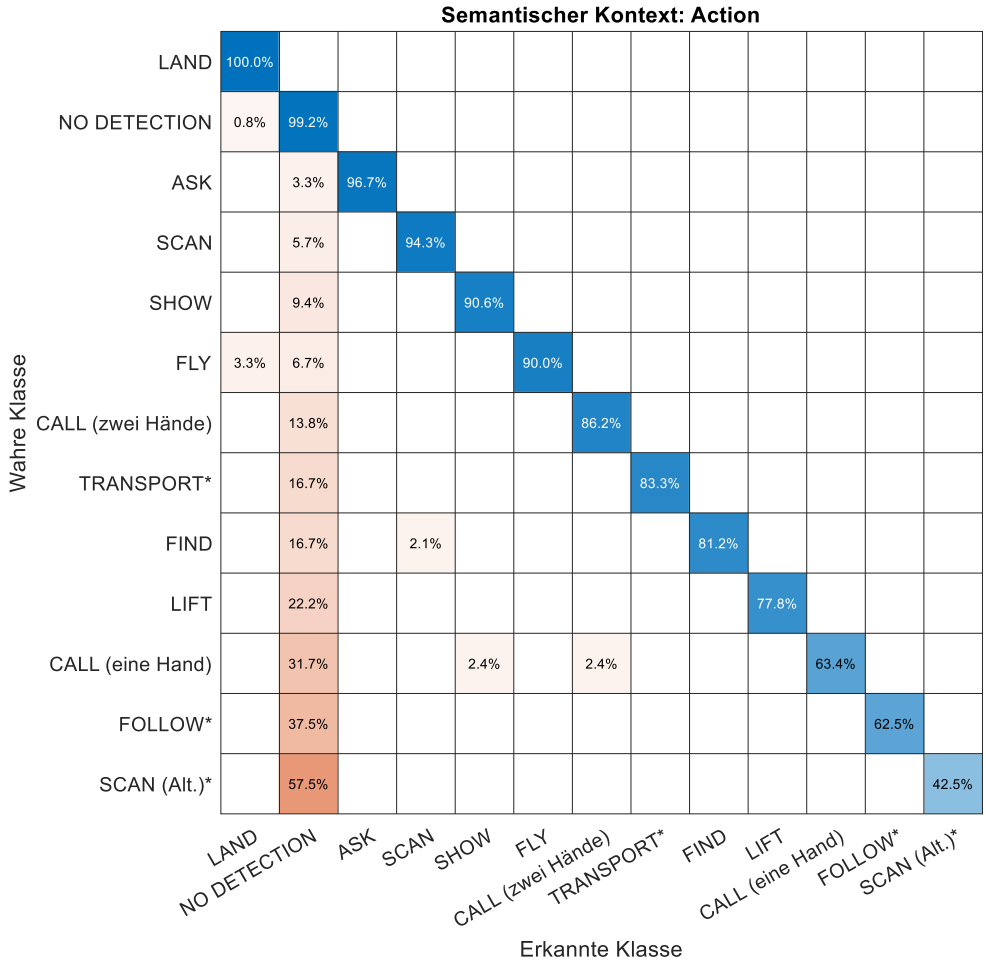


Abb. 82 Konfusionsmatrix mit der Klassifikationsleistung des Experimentalsystems im semantischen Kontext *Action,* Signale mit dynamischen Komponenten sind mit einem Stern markiert, Alternativen sind mit Alt. abgekürzt

Zu den drei am wenigsten erkannten Signalen dieses Kontextes gehören zwei Klassen mit einer dynamischen Komponente (gekennzeichnet durch einen Stern) - *FOLLOW** und *SCAN (Alt.)**. Die Erkennung dieser Signale basiert primär auf dem Klassifikationsergebnis des DTW-Detektors, der für die Analyse der Bewegungspfade der Hände zuständig ist. Eine mögliche Ursache für die niedrige Detektionsleistung dieser Signale kann daher im verwendeten, womöglich zu kleinen Datensatz vermutet werden, mit dem der Detektor angelernt wurde (siehe Kapitel 5.2.8). Gleichfalls stellt die Erhebungsmethode einen möglichen weiteren Faktor dar. Die Messungen wurden während des Vokabeltrainings der Probanden erhoben, sodass ein Großteil die entsprechenden nonverbalen Signale überhaupt zum ersten Mal produzierte. Es kann daher nicht gänzlich ausgeschlossen werden, dass die niedrige Detektionsleistung dieser Signale auch durch eine ungenaue Darstellung der Probanden beeinflusst wurde. Das dritte, am wenigsten erkannte, Signal stellt *CALL (eine Hand)* dar, welches mittels der Handform (ausgestreckter Daumen und kleiner Finger), dem aktiven Gestenraum (links vom Körper) und dem Bewegungszustand der Hand (keine Bewegung) beschrieben wird. Einige der Probanden hielten ihre Hand zu nah am Kopf und aktivierten dadurch den Gestenraum „vor dem Kopf", wodurch die Erkennung des Signals verhindert wurde. Wiederum anderen bereitete die korrekte Produktion der notwendigen Handform mit der linken Hand Schwierigkeiten. Die Ursache hierfür könnte somit unter anderem in einer zu strickten Formulierung des Signals liegen. Für den Großteil der problematischen Signale enthält das Vokabular jedoch Synonyme in Form von alternativen Darstellungsformen, beispielsweise als Pose (siehe *SCAN*) oder durch den alternativen Gebrauch beider Hände (siehe *CALL*).

Bei der für die Steuerung des Dialogs wichtigen Gruppe an Signalen aus dem semantischen Kontext *Dialog* lag die Fehlerquote bei ca. 1,7 %. Ein Blick in die entsprechende Konfusionsmatrix (Abb. 83) verdeutlich auch hier, dass es sich dabei ausschließlich um Nicht-Detektionen handelt.

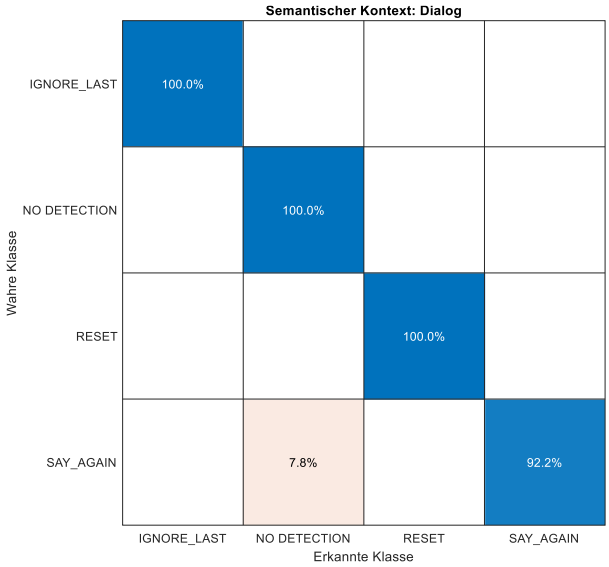


Abb. 83 Konfusionsmatrix mit der Klassifikationsleistung des Experimentalsystems im semantischen Kontext *Dialog*

Zusammenfassend kann das Detektionsverhalten des Experimentalsystems als eher konservativ charakterisiert werden, d.h. es übersieht zwar vereinzelt Körpersignale (Recall von 86,29 %), zeigt aber eine hohe Genauigkeit bei den erkannten Signalen (Precision von 95,17 %). Dieses Verhalten ist bei einem nonverbalen Mensch-UAV-Kommunikationssystem wünschenswert, weil es dem Nutzer einen gewissen „Bewegungsspielraum" lässt, ohne dass eine Systemreaktion durch eventuell unbewusst übermittelte Körpersignale ausgelöst wird. Falls das System ein bewusst platziertes Signal nicht er-

kennt, reicht es meistens aus, dieses zu wiederholen. Hätte das Klassifikationssystem jedoch die Tendenz, Klassen zu verwechseln, bestünde ein höheres Risiko für einen negativen Dialogverlauf und im Falle der nonverbalen Einweisung eines fliegenden Systems, auch ein höheres Risiko für die Unversehrtheit des Nutzers und der Flugplattform.

6.2 Experiment A: Nonverbale Körpersignale von Nutzern ohne Systemerfahrung

2020 verzeichnete der Deutsche Alpenverein 920 Unfälle und Notfälle im deutschen Alpenraum [352]. Neben Stürzen führten insbesondere Blockierungen[81] beim Bergsteigen, Klettern oder Wandern zur Alarmierung der Luftrettung. Schlechte Wetterbedingungen in Form von hohen Windgeschwindigkeiten, Gewitter, gefrierendem Regen oder starken Sichteinschränkungen durch Nebel und niedrige Wolkenuntergrenzen können jedoch auch eine Gefahr für die Besatzung von Rettungstransporthubschraubern darstellen [353, S. 130], sodass in diesen Situationen keine Rettung aus der Luft erfolgt und die Hilfesuchenden stattdessen von Bodenkräften gefunden und versorgt werden müssen. Unbemannte Fluggeräte sind zum gegenwärtigen Zeitpunkt noch nicht imstande, diese Lücke zu schließen und Menschen automatisiert aus einem Gefahrengebiet zu fliegen, aber sie können zumindest dabei helfen, die Position von Vermissten und Hilfesuchenden zu bestimmen und die Koordinaten an eine Rettungsleitstelle zu übermitteln. Um die Dringlichkeit einer Notlage festzustellen, reichen Standortinformationen alleine jedoch nicht aus und erfordern eine genauere Analyse der Situation. Sofern der Hilfesuchende noch ansprechbar ist, kann eine Befragung ein hilfreiches Werkzeug zur Feststellung der Kritikalität darstellen. Hierzu muss ein UAV jedoch in der Lage sein, einen (nonverbalen) rudimentären Dialog mit einer Person führen zu können, was voraussetzt, dass das UAV menschliche Körpersignale nicht nur erfassen, sondern diese auch korrekt interpretieren können muss. Um diese Fähigkeitslücke adressieren zu können, muss daher zunächst bekannt sein, auf welche nonverbalen Signalisierungsmuster Menschen in solchen Notlagen zurückgreifen.

6.2.1 Untersuchungsinhalte

Diese Untersuchung wurde durchgeführt, um Antworten auf die Forschungsfragen aus dem Bereich der Wissenslücke 1 (WL1) zu finden und die Hypothese 2 (H2) zu überprüfen. Im Detail wurden folgende Fragestellungen untersucht:

WL1: Nonverbale Kommunikationssignale von Nutzern ohne Systemerfahrung im Kontext einer simulierten Notsituation

Es existieren zwar offizielle alpine Boden-Luft-Notsignale (vgl. Kapitel 2.2.6), jedoch kann nicht vorausgesetzt werden, dass diese allen Menschen bekannt sind. Auch muss davon ausgegangen werden, dass Menschen (insbesondere solche ohne jegliche Erfahrung mit Zeichensprachen) in Notsituationen auf bekannte Verhaltens- und Darstellungsmuster (z.B. großräumiges Winken mit den Armen) zurückgreifen, die sich an Alltagserfahrungen orientieren. Insofern stellt sich die Frage, welche visuell erfassbaren Körpersignale ungeschulte Nutzer in (simulierten) Notlagen verwenden, um insbesondere:

- einen Interaktionswunsch zu signalisieren (WL1.1),
- zustimmende oder ablehnende Rückmeldungen zu geben (WL1.2) sowie

[81] Damit werden diejenigen Notfälle bezeichnet, bei denen Bergsportler zwar unverletzt sind, sich jedoch verlaufen haben oder zu erschöpft sind, um den Rückweg aus eigener Kraft anzutreten.

- auf körperliche Probleme hinzuweisen (WL1.3).

H2 Eignung der implementierten Verfahren zur Erkennung von generischen Eingaben

Das VisCom-Experimentalsystem kann eine Reihe von visuellen Merkmalen erfassen und klassifizieren. Die Merkmale „Gestikulationsstärke“ und „Mundbild“ sind zwei davon, welche in Kapitel 5.2.5 und 5.2.6 näher beschrieben wurden und zur Erkennung von unspezifischen Interaktionswünschen genutzt werden können. Ein weiterer Untersuchungspunkt ist daher die Frage, wie gut sich diese Merkmalbeschreiber zur Erkennung von Interaktionswünschen eignen. Als Bewertungsmetrik wurden daher der Anteil erkannter Interaktionswünsche (H2.1) sowie die Zeitdauer für deren Erkennung (H2.2) herangezogen.

Um generische zustimmende oder ablehnende Nutzerrückmeldungen detektieren zu können, unterstützt das Experimentalsystem auch die Erkennung von Nick- und Schüttelbewegungen des Kopfes (Kapitel 5.2.4) sowie bestimmte Handformen (z.B. Daumen hoch/Daumen runter in Kapitel 5.2.10). Als Bewertungsmetrik wurde auch hier der Anteil der erkannten Rückmeldungen genutzt (H2.3).

Im Körpermodell des Experimentalsystems sind zudem mehrere Gliedmaßen hinterlegt (siehe Kapitel 5.2.9), auf die der Nutzer Bezug nehmen kann, sofern das entsprechende Körperteil mit einer Hand berührt wird. Die dritte Fragestellung lautet daher, wie gut sich das implementierte Annäherungsverfahren zur Erkennung von körperlichen Problemen eignet (H2.4). Darüber hinaus soll der Interaktionserfolg, d.h. ob ein Notruf erfolgreich vom UAV abgesetzt wurde (H2.5), als auch die Interaktionsdauer (H2.6) ermittelt werden.

6.2.2 Untersuchungsdesign

In einer kontrollierten Laborumgebung wurde eine Notfallsituation nachgestellt und folgendes Narrativ verwendet: Bei einer Wanderung in den Alpen ist ein körperliches Problem aufgetreten, welches ein Weiterlaufen unmöglich macht und daher externe Hilfe benötigt wird. Die Probanden sind jedoch auf sich alleine gestellt und von der Außenwelt abgeschnitten. Am Himmel erkennen sie ein (möglicherweise) bemanntes Flugobjekt, welches auf sie zuzufliegen scheint und in der momentanen Lage die einzige Rettungsmöglichkeit darstellt.

Die Untersuchung gliederte sich dabei in zwei aufeinander aufbauende Abschnitte. Im ersten Teil sollten die Probanden innerhalb eines definierten Zeitfensters von 20 Sekunden so viel Aufmerksamkeit wie möglich auf sich ziehen, um den Vorbeiflug des Flugobjektes zu verhindern und eine Interaktion aufzubauen. Dabei waren ihnen alle Modalitäten und Darstellungsformen freigestellt. Nachdem der Interaktionswunsch erfolgreich erkannt wurde, schwebte das Flugobjekt im zweiten Teil der Untersuchung imaginär vor den Probanden und initiierte eine textbasierte Konversation mittels eines Displays, um festzustellen, ob bei den Probanden ein medizinisches Problem vorlag. Die Probanden wurden dabei angewiesen auf ein selbstgewähltes, körperliches Problem hinzuweisen und die Konversation dahingehend zu lenken, dass das UAV einen Notruf absetzt. Hierzu mussten Probanden auch Systemrückfragen im Laufe der Konversation beantworten.

Versuchsaufbau

Die Experimente wurden in einem Versuchslabor der Universität der Bundeswehr München in Neubiberg durchgeführt. Die Probanden standen ca. drei Meter vom Experimentalsystem (ohne fliegende

Komponente) entfernt und blickten auf einen 25“-Monitor[82], der auf Augenhöhe befestigt war (Abb. 84). Die Versuchsleitung befand sich dabei in einiger Entfernung links vom Probanden.

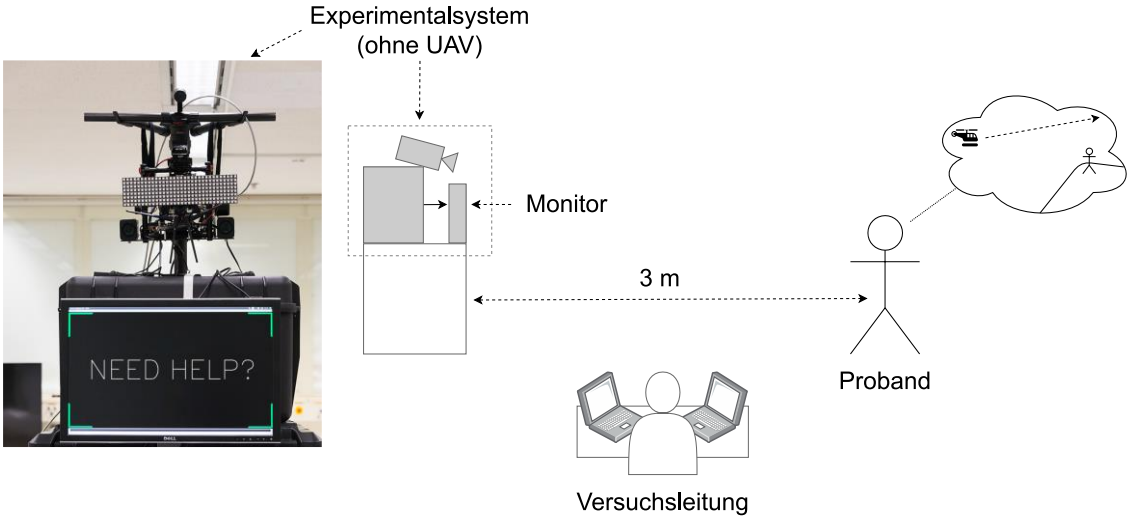


Abb. 84 Laboraufbau im Experiment A

Systemverhalten

Für diese Untersuchung wurde der *Interview Modus* des VisCom-Experimentalsystems genutzt. Dabei wird nach der Erkennung einer hohen Gestikulationsstärke eine einfache Befragungsroutine zur Feststellung der Hilfsbedürftigkeit aktiviert (Abb. 85). Positive Rückmeldungen (Kopfnicken, Daumen-Hoch-Geste) führen die Konversation in Richtung Notruf fort, während negative Rückmeldung in Form von Kopfschütteln, ablehnenden Armbewegungen oder einer Daumen-runter-Geste die Routine vorzeitig beenden und so die Konversation zurücksetzen. Zur Erkennung von problembehafteten Körperteilen wurde der Klassifizierer für Berührungen aus Kapitel 5.2.9 in Kombination mit der Bewegungsanalyse der Hände aus Kapitel 5.2.4 genutzt. Vor dem Beginn der Untersuchungen wurde ein Probelauf mit einer Freiwilligen durchgeführt und das Untersuchungsdesign getestet.

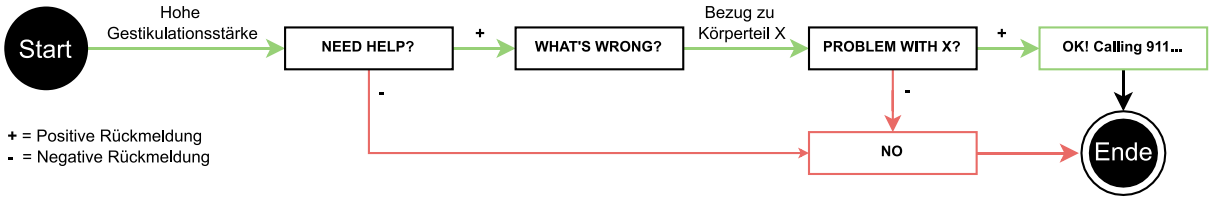


Abb. 85 Systemverhalten im Experiment A, Rechtecke stellen dabei die Systemantworten dar, die den Probanden auf dem Monitor dargestellt wurden (der rote Pfad führt zu einem Neustart des Dialogs, der grüne zur Auslösung des Notrufs)

6.2.3 Stichprobenbeschreibung

An diesem Experiment nahmen 22 Probanden teil (zwei davon weiblich). Alle Teilnehmer verfügten über einen direkten Luftfahrtbezug durch Studium der Luft- und Raumfahrttechnik bzw. durch die berufliche Tätigkeit am Institut für Flugsysteme. Drei der Probanden verfügten zudem über eine mi-

[82] Um einen möglichen negativen Einfluss der begrenzten Anzeigefläche der LED-Matrix auf die Interaktion auszuschließen, wurde bei diesem Experiment ein alternatives Ausgabegerät mit einer größeren Anzeigefläche gewählt.

litärische Ausbildung. Keiner der Teilnehmer hatte Erfahrungen mit Handzeichensystemen. Die vertretene Altersspanne reichte von 21 bis 55 Jahren, mit einem durchschnittlichen Alter von 30 Jahren (SD=7,44).

6.2.4 Datenerhebung

Die Daten wurden im Zeitraum vom 27.04.2021 bis 25.05.2021 erhoben. Zu Beginn des Experiments wurden demographische Daten der Teilnehmer per Fragebogen erfasst und die anschließende Interaktion per Audio- und Videoaufzeichnung dokumentiert. Wichtige Systemparameter (erkannte Eingaben, Klassifikationsergebnisse sowie Systemausgaben) des Experimentalsystems wurden ebenfalls aufgezeichnet. Eine Übersicht der ermittelten Daten befindet sich in Anhang A10.

6.2.5 Versuchsdurchführung

Nach einem Briefing[83] wurde die Probanden gebeten, sich in die beschriebene alpine Notlage hineinzuversetzen und sich möglichst authentisch zu verhalten sowie mit dem Experimentalsystem zu interagieren. Mittels eines standardisierten Ablaufplans wurden die Probanden durch beide Interaktionsphasen geleitet. Im Anschluss erfolgte ein offenes Debriefing.

6.2.6 Operationalisierung und Messmethoden

Zur Charakterisierung der visuellen Notsignale wurde die bevorzugte Darstellungsform[84] als auch die Subsignale Arm-/Handbeteiligung, Mundbeteiligung sowie die Handform anhand der aufgezeichneten Videoaufnahmen manuell analysiert und zu Gruppen ähnlicher Ausprägungen zusammengefasst (WL1.1, WL1.2, WL1.3). Zur Feststellung der Eignung der visuellen Merkmale „Gestikulationsstärke“ und „Mundbild“ zur Erkennung von Interaktionsabsichten wurde zum einen der Anteil der erkannten Interaktionswünsche bestimmt (H2.1) und zum anderen die Zeitspanne ab dem Beginn der Probandenbewegung bis zur ersten Reaktion des Experimentalsystems manuell anhand des Timecodes der Aufnahmen ermittelt (H2.2). Zusätzlich wurde der Anteil der erkannten Rückmeldungen überprüft (H2.3).

Der Interaktionserfolg wurde am Anteil der korrekt als problematisch klassifizierten Körperteile (H2.4) sowie dem Anteil der erfolgreich abgesetzten Notrufe gemessen (H2.5). Zudem wurde die Interaktionsdauer von der ersten Probandenbewegung bis zur Anzeige der Erfolgsmeldung erfasst (H2.6). Im Falle einer erfolglosen Interaktion wurde die Zeit bis zum Interaktionsabbruch durch den Probanden gemessen. Tab. 21 zeigt die in diesem Experiment verwendeten Messgrößen in der Übersicht.

[83] Das Briefing umfasste die Vorstellung der Untersuchungsziele, die Art der Datenerhebung und -aufbewahrung als auch die Aufklärung über die Rechte des Probanden sowie die Unterzeichnung der Einverständniserklärung zur Verarbeitung und Veröffentlichung der erhobenen und pseudonymisierten Daten.
[84] Diejenige Darstellungsform, welche als erstes produziert wurde.

Tab. 21 Übersicht der Messgrößen zur Operationalisierung im Experiment A

Untersuchungsinhalt und Messgrößen	Ausprägung				
WL Nonverbale Kommunikationssignale von Nutzern ohne Systemerfahrung					
	Darstellungsform	Arm-/Handbeteiligung	Mundbeteiligung	Kopfbeteiligung	Handform
WL1.1 Interaktionswunsch	•	•	•	-	-
WL1.2 Rückmeldung	•	•	•	•	-
WL1.3 Problembeschreibung	•	•	-	-	•
H2 Eignung der visuellen Merkmale zur Erkennung generischer Eingaben					
H2.1 Anteil erkannter Interaktionswünsche	0 … 100 %				
H2.2 Erkennungsdauer der Interaktionswünsche	0 … x Sekunden				
H2.3 Anteil erkannter Rückmeldungen	0 … 100 %				
H2.4 Anteil erkannter Problembeschreibungen	0 … 100 %				
Interaktionserfolg					
H2.5 Anteil erfolgreich abgesetzter Notrufe	0 … 100 %				
H2.6 Interaktionsdauer	0 … x Sekunden				

6.2.7 Ergebnisse

Die nachfolgenden Abschnitte beschreiben die ermittelten Ergebnisse. Das Signifikanzniveau für die statistische Auswertung wurde auf 0,05 festgelegt.

Visuelle Signalisierung eines Interaktionswunsches (WL1.1)

Zur Erzeugung von Aufmerksamkeit wählten alle Probanden eine dynamische Darstellung. Dabei bevorzugte der Großteil ein großräumiges Winken (N=16). Zwei Probanden sprangen zusätzlich zum Winken, um sich „größer zu machen“ (P12). Jeweils einmal wurde Winken mit einer heranholenden Bewegung sowie mit einer Zeigegeste auf den eigenen Oberkörper kombiniert. Ebenso wurde einmal eine Kombination aus kreisenden und heranholenden Bewegungen als auch nur heranholende Bewegungen (Abb. 86 links) gewählt. In fast dreiviertel der Fälle (N=16) wurde die Bewegung unter Verwendung beider Arme durchgeführt. Vier der Probanden begannen mit einer einarmigen Bewegung und wechselten dann auf eine zweiarmige Darstellung. Zwei Probanden bevorzugten durchgehend nur einen Arm für die Signalisierung (Abb. 86 rechts oben). Etwas mehr als die Hälfte der Probanden (N=12) nutzte den akustischen Kanal aktiv (N=10), um das UAV heranzurufen oder deutete Mundbewegungen zumindest an (N=2). Die Minderheit (N=10) verwendete nur den visuellen Kanal und verzichtete auf eine Beteiligung des Mundes zur Signalisierung eines Interaktionswunsches (Abb. 86 rechts unten).

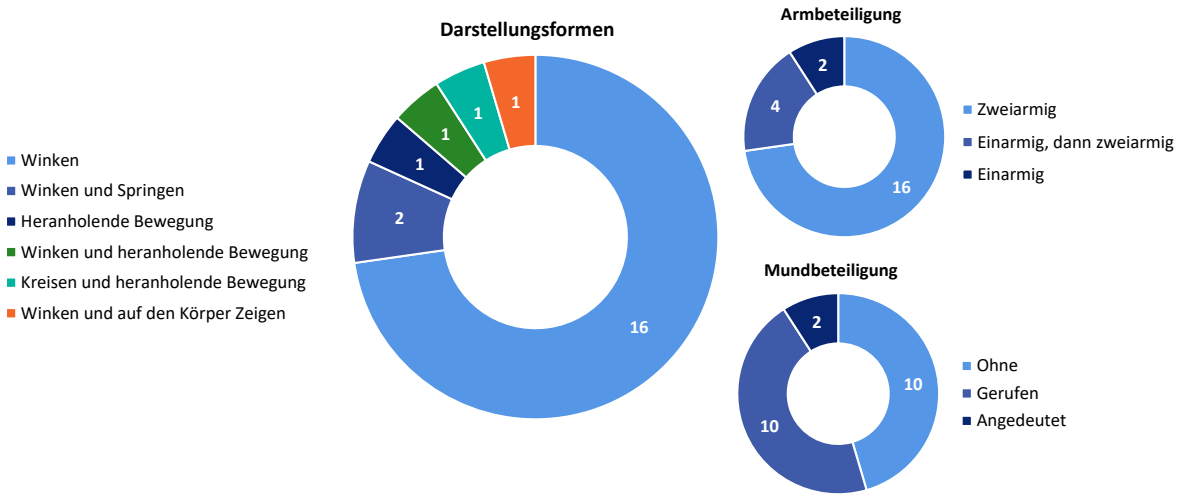


Abb. 86 Verteilung der Darstellungsformen (links) sowie der Beteilung der Arme (rechts oben) und des Mundes (rechts unten) zur Signalisierung eines Interaktionswunsches

Visuelle Rückmeldungssignale (WL1.2)

Für die visuelle Rückmeldung wurden vielfältige Darstellungsformen (Abb. 87) gewählt. Die Daumen-Hoch-Geste war dabei häufig vertreten (N=6) und wurde von einem Teilnehmer mit einem Kopfnicken kombiniert. Drei Probanden versuchten mit einer heranholenden Bewegung ihre Zustimmung zu zeigen, während zwei Teilnehmer dafür eine seitwärts gerichtete Winkbewegung verwendeten. Ein Proband (P17) verzichtete auf eine nonverbale Darstellung komplett und gab stattdessen eine rein verbale Antwort. Lediglich ein Proband (P11) nutzte das offizielle Notsignal (Y-Pose) zur positiven Rückmeldung. Ein großer Teil der Probanden (N=6) wählte jedoch dieselbe Darstellung wie zuvor für die Signalisierung des Interaktionswunsches. Aus dieser Gruppe fügte die Mehrheit (N=5) ihrer Darstellung ein Kopfnicken hinzu oder kommunizierte zusätzlich über den akustischen Kanal (N=4).

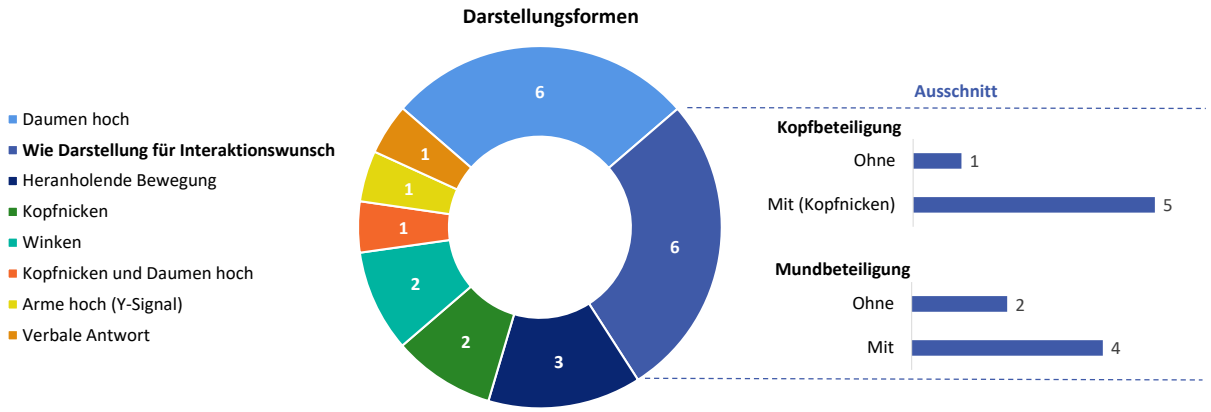


Abb. 87 Die verwendeten Darstellungsformen zur visuellen Rückmeldung

Im Allgemeinen wurde in mehr als der Hälfte der Fälle beide Hände (N=12) für die Gestikulation genutzt, sechs Teilnehmer kommunizierten mit einer Hand und vier kamen ganz ohne Handbeteiligung aus. Während der Rückmeldung nickte die Mehrheit (N=15) mit dem Kopf und 50 Prozent der Probanden (N=11) sprachen dabei mit dem System (Abb. 88).

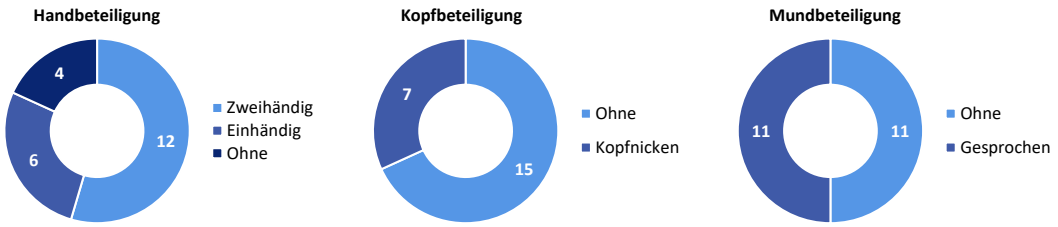


Abb. 88 Die Verteilung der Subsignale während der Rückmeldung

Visuelle Signale zur Beschreibung eines körperlichen Problems (WL1.3)

Zur Beschreibung eines körperlichen Problems nutzten die Probanden eine Vielzahl von Darstellungen (Abb. 89). Am häufigsten (N=5) wurde auf das problematische Körperteil gezeigt, gefolgt von Klopfen (N=3) oder Klopfen in Kombination mit Umgreifen des betreffenden Körperteils (N=3). Jeweils zwei Probanden wählten Mischformen aus Zeigen und Klopfen sowie Klopfen und Streichen. Aber auch singuläre Streichbewegungen (N=2) oder Berührungen (N=2) wurden zur Signalisierung verwendet. Jeweils ein Proband nutzte die Kombination aus Zeigen und Umgreifen sowie Zeigen und Streichen. In den meisten Fällen wurde das problematische Körperteil einhändig (N=10) oder zweihändig beschrieben. In einigen wenigen Fällen (N=2) wurde von einhändig zu zweihändig im Laufe der Beschreibung gewechselt. Die Mehrheit der Teilnehmer nutzte für klopfende, streichende oder umgreifende Signale eine flache Handform (N=11), für deiktische Darstellung wurde dagegen der Zeigefinger (N=10) bevorzugt.

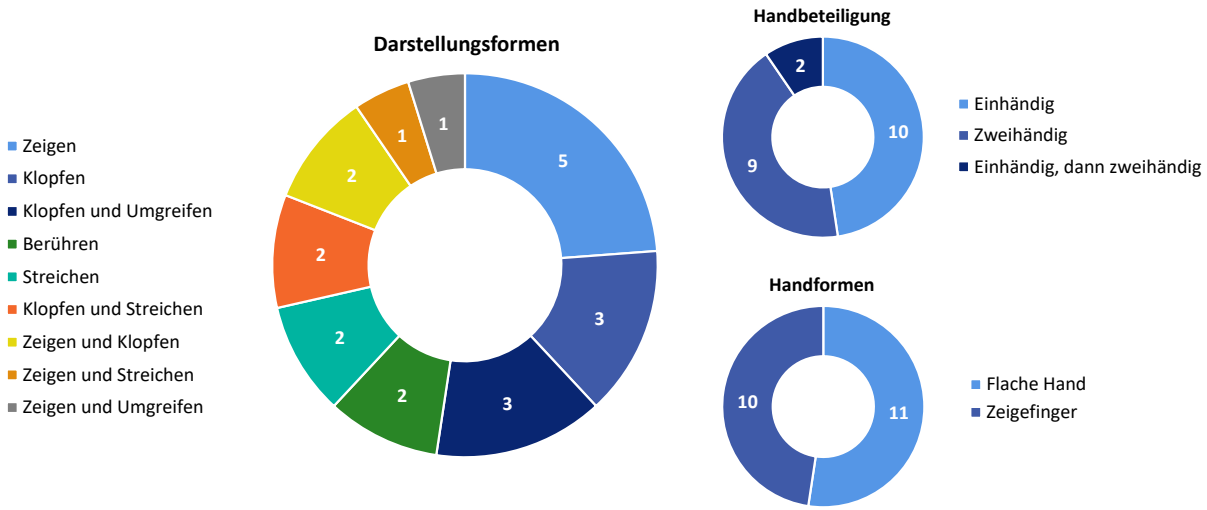


Abb. 89 Verteilung der Darstellungsformen zur Beschreibung eines körperlichen Problems (links), der Handbeteiligung (rechts oben) und der genutzten Handformen (rechts unten)[85]

Anteil erkannter Interaktionswünsche und Erkennungsdauer (H2.1 und H2.2)

Mittels der Gestikulationsstärke und des Mundbilds konnten bei 19 Probanden (86 %) der Interaktionswunsch innerhalb eines Zeitfensters von 20 Sekunden erkannt werden. Nach spätestens 38 Sekunden wurden alle Interaktionswünsche erfasst. Im Durchschnitt dauerte es 6,77 s (SD = 9,89 s), bis das Experimentalsystem eine Reaktion zeigte. Die drei Probanden mit den längsten Detektionszeiten (Abb. 90) nutzten zunächst nur einen Arm zur Signalisierung und gingen erst danach auf eine zweiarmige

[85] Diese Auswertung enthält nur die Daten von 21 der 22 Probanden. Das Experimentalsystem hat bei einem Teilnehmer (P22) nicht auf dessen Rückmeldesignale reagiert, sodass dieser Teil des Experiments vom Probanden vorzeitig beendet wurde und daher keine Daten diesbezüglich vorliegen.

Darstellungsform über. Im Gegensatz zu P03 und P09, fügte P13 zum Winken eine heranholende Bewegung hinzu, um auf sich aufmerksam zu machen.

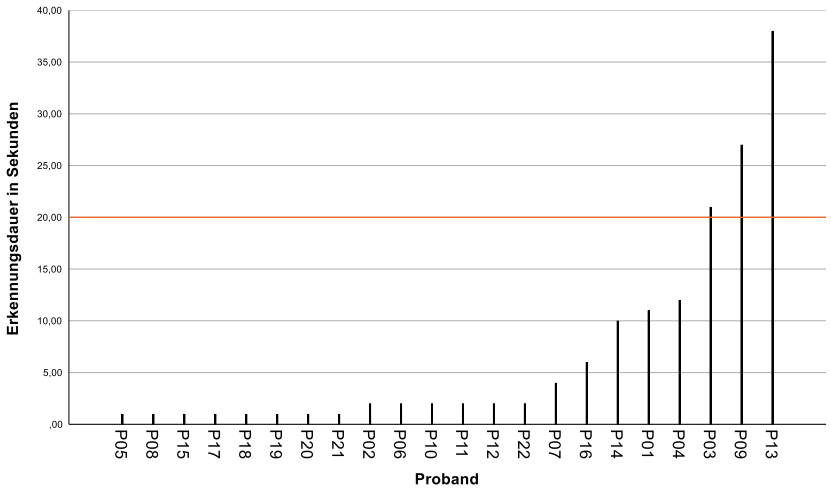


Abb. 90 Erkennungsdauer von Notsignalen mittels der Gestikulationsstärke mit Markierung bei 20 s

Bei einer genaueren Betrachtung zeigt sich, dass Darstellungen mit Beteiligung beider Arme und Mundbewegungen (N=16) im Durchschnitt nach 2,50 s (SD = 3,07 s) erkannt wurden, wohingegen zunächst einarmig, dann zweiarmig durchgeführte Darstellungsformen mit Mundbeteiligung (N=3) nach durchschnittlich 28,67 s (SD = 8,62 s) vom System wahrgenommen wurden. Tab. 22 zeigt die Mittelwerte aller Varianten in der Übersicht.

Tab. 22 Erkennungsdauer in Sekunden für nonverbale Notsignale mittels der Gestikulationsstärke, Standardabweichung in Klammern

		Armbeteiligung					
		Einarmig	N	Zweiarmig	N	Einarmig, dann zweiarmig	N
Mundbeteiligung	Ohne	2,00 (*)	1	3,13 (3,60)	8	4,00 (*)	1
	Mit	12,00 (*)	1	2,50 (3,07)	8	28,67 (8,62)	3

*: nur ein Messwert verfügbar

Anteil erkannter zustimmender oder ablehnender Rückmeldungen (H2.3)

Innerhalb der Untersuchung mussten alle Probanden zweimal auf Systemrückfragen mit einer zustimmenden oder ablehnenden Rückmeldung antworten. Von den insgesamt 44 Eingaben wurden 41 erkannt (93 %).

Anteil Erkannter Problembeschreibungen (H2.4)

Von den 21 Darstellungen eines problematischen Körperteils konnte das Experimentalsystem 18 (86 %) korrekt klassifizieren. Die Klassifizierung galt als korrekt, wenn das erkannte Körperteil zur selben Extremität (Arm, Bein) bzw. zum selben Teil des Körperstamms (Kopf, Hals, Rumpf) des problematischen Körperteils gehörte. Lag beispielsweise ein Problem mit einem Fuß vor, wurde aber vom System als „Problem mit Bein“ erkannt, so wurde die Klassifizierung als korrekt gewertet.

Anteil erfolgreich abgesetzter Notrufe (H2.5)

Von den 22 Interaktionen führten 20 Interaktionen (90 %) zum Absetzen des Notrufs und damit zum Erfolg der Interaktion. Die Ursachen der zwei abgebrochenen Interaktionen (P19, P22) lagen in nicht erkannten Rückmeldungen, die eine Fortführung des Dialogs verhinderten.

Interaktionsdauer (H2.6)

Die Interaktionsdauer der erfolgreichen Interaktionen (N=20) lag im Bereich von 15 bis 168 Sekunden und dauerte im Durchschnitt 66,2 Sekunden (SD = 44 s). Zwei Probanden brachen die Interaktion nach jeweils 105 Sekunden (P19) bzw. 63 Sekunden (P22) ab, nachdem das Experimentalsystem auf mehrfache zustimmende Rückmeldeversuche nicht reagierte.

Fehlerbetrachtung

Der Auswerteprozess wurde so weit wie möglich automatisiert, um Fehler bei der Datenverarbeitung zu minimieren. Es kann jedoch nicht ausgeschlossen werden, dass bei der manuellen Annotation der Beobachtungen Fehler eingebracht wurden. Es wurden aber Gegenmaßnahmen getroffen (Plausibilitätsüberprüfungen, Ausreißeranalyse), um mögliche Fehler zu identifizieren.

6.2.8 Diskussion

Während der Untersuchung haben sich zwei Gruppen von Probanden herauskristallisiert. Während ein Teil der Probanden (N=12) sich in die vorgestellte Notsituation gut hineinversetzen konnte und ein authentisches Verhalten (laute Rufe, großräumige und hektische Körperbewegungen) zeigte, wies der andere Teil (N=10) in der Laborsituation ein eher zurückhaltendes Verhalten auf mit langen Pausen zwischen den Produktionen und wenig bis keinem Stimmeinsatz.

Interaktionswunsch und Gestikulationsstärke

Erwartungsgemäß nutzte ein Großteil der Probanden großräumige, mit beiden Armen durchgeführte Winkbewegungen zur Signalisierung eines Interaktionswunsches. Aber auch andere Signalisierungsformen mit hohen Bewegungsamplituden konnten mittels des visuellen Merkmals „Gestikulationsstärke“ erfasst werden. Einarmige, langsam ausgeführte Darstellungen enthielten jedoch zu wenig Dynamik und wurden vom Experimentalsystem erst erkannt, nachdem der zweite Arm hinzugenommen wurde. Insofern sollte das Verhalten der Arme/Hände separat betrachtet und die Sensitivität erhöht werden, um weniger deutliche Körpersignale von tatsächlich verletzten und in der Bewegungsfreiheit eingeschränkten Nutzern detektieren zu können. In diesem Experiment standen alle Probanden während der Interaktion. In einem realistischen Szenario muss das Klassifizierungssystem auch in der Lage sein, Körpersignale von Sitzenden oder Liegenden zu erkennen.

Rückmeldungen

Trotz der relativ geringen Stichprobe, konnte eine große Bandbreite an nonverbalen Rückmeldungssignalen beobachtet werden. Nicht immer waren die Signale so offensichtlich und eindeutig wie die Daumen-Hoch-Geste[86]. Mehrfach wurden dieselben Hand- und Armbewegungen für unterschiedliche Signale genutzt und unterschieden sich nur in den Subsignalen (Kopfbewegungen, verbale Rückmeldungen), was die Notwendigkeit für einen multimodalen Klassifizierungsansatz verdeutlicht, um relevante Subsignale (Kopfnicken) von irrelevanten (Winkbewegungen mit den Armen) im situativen Kontext unterscheiden zu können. Im Laufe der Interaktion stellten einige Probanden (P04) die nonverbale Kommunikation zeitweise sogar komplett ein und gingen zu einer rein verbalen Form der Kommunikation über. Insofern war die Entscheidung, eine orale Komponente in das Klassifizierungssystem miteinzubeziehen richtig. Die implementierte, reine binäre Klassifizierung des Mundbildes (geöffnet/geschlossen) stößt hierbei jedoch an ihre Grenzen, sodass alternative Methoden benötigt wer-

[86] Selbst dieses Signal ist nicht so eindeutig wie es zunächst scheint. Je nach Intention des Signalproduzenten kann es als „Ja, ich brauche Hilfe“ oder „Nein, bei mir ist alles in Ordnung“ gedeutet werden.

den, die eine genauere Analyse des verbalen Kanals ermöglichen. Aufgrund der Lautstärkeproblematik (siehe Kapitel 3.1.2) kommen akustische Schnittstellen bei der aktuellen Generation von UAVs nicht in Frage. Jedoch verzeichnen DNN-basierte Methoden zum Lippenlesen in den letzten Jahren große Fortschritte [354]. Insofern könnten diese Verfahren in zukünftigen Klassifizierungssystemen zum Einsatz kommen.

Körperausrichtung

Die Designentscheidung, Nutzersignale erst dann zu verarbeiten, wenn sowohl Oberkörper als auch Kopf in Richtung des Systems zeigen, stellte sich bei dieser Nutzergruppe in einigen Fällen als problematisch heraus. So drehte ein Proband (P03) seinen Körper seitlich zum Experimentalsystem und mimte eine unnatürliche Laufbewegung nach, um zu signalisieren, dass etwas mit dem Bein nicht in Ordnung sei. Durch das Wegdrehen vom System wurden die Bewegungen vom Klassifizierungssystem jedoch nicht analysiert und somit auch keine Reaktion angezeigt. Einige Probanden (P02, P20, P21) hielten zudem nicht immer ihren Blickkontakt zum System aufrecht und richteten diesen zeitweise auf ihr (simuliertes) verletztes Bein. Insofern stellt die Blickrichtung des Nutzers auch abseits des Systems ein wertvolles visuelles Merkmal dar, welches einer permanenten Analyse bedarf, um ein besseres Verständnis der Nutzersituation und damit der Kritikalität zu erreichen.

Dialogführung

Die textbasierte Ausgabe des Experimentalsystems war sehr reduziert, vom Informationsgehalt jedoch ausreichend, da alle Probanden die Systemausgaben verstehen und einen minimalistischen Dialog führen konnten. Die lineare Dialogführung führte jedoch teilweise zu Missverständnissen, wenn Probanden bewusst oder unbewusst Systemfragen verneinten (z.B. Kopfschütteln) und somit erst die Aufmerksamkeit des Systems wieder auf sich ziehen mussten, bevor sie beim letzten Dialogpunkt anknüpfen konnten. So erkannte das Experimentalsystem in einigen Fällen andere als von den Probanden ausgewählte Körperteile als problematisch, sodass manche Teilnehmer die nachfolgende Systemrückfrage ablehnten, was zu einem Dialogabbruch führte und sie damit gezwungen waren, den Dialog von vorne zu beginnen. Eine flexiblere Dialogführung hätte den Dialogverlauf in dieser Gruppe womöglich beschleunigen können.

Schlussfolgerungen

Die Stichprobe in dieser Untersuchung stellt nur einen sehr kleinen und spezifischen Ausschnitt aus einem westlich geprägten Nutzerkreis dar. Die gewonnen Erkenntnisse können daher nicht ohne Weiteres auf andere Bevölkerungsgruppen oder Kulturen übertragen werden, sodass weitere Forschung zu nonverbalen Signalisierungsmustern notwendig erscheint. Das Experimentalsystem des VisCom-Konzepts war jedoch bei der untersuchten Probandengruppe in der Lage, einen Großteil der nonverbalen Nutzersignale korrekt zu interpretieren sowie eine (simulierte) Alarmierung der Rettungskräfte in 90 % der Interaktionen erfolgreich einzuleiten. Weiteres Verbesserungspotential liegt insbesondere in der visuellen Analyse des auditiven Kanals (Mundbild) sowie in einer intelligenten Dialogführung über zusätzliche, multimodale Rückkanäle.

6.3 Experiment B: kollokierte UAV-Führung im Realflug

Die nonverbale Interaktion zwischen Flugbesatzung und Bodenpersonal während des Rollvorgangs auf dem Vorfeld ist ein standardisierter Vorgang, bei dem ein Marshaller das Luftfahrzeug mittels visueller Längs- und Querführungsanweisungen zur optimalen Parkposition leitet. Bei der Einweisung von schwebefähigen Fluggeräten in Bodennähe (z.B. Helikopter) muss der Marshaller darüber hinaus auch vertikale Führungsanweisungen übermitteln. Damit unbemannte Transportsysteme zukünftig ebenfalls in diese Abläufe eingegliedert werden können, müssen sie in der Lage sein, ebensolche menschliche Einweisungssignale interpretieren und in Positions- und Fluglageänderungen umsetzen zu können. Unter dem Begriff der *kollokierten UAV-Führung* ein entsprechender Anwendungsfall als möglichst realitätsnahes Experiment nachgebildet.

6.3.1 Untersuchungsinhalte

Diese Untersuchung wurde durchgeführt, um Antworten auf die Forschungsfragen aus dem Bereich der Wissenslücke 2 (WL2) zu finden. Im Detail wurden folgende Fragestellungen untersucht.

WL2.1 Flugführungsverfahren zur Fernführung eines Multikoptersystems über eine visuelle Schnittstelle

In der zwischenmenschlichen Interaktion wird beim Rangieren von Fahr- und Flugzeugen mit der Unterstützung durch Hilfspersonal oftmals ein zweiteiliges Signalisierungsverfahren angewendet [117, 118, S. 42-54]: mit einem Arm gibt der Einweiser in Form einer statischen Zeigegeste die Bewegungs- oder Steuerrichtung vor, während er mit dem anderen eine periodische Bewegung vollführt. Über die Frequenz dieser Bewegung wird dem Steuerer mitgeteilt, mit welcher Intensität die Bewegungs- oder Richtungsänderung zu erfolgen hat. Eine spezielle Unterbrechungsgeste oder der Stillstand der Gestikulation signalisiert üblicherweise dem Steuerer das Ende der Bewegung. Idealerweise besteht während der gesamten Interaktion ein Sichtkontakt zwischen den Beteiligten, sodass beide Interaktionspartner während des Prozesses mehr oder minder stark aneinandergekoppelt sind. Die Führungsanweisung bezieht sich dabei immer nur auf eine Steuerungsdimension [355, S. 248-249]. Diese Form der Kommunikation wird im Allgemeinen nur innerhalb eines kurzen Zeitfensters genutzt, d.h. für die Dauer der Hilfestellung (Einparken, Rückwärtsfahrt, Umsetzen, etc.). Wie steht es jedoch um die Beanspruchung des Einweisenden, wenn dieser Vorgang länger dauert und der Interaktionspartner kein Mensch, sondern ein UAV ist? Gibt es eventuell ein besseres Führungsverfahren, um ein unbemanntes Fluggerät einzuweisen?

Im Experimentalsystem wurden daher zwei unterschiedliche Führungsvarianten implementiert, welche in idealisierter Form in Abb. 91 dargestellt sind. Die erste Variante orientiert sich am oben beschriebenen Verfahren, sodass das UAV eine Flugbewegung nur dann ausführt, solange ein Flugführungskommando produziert wird. Dieses Verfahren wird im Folgenden als *kontinuierliches Führungsverfahren* (KF-Verfahren) bezeichnet, weil der Nutzer für die Führung des UAVs kontinuierlich Eingaben produzieren muss. Stoppt die Produktion, beendet auch das UAV seine Bewegung zeitnah. Hierbei ist eine hohe Mitwirkung des Nutzers notwendig, da während der kontinuierlichen Gestikulation auch ein Blickkontakt zum UAV gehalten werden muss, um gültige Kommandos zu übermitteln.

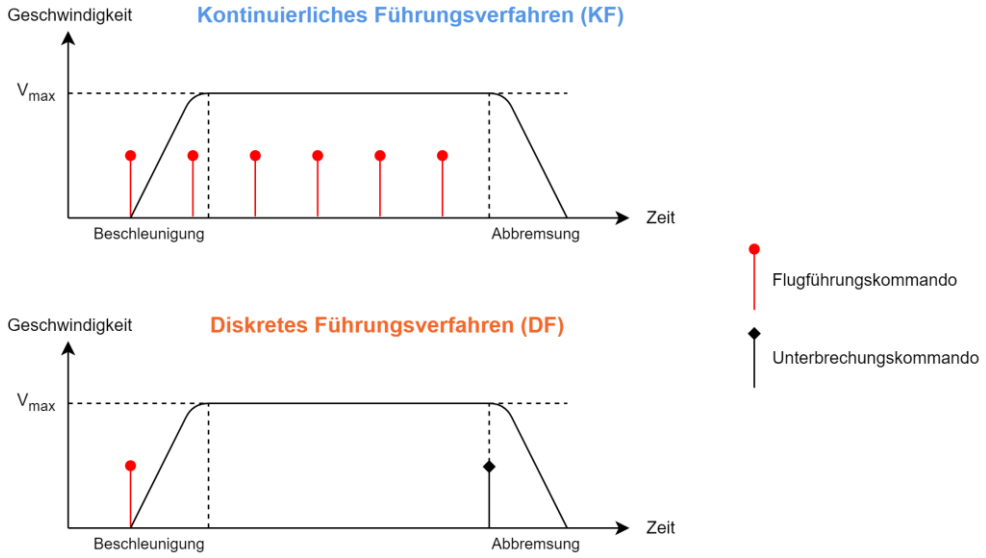


Abb. 91 Kontinuierliches und diskretes Führungsverfahren im Zeitverlauf

Bei der zweiten Variante wird eine Flugbewegung ebenfalls durch ein Flugführungskommando eingeleitet, jedoch mit dem Unterschied, dass dieses solange ausgeführt wird, bis ein anderslautendes oder ein spezielles Unterbrechungskommando empfangen wird. Der Nutzer muss bei diesem Verfahren also nur zu diskreten Zeitpunkten Flugführungskommandos übermitteln (d.h. zu Beginn und am Ende einer Flugbewegung) und auch nur währenddessen den Blickkontakt zum UAV aufrechterhalten. Dieses Verfahren wird im Folgenden als *diskretes Führungsverfahren*[87] (DF-Verfahren) bezeichnet. Beide Verfahren wurden im Experimentalsystem implementiert und im Hinblick auf die subjektive Beanspruchung als auch auf die Effektivität der UAV-Führung während der Interaktion untersucht.

WL2.2 Raumnutzungsverhalten während kollokierter Interaktion

Um ein Fluggerät mit einer Fernsteuerung zu führen, muss der Nutzer seinen Standort nicht verändern, solange er dessen Position und Fluglage sicher einschätzen kann. Anders ist es hingegen bei der kollokierten Interaktion, bei der sich der Nutzer in der Regel in der Nähe des UAVs aufhalten muss, um die Sichtverbindung aufrecht zu erhalten. Daher gibt es prinzipiell drei mögliche Spielarten, um das UAV zu versetzen, ohne die Sichtverbindung zum Fluggerät für einen längeren Zeitraum aufgeben zu müssen (Abb. 92). Dabei kann die Hauptbewegungsrichtung des UAVs zum Nutzer hin (*Pull-Methode*), vom Nutzer weg (*Push-Methode*) oder seitlich zum Nutzer (*Parallel-Methode*) erfolgen.

[87] Um zu verhindern, dass das UAV bei länger anhaltendem Verlust der Sichtverbindung unkontrolliert weiterfliegt, wurde ein automatischer Übergang in den Schwebeflug implementiert, der ausgelöst wird, wenn nach Ablauf einer bestimmten Zeitspanne (in diesem Fall ca. 12 Sekunden) kein Unterbrechungskommando empfangen wird.

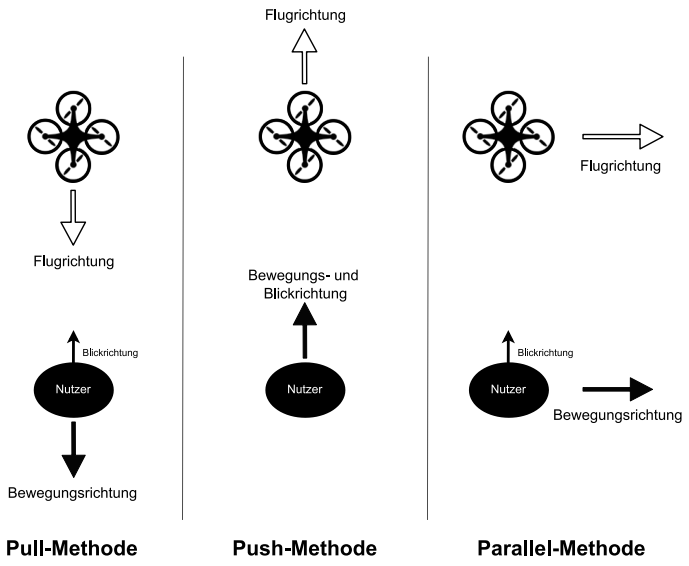


Abb. 92 Räumliche Flugführungspräferenzen unter Beibehaltung einer Sichtverbindung

Bei der Pull-Methode muss der Nutzer die Bewegungen des UAVs im Blick behalten, gleichzeitig aber auch den rückwärtigen Raum beobachten, um bei Rückwärtsbewegungen nicht zu stolpern. Bei Kommunikationsproblemen kann bei dieser Methode eine sichere Distanz zum Fluggerät unterschritten werden, wenn beispielsweise ein Unterbrechungskommando während einer Vorwärtsbewegung auf den Nutzer zu vom UAV nicht erkannt wird.

Die Push-Methode bietet eine komfortable Möglichkeit, um das UAV zu führen und gleichzeitig den Bewegungspfad zu beobachten ohne den Blick vom UAV abwenden zu müssen. Die Sicherstellung einer sicheren Entfernung zum UAV steuert der Nutzer. Der Luftraum hinter dem UAV ist jedoch für den Nutzer nur bedingt einsehbar.

Bei der Parallel-Methode bleibt aufgrund der lateralen Bewegungsrichtung der Abstand zum UAV gleich. Der Nutzer muss jedoch ebenfalls den Blick vom UAV zeitweise abwenden, um den seitlichen Raum beobachten zu können. In drei unterschiedlichen Flugphasen (Abflug, Transit, Landung) sollten die bevorzugten Flugführungsstile identifiziert werden.

WL2.3 Einfluss der Erfahrung auf die Interaktionsdistanz

Multikopter können aufgrund der schnell rotierenden Propeller und der potentiellen und kinetischen Energie eine Gefahr für den Nutzer darstellen. Während die Erforschung der Proxemik in der Mensch-Roboter-Interaktion bereits weit vorangeschritten ist, ist das Raumnutzungsverhalten in der Mensch-UAV-Interaktion noch wenig erforscht. In der Literatur finden sich einige Untersuchungen zur Proxemik zumeist nur im Zusammenhang mit kleineren Quadkoptersystemen bis 2 kg und vorrangig unerfahrenen Nutzern [207]. In diesem Experiment sollte daher näher untersucht werden, welche Abstände bei der Interaktion mit einem Hexakoptersystem der 15 kg Klasse bevorzugt werden und ob das Erfahrungsniveau sich auch im Interaktionsabstand bemerkbar macht. Die Kenntnis über eine (möglicherweise) bevorzugte Interaktionsdistanz kann Hinweise darauf geben, welche Anforderungen an das Nutzererfassungssystem gestellt und welche Sicherheitsvorkehrungen für die jeweilige Nutzergruppe getroffen werden müssen. Im verwendeten Experimentalsystem wurde ein elektrooptischer Sensor mit variabler Brennweite[88] genutzt, der Interaktionsdistanzen von einem bis etwa 50 m ermöglichte.

[88] Variabler Öffnungswinkel von 55,4° bis 2,9°, entspricht in etwa einer 20-fachen optischen Vergrößerung.

6.3.2 Untersuchungsdesign

Die Aufgabe der Probanden bestand darin, das Experimentalsystem von einem 5x5 Meter großen Teil einer Wiese (Zone A) zu einem 50 m entfernt gelegenen gleichgroßen Bereich (Zone B) nonverbal zu führen, ohne dabei ein dazwischen befindliches Hindernis zu überfliegen (siehe Abb. 93). Die entsprechenden Bereiche wurden am Boden mit Absperrband markiert. Neben einer möglichst zentralen Landung in der entsprechenden Zone, mussten auch die Vorgaben zur Orientierung des Fluggerätes nach der Landung (Nord- bzw. Südausrichtung) sowie weitere Flugparameter (Tab. 23) beachtet werden. Start und Landung wurden durch mehrere High-Level Befehle kommandiert und erst nach einer zusätzlichen Rückbestätigung durch den Nutzer vollautomatisiert ausgeführt. Im Gegensatz dazu wurden während des Fluges alle anderen Flugführungskommandos direkt vom UAV in eine Fluglageänderung umgesetzt. Das dabei verwendete Gestenvokabular orientiert sich an den Einweiserzeichen aus der Luftfahrt [118, S. 42-54]. Diejenigen Handzeichen, die Bewegungen in der frontalen Ebene enthielten, wurden auf eine zweidimensionale Darstellungsform umgeändert. Eine Übersicht mit den implementierten nonverbalen Flugführungskommandos und der verwendeten Syntax befindet sich in Anhang A6.

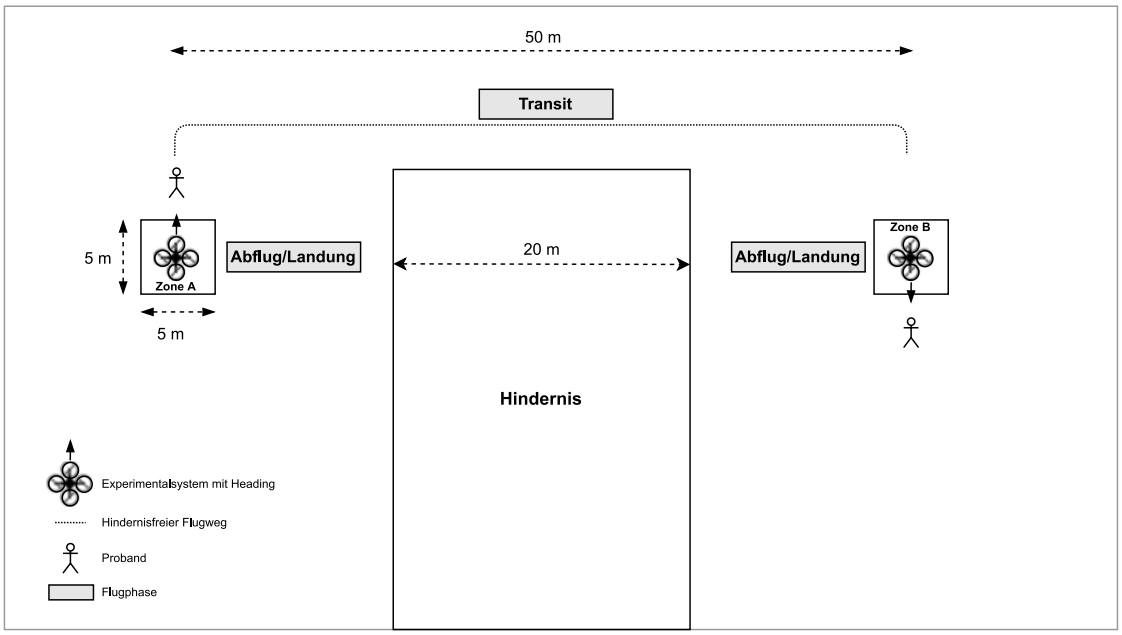


Abb. 93 Versuchsaufbau im Experiment B

Erhöhung des Realitätsgrades

Menschen haben im Allgemeinen eine gute Vorstellung darüber, was ihr Gegenüber sehen kann oder ob ein Gegenstand eventuell die Sicht auf einen bestimmten Bereich versperrt. In der Psychologie spricht man hierbei von der Fähigkeit zur Perspektivenübernahme. Bei robotischen Systemen ist von außen nicht immer erkennbar, wo der Erfassungsbereich der verbauten Sensorik beginnt und wo er endet. Auch kann es „tote Winkel“ geben, die von keinem der Sensoren abgedeckt werden. Um den Realitätsgrad der Interaktion weiter zu erhöhen, wurde daher der Arbeitsbereich des verwendeten dreiachsigen Gimbalsystems horizontal auf etwa +/- 30 Grad und vertikal auf ca. +/- 15 Grad softwareseitig limitiert, um einen eingeschränkten Erfassungsbereich zu simulieren. Als Orientierungshilfe dienten den Versuchsteilnehmern die beiden vorderen, roten Rotorarme des Fluggerätes, die in etwa den Begrenzungen des nutzbaren Sensorbereichs entsprachen. Farblich kodierte Ecken an der LED-Matrix signalisierten den Probanden zudem, ob sie sich innerhalb oder außerhalb des Erfassungsbereichs des optischen Systems befanden.

Neben sensorischen Unzulänglichkeiten wurden auch zeitkritische Faktoren in das Experimentaldesign eingebaut. Für beide Flüge standen den Probanden 80 % der verfügbaren Kapazität zur Verfügung, was einer Flugzeit von ca. 15 Minuten entspricht. Das Experimentalsystem wurde so konfiguriert, dass bei einem Absinken der verbleibenden Kapazität auf unter 20 % eine automatische Sicherheitslandung eingeleitet wurde, um eine Tiefenentladung der Flugakkus zu vermeiden. In Tab. 23 sind die Zielvorgaben für die Probanden mit den entsprechenden Auswirkungen bei Missachtung dargestellt.

Tab. 23 Vorgaben für Probanden in Experiment B

Parameter		Vorgabe	Auswirkung bei Missachtung
Entfernung zum UAV		1 – 50m	Keine
Flughöhe		1,5 – 5m	Unterschreiten: Kollision mit Boden Überschreiten: Verlust der Sichtverbindung
Position des Nutzers relativ zum UAV		Zwischen Rotorarm 1 und 2	Verlust der Sichtverbindung
Landung in Zone B	Ausrichtung	Süd	Nichterreichung der Aufgabe
(Flug 1)	Akkukapazität	> 60%	Keine
Landung in Zone A	Ausrichtung	Nord	Nichterreichung der Aufgabe
(Flug 2)	Akkukapazität	> 20%	Einleitung der automatischen Landung

Vortest

Vor dem Beginn der Erprobungskampagne wurde ein Probelauf mit zwei Freiwilligen durchgeführt und einzelne Systemparameter (Schwellwerte für Klassifizierungssystem, Helligkeit der LED-Matrix) angepasst. Im Vergleich zum DF-Verfahren wurde für das KF-Verfahren eine höhere maximale Horizontalgeschwindigkeit gewählt (2,00 m/s statt 0,75 m/s), um die Anzahl der notwendigen Kommandos pro Flugstrecke auf ein vertretbares Maß zu reduzieren und um sicherzustellen, dass die verfügbare Kapazität der Flugakkus für zwei komplette Flüge ausreicht.

6.3.3 Stichprobenbeschreibung

An diesem Experiment nahmen 15 Probanden teil (davon einmal weiblich). Alle Teilnehmer verfügten über einen direkten Luftfahrtbezug durch Studium der Luft- und Raumfahrttechnik bzw. durch die berufliche Tätigkeit am Institut für Flugsysteme als wissenschaftliche Mitarbeitende oder technische Angestellte. Zwei der Probanden verfügten zudem über eine militärische Ausbildung. Die vertretene Altersspanne reichte von 25 bis 58 Jahren, mit einem durchschnittlichen Alter von 32,7 Jahren (SD = 8,7). Die praktische Erfahrung mit der Fernsteuerung von Fluggeräten wurde mit null bis 47 Jahren angegeben und lag im Mittel bei 9,1 Jahren (SD = 13,1). Alle Teilnehmer waren Rechtshänder und verfügten über eine (korrigierte) Sehleistung von 100 %.

6.3.4 Datenerhebung

Die Daten wurden im Zeitraum vom 18.06.2021 bis 22.07.2021 auf dem Testgelände der Universität der Bundeswehr München in Neubiberg (Abb. 94) erhoben. Zu Beginn des Experiments wurden demographische Daten der Teilnehmer per Fragebogen erfasst. Die Interaktion wurde aus Sicht des Probanden (Point-of-View-Kamera auf Brusthöhe) als auch aus der des Fluggerätes per Videoaufzeichnung dokumentiert. Wichtige Systemparameter (Telemetrie, Sensordaten sowie diverse Klassifikationsergebnisse) wurden ebenfalls an Bord des Flugsystems aufgezeichnet. Zusätzlich wurde nach jeder Landung

die Distanz des Probanden zum UAV manuell gemessen. Eine Übersicht aller ermittelten Daten befindet sich in Anhang A11.

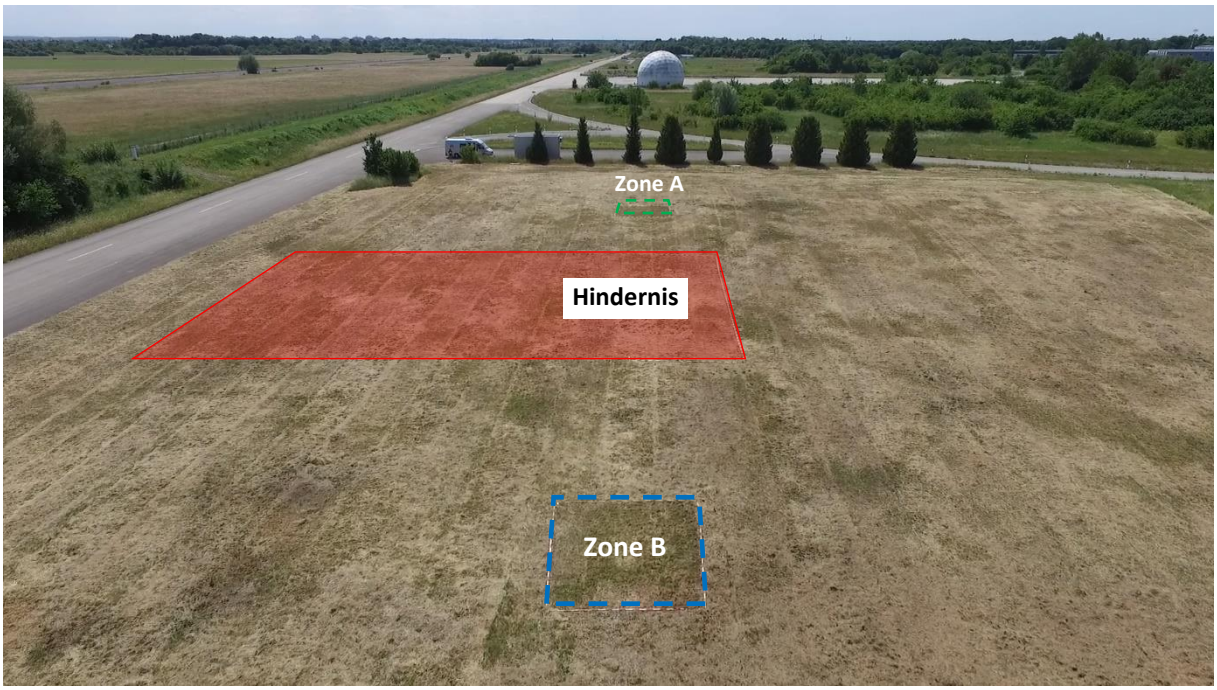


Abb. 94 Das für Experiment B verwendete Testgelände mit hervorgehobenen Abflug-/Landezonen A und B sowie dem fiktiven Hindernis dazwischen

6.3.5 Versuchsdurchführung

Nach dem Briefing wurde jeder Proband in die Funktionsweise des Experimentalsystems eingewiesen und mit den Betriebsgrenzen des Kommunikationssystems (Erfassungsbereiche, minimale und maximale Flughöhen) als auch dem ersten Führungsverfahren[89] vertraut gemacht. Im Anschluss wurden die verfügbaren Flugführungskommandos vorgestellt, gefolgt von einem ausgiebigen, standardisierten Training. Kurz vor dem ersten Flug wurden die entsprechenden Körpersignale in der Zone A nochmals abgefragt und deren Erkennung im Zusammenspiel mit dem Experimentalsystem auf eine korrekte Ausführung hin überprüft.

Daraufhin kommandierten die Probanden selbständig den Start, dirigierten das UAV entlang der Transitstrecke am Hindernis vorbei, stellten die vorgeschriebene Ausrichtung des UAVs her und leiteten die Landung in Zone B ein. Direkt im Anschluss an die Landung erhielt jeder Proband einen NASA RTLX Fragebogen zum Ausfüllen. Vor dem Rückflug zu Zone A wurde das zweite Steuerungsverfahren erläutert und aktiviert. Alle Flüge wurden von einem Sicherheitspiloten überwacht und die verbleibende Akkukapazität in 10 %-Schritten von der Versuchsleitung angesagt. Abb. 95 zeigt die Interaktion kurz vor der Landung in Zone A aus der Sicht eines Probanden.

[89] Die Zuweisung der Führungsverfahren erfolgte auf Grundlage der pseudonymisierten Probandennummer. In Anlehnung an die Latin-Square-Methode [356, S. 177-181] wurde die Reihenfolge der Führungsverfahren nach der Hälfte der Stichprobe gewechselt, um den Einfluss von Lerneffekten auf das Messergebnis zu minimieren.

Abb. 95 Das Experimentalsystem aus der Sicht eines Probanden kurz vor der Landung in Zone A, das LED-Display zeigt die Rückfrage des Systems „LAND?“

6.3.6 Operationalisierung und Messmethoden

Die Interaktion wurde sowohl mit subjektiven als auch objektiven Messmethoden evaluiert. Die Messung der subjektiven Beanspruchung (WL2.1.1) während der Interaktion erfolgte mittels des in der Luftfahrt etablierten *NASA Task Load Index* (NASA-TLX) in der „Raw“-Variante[90] (d.h. ohne individuelle Gewichtung der Beanspruchungsdimensionen durch den Probanden in einem zweiten Schritt) und in deutscher Ausführung [358]. Dieses als standardisierter Fragebogen ausgeführte Bewertungsinstrument wurde von allen Probanden zeitnah nach jeder Landung ausgefüllt und lieferte ein subjektives Maß für die Dimensionen *geistige Anforderungen*, *körperliche Anforderungen*, *zeitliche Anforderungen*, *Leistung*, *Anstrengung* und *Frustration*. Zur Berechnung der subjektiven Beanspruchung wurde die Summe aller Dimensionen (Wertebereich jeweils von 0 – 20) gebildet. Der verwendete Fragebogen befindet sich in Anhang A9.

Als objektive Bewertungsmaße zur Bestimmung der Effektivität des Führungsverfahrens wurden darüber hinaus erfasst:

- die Anzahl der vom Experimentalsystem empfangenen Flugführungskommandos von Abflug bis Landung (WL2.1.2),
- die Abweichung des Landeortes vom Zentrum der jeweiligen Landezone (WL2.1.3) sowie
- die Flugdauer, gemessen vom empfangenen Abflugkommando bis zum Landekommando (WL2.1.4).

Die Flugführungspräferenzen bei Abflug (WL2.2.1), Transit (WL2.2.2) und Landung (WL2.2.3) wurden anhand der aufgezeichneten Videodaten bestimmt. Das Erfahrungsniveau der Teilnehmer bezüglich der Steuerung von UAVs (WL2.3.1) wurde über den demographischen Fragebogen erfasst und einem von drei Erfahrungsniveaus zugewiesen:

- Erfahrungsniveau A (null bis ein Jahr),
- Erfahrungsniveau B (zwei bis zehn Jahre),
- Erfahrungsniveau C (mehr als zehn Jahre).

[90] In der Literatur finden sich Hinweise darauf, dass die Sensitivität des Raw TLX (RTLX) zuweilen höher, zumindest aber gleichwertig zum originären zweistufigen TLX ist [357]. Im Sinne einer kürzeren Befragungsphase wurde daher der RTLX gewählt.

Die Entfernung der Probanden zum UAV wurde nach jeder Landung manuell gemessen (WL2.3.2). Alle Untersuchungsinhalte und erfassten Messgrößen sind in Tab. 24 dargestellt.

Tab. 24 Übersicht der Messgrößen zur Operationalisierung im Experiment B

Untersuchungsinhalt und Messgrößen	Ausprägung
WL2.1 Flugführungsverfahren	
WL2.1.1 Subjektive Beanspruchung	0 … 120
WL2.1.2 Anzahl erkannter Flugführungskommandos	0 … x
WL2.1.3 Landegenauigkeit	0 … x Meter
WL2.1.4 Flugdauer	0 … x Minuten
WL2.2 Raumnutzungsverhalten	
WL2.2.1 Abflug	Pull-, Push-, Parallel-Methode
WL2.2.2 Transit	
WL2.2.3 Landung	
WL2.3 Einfluss der Erfahrung auf Interaktionsdistanz	
WL2.3.1 Erfahrung mit Steuerung von UAVs	0-1, 2-10, über 10 Jahre
WL2.3.2 Entfernung des Probanden zum UAV nach Landung	0 … x Meter

Falls nicht anders angegeben, wurden alle metrisch skalierten Messgrößen zunächst mittels Shapiro-Wilk-Test (SW-Test) auf Normalverteilung untersucht[91] und anschließend die Mittelwerte der jeweiligen Gruppen über einen gepaarten t-Test (Normalverteilung) bzw. einen Wilcoxon-Test (keine Normalverteilung) auf statistisch signifikante Unterschiede geprüft.

6.3.7 Ergebnisse

Die nachfolgenden Abschnitte beschreiben die ermittelten Ergebnisse der 30 durchgeführten Flüge im Detail. Abb. 96 (KF-Verfahren) und Abb. 97 (DF-Verfahren) zeigen die Trajektorien der durchgeführten Flüge. Dabei konnten alle Probanden die Auftragsziele erreichen, d.h. das Fluggerät von einer Zone in die andere führen, ohne dabei das virtuelle Hindernis zu überfliegen. Vereinzelt wurde die optimale Flughöhe überschritten oder der Sensorbereich verlassen, wodurch die Sichtverbindung kurzzeitig unterbrochen und dadurch einige Eingaben nicht erkannt wurden. Manche Probanden (P04, P14, P15) verwechselten, vergaßen oder führten Flugführungskommandos auch falsch aus, was zu einem unbeabsichtigten Abdriften des UAVs führte (z.B. P15 (pink) während des Transits in Abb. 96). Einige Probanden übersahen auch die Landezone und überschossen den markierten Bereich beim ersten Überflug (z.B. P07 (ockerfarben) bei Zone B in Abb. 97).

[91] Signifikanzniveau: 0,05.

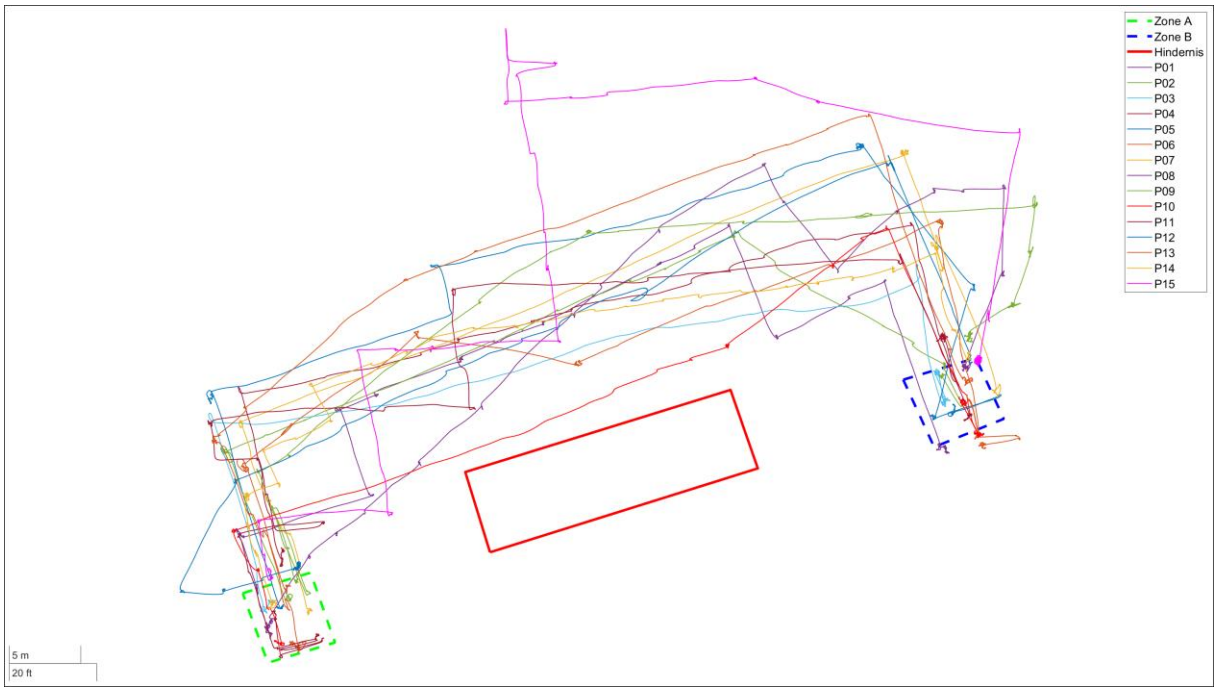


Abb. 96 Trajektorien der Flüge mit kontinuierlichem Führungsverfahren (KF-Verfahren)[92]

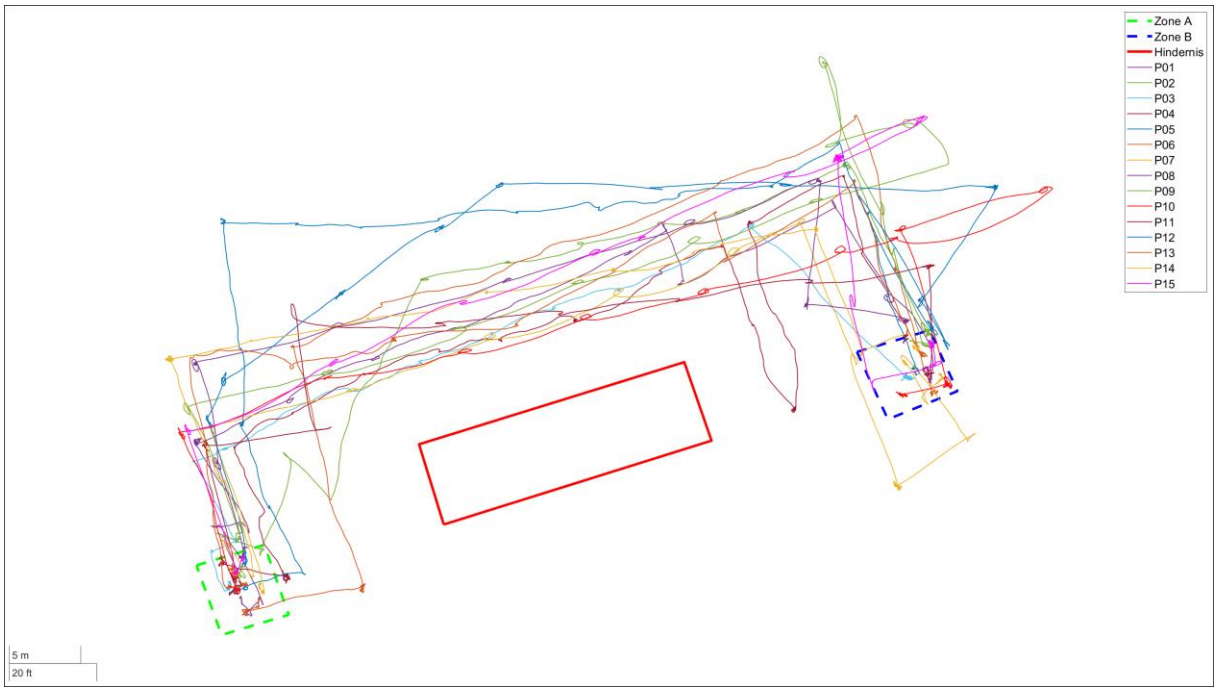


Abb. 97 Trajektorien der Flüge mit diskretem Führungsverfahren (DF-Verfahren)

[92] Die dargestellten Pfade suggerieren Abflüge/Landungen außerhalb der eingezeichneten Zonen A und B. Tatsächlich erfolgten jedoch alle Starts und Landungen innerhalb der Zonenmarkierungen. Die Abweichungen sind auf Ungenauigkeiten in den GPS-Daten während des Schwebeflugs zurückzuführen. Die Messung der Entfernung des Landeortes vom Zentrum der Landezone erfolge nicht auf Basis der GPS-Daten, sondern wurde manuell nach jeder Landung ermittelt.

Subjektive Beanspruchung in Abhängigkeit vom Führungsverfahren (WL2.1.1)

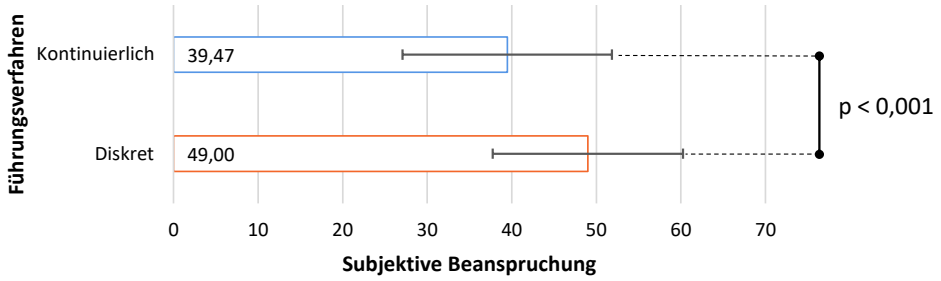


Abb. 98 Subjektive Beanspruchung in Abhängigkeit vom verwendeten Führungsverfahren

In beiden Führungsverfahren lagen normalverteilte Daten ohne Ausreißer vor. In der Gruppe mit dem KF-Verfahren ($M = 39{,}47; SD = 24{,}77$) konnte ein niedrigerer Mittelwert beobachtet werden als in der Gruppe mit dem DF-Verfahren ($M = 49; SD = 22{,}49$). Hierbei konnte ein signifikanter Unterschied ($t(29) = -9{,}81; p < 0{,}001$) in der subjektiven Beanspruchung nachgewiesen werden (Abb. 98).

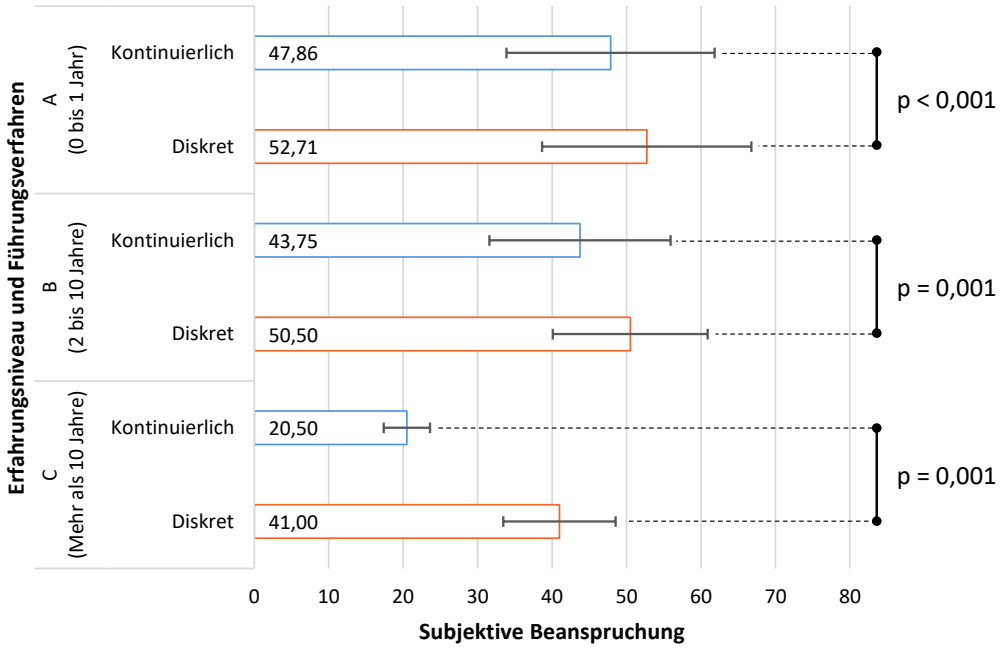


Abb. 99 Subjektive Beanspruchung in Abhängigkeit vom Führungsverfahren gruppiert nach Erfahrungsniveau

Bei einer separaten Betrachtung der Erfahrungsniveaus zeigt sich ein abnehmender Trend bei der subjektiven Beanspruchung je höher das Erfahrungsniveau ausfällt (Abb. 99). Sowohl mit dem Erfahrungsniveau A ($t(13) = -6{,}73; p < 0{,}001$) als auch mit dem Erfahrungsniveau B ($t(7) = -6{,}04; p = 0{,}001$) lag die subjektive Beanspruchung beim KF-Verfahren niedriger. Insbesondere bei den Teilnehmern mit dem Erfahrungsniveau C verringerte das KF-Verfahren die subjektive Beanspruchung um den Faktor zwei ($t(7) = -5{,}27; p = 0{,}001$).

Anzahl empfangener Flugführungskommandos Abhängigkeit vom Führungsverfahren (WL2.1.2)

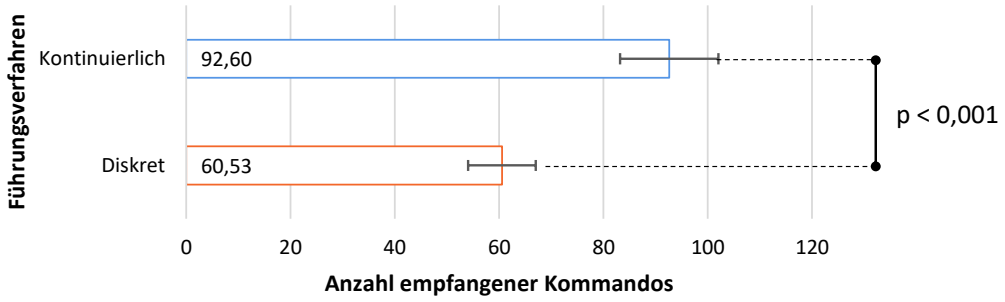


Abb. 100 Anzahl der empfangenen Kommandos pro Flug in Abhängigkeit vom verwendeten Führungsverfahren

In beiden Gruppen lagen normalverteilte Daten ohne Ausreißer vor. In der Gruppe mit dem KF-Verfahren ($M = 92{,}60; SD = 9{,}34$) lag der Mittelwert erwartungsgemäß statistisch signifikant höher ($t(29) = -9{,}81; p < 0{,}001$) als in der Gruppe mit dem DF-Verfahren ($M = 60{,}53$; $SD = 6{,}48$). Die tatsächliche Anzahl der produzierten Kommandos dürfte in beiden Gruppen höher liegen, da nicht alle Kommandos vom Experimentalsystem erfasst wurden.

Landegenauigkeit in Abhängigkeit vom Führungsverfahren (WL2.1.3)

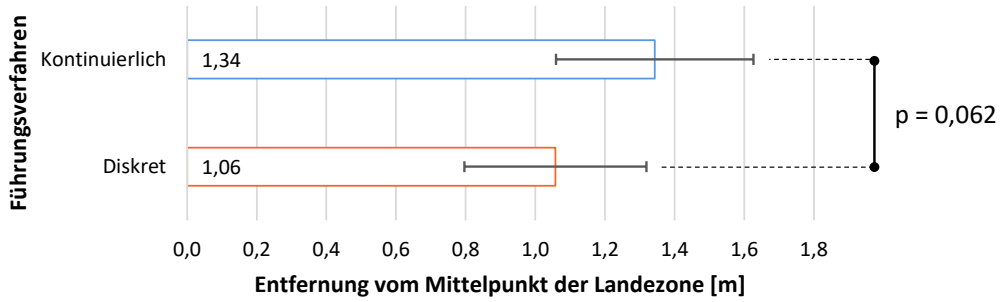


Abb. 101 Landegenauigkeit in Abhängigkeit vom verwendeten Führungsverfahren

In beiden Gruppen lagen normalverteilte Daten ohne Ausreißer vor. Die mittlere Abweichung vom Mittelpunkt der Landezone war beim KF-Verfahren etwas größer ($M = 1{,}34\,m$; $SD = 0{,}28\,m$) als beim DF-Verfahren ($M = 1{,}06\,m$; $SD = 0{,}26\,m$), es konnte jedoch kein statistisch signifikanter Unterschied nachgewiesen werden ($t(29) = 1{,}93; p = 0{,}062$).

Flugdauer in Abhängigkeit vom Führungsverfahren (WL2.1.4)

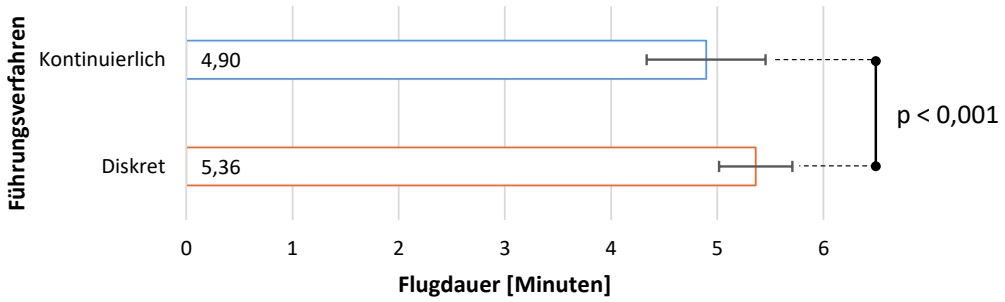


Abb. 102 Flugdauer in Abhängigkeit vom verwendeten Führungsverfahren

In beiden Gruppen lag eine Normalverteilung bei der Flugdauer vor. In der Gruppe mit dem DF-Verfahren gab es einen Ausreißer (P04) mit 8,29 Minuten, der aus der Auswertung ausgeschlossen wurde. Der Mittelwert für einen Flug (Zone A nach B bzw. Zone B nach A) war beim KF-Verfahren ($M = 4{,}9\ Min;\ SD = 0{,}56\ Min$) signifikant niedriger ($t(28) = -20{,}43; p < 0{,}001$) als beim DF-Verfahren ($M = 5{,}36\ Min;\ SD = 0{,}34\ Min$). In Abb. 102 sind die Unterschiede graphisch dargestellt.

Die längere Flugzeit des DF-Verfahrens von etwa 9,4 % lässt sich jedoch nicht eindeutig auf das verwendete Führungsverfahren zurückführen, da beim KF-Verfahren höhere horizontale Fluggeschwindigkeiten genutzt wurden. Ein Vergleich der durchschnittlichen Fluggeschwindigkeiten zeigt einen geringen, aber statistisch signifikanten ($t(28) = 10{,}7; p < 0{,}001$) Unterschied von 10 % (Tab. 25), der in etwa mit der längeren Flugzeit übereinstimmt.

Tab. 25 Vergleich der mittleren Fluggeschwindigkeiten (SD in Klammern)

	KF-Verfahren	DF-Verfahren	Differenz
Mittlere Fluggeschwindigkeit [m/s]	0,44 (0,073)	0,40 (0,050)	10%

Flugführungspräferenzen bei Abflug, Transit und Landung (WL2.2.1, WL2.2.2 und WL2.2.3)

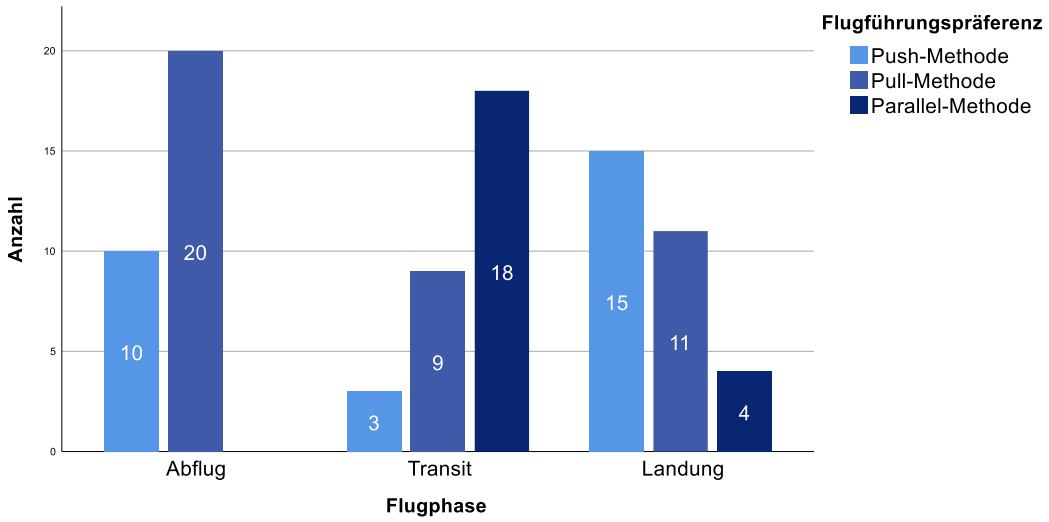


Abb. 103 Flugführungspräferenzen in Abhängigkeit von der Flugphase

Bei den insgesamt 30 durchgeführten Flügen wählten beim Abflug zwei Drittel der Probanden die Pull-Methode, die Verbleibenden nutzten die Push-Methode. Die Parallel-Methode wurde hingegen primär in der Transitphase gewählt (60%), gefolgt von der Pull-Methode (30%). In einigen wenigen Fällen wurde die Push-Methode für den Transit genutzt (10%). In der Landephase drehte die Hälfte der Teilnehmer das Fluggerät zunächst um die Hochachse und steuerte es dann von sich weg in den Landebereich (Push-Methode) hinein. Knapp 37% bevorzugten es, das Fluggerät hinter sich herziehend in der Landezone zu platzieren (Pull-Methode). Ca. 13% wählten hierfür stattdessen die Parallel-Methode.

Interaktionsentfernungen (WL2.3.1 und WL2.3.2)

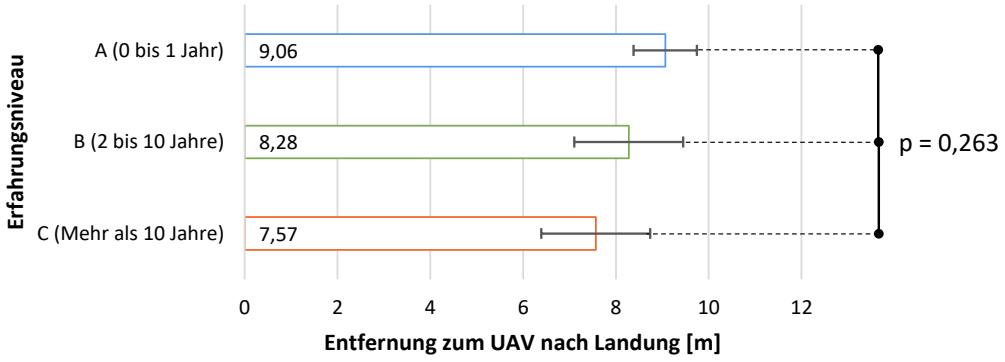


Abb. 104 Entfernung zum UAV nach der Landung in Abhängigkeit vom Erfahrungsniveau

Nach der Landung befanden sich die Probanden mindestens 4,7 m und maximal 16,1 m vom UAV entfernt. Aus dem Erfahrungsniveau A (N=13) wurde ein Ausreißer (P03) mit 16,1 Metern aus der Auswertung ausgeschlossen. In der Gruppe mit dem Erfahrungsniveau C (N=7) wurde bei einem Flug die Entfernung nicht erfasst und daher der Messwert ebenfalls aus der Auswertung ausgeschlossen. Alle übrigen Stichproben wurden dem mittleren Erfahrungsniveau B zugeordnet (N=8). Ein SW-Test konnte in allen Gruppen eine Normalverteilung der Daten nachweisen, jedoch lieferte eine einfaktorielle Varianzanalyse (ANOVA) keine statistisch signifikanten Gruppenunterschiede bezüglich der Entfernung zum UAV nach dessen Landung ($p = 0{,}263$).

6.3.8 Diskussion

Die kollokierte Interaktion mit dem fliegenden Experimentalsystem wurde von den Probanden insgesamt positiv aufgenommen (P13: „Hat Spaß gemacht.“) und die gewählten Fluggeschwindigkeiten auch von flugerfahrenen Teilnehmern als angemessen bewertet (P14).

Verzögerungen während der Interaktion

Während der Interaktion kam es in einigen Fällen zu Verzögerungen, deren Ursachen in den Umgebungsbedingungen, dem Experimentalsystem und dem Verhalten der Probanden verortet werden können. Bei realitätsnahen Experimenten im Freien können nicht alle Umgebungsbedingungen (Bewölkungsgrad, Sonnenstand, Windverhältnisse) kontrolliert werden. Um dennoch vergleichbare Versuchsbedingungen für alle Probanden zu ermöglichen, wurden Flüge nur an solchen Tagen durchgeführt, an denen die zulässigen Betriebsgrenzen[93] der Flugplattform eingehalten werden konnten. Die Landungen erfolgten bei allen Flügen automatisiert, sodass, je nach vorherrschender Windbedingung, die finale Landeposition des UAVs von der beabsichtigten Position abweichen konnte. Mit der gewählten Systemkonfiguration konnte jedoch jeder Versuchsteilnehmer – unabhängig vom Erfahrungsniveau – den Multikopter innerhalb des Ziellandebereichs von 5x5 Metern platzieren. Einige Probanden nahmen hierbei eine Priorisierung der Missionsziele vor und bevorzugten eine zügige Landung gegenüber einer genauen (P15: *„Wenn der Akku nicht so niedrig gewesen wäre, hätte ich die Mitte besser getroffen.“*). Im Laufe der Studie nahm zudem der Grasbewuchs auf dem Testgelände zu, wodurch die Erkennung der bodenseitig mit Absperrband markierten Landezonen für einige Probanden (P11 bis P15) erschwert wurde. In diesen Fällen erfolgte jedoch eine Hilfestellung durch die Versuchsleitung bei der Orientierung.

[93] Die vom Experimentalsystems unterstützte maximale Windgeschwindigkeit lag bei 8 m/s.

Systembedingt gab es zudem Verzögerungen bei sehr geringen Interaktionsabständen (unter 5 m), wobei die Sensornachführung für einige Probanden nicht schnell genug reagierte, sodass vereinzelt der Wunsch nach einer besseren Nutzerverfolgung geäußert wurde (P12: *„Wenn der Gimbal einen noch mehr verfolgen würde, wäre es einfacher.“*). Darüber hinaus wurden bei einigen wenigen Probanden Flugführungskommandos falsch vom Experimentalsystem interpretiert (z.B. `CCW` als `FRONT` bei P10) oder erst nach mehrmaliger Wiederholung erkannt, wodurch Korrekturen notwendig wurden.

Verzögerungen traten aber auch aufgrund des Verhaltens der Probanden auf. So wurde während der Produktion nicht immer der Blickkontakt zum UAV aufrechterhalten (P08) oder der Erfassungsbereich des Sensors verlassen. Einigen bereitete auch die korrekte Einschätzung der Orientierung des UAVs Probleme, sodass Drehrichtungen verwechselt wurden (P08, P14). Einzelne Kommandos wurden auch vergessen (P04, P15), verwechselt (`FRONT` mit `DOWN` bei P15) oder zu ungenau, d.h. mit einer falschen Bewegungsform und -geschwindigkeit, ausgeführt (P04). Das Klassifizierungssystem war jedoch in der Lage, auch leicht abgewandelte Darstellungen (z.B. `UP/DOWN` Kommandos von P13 und P15) bis zu einem gewissen Grad korrekt zu identifizieren.

Sichtverbindung und visuelle Rückmeldeverfahren

Der visuelle Rückkanal und das gewählte Anzeigekonzept wurde von den Probanden als „guter Indikator“ (P09) für die Rückmeldung über den Systemstatus bewertet. Jedoch wurde auch der Wunsch geäußert, eine zusätzliche Signalisierung zu integrieren, falls der Nutzer zu nah oder zu weit weg vom UAV und damit außerhalb des Sensorbereichs ist (P08). Es wurde zudem angemerkt, dass das UAV in den Schwebeflug übergehen sollte, sobald die Sichtverbindung zum Nutzer während der Interaktion unterbrochen ist (P06).

Einfluss des Führungsverfahrens auf die subjektive Beanspruchung, Kontrollierbarkeit und Vorhersagbarkeit

Bei der offenen Befragung beim Debriefing war keine eindeutige Tendenz bei der Bevorzugung eines Führungsverfahrens erkennbar. Das kontinuierliche Führungsverfahren wurde von den Probanden als „einfacher zu handhaben“ (P14) und „fehlertoleranter“ (P10) bewertet und ermöglichte dadurch eine bessere „Feinjustierung“ (P15) und „mehr Vertrauen ins Fluggerät“ (P14). P10 vertrat die Meinung, dass die Erkennung der Flugführungskommandos in diesem Modus besser funktioniert hätte. Gleichzeitig wurde die Interaktion als „stressiger“ und „anstrengender“ (P15) empfunden. Die automatische Unterbrechung der Flugbewegungen und das daraus resultierende schaukelnde Flugverhalten des UAVs wurde zudem als „fehlendes Verständnis für kontinuierliche Bewegungen“ (P11) interpretiert. P12 beschrieb das Verfahren als „wesentlich besser zu steuern (...), aber anspruchsvoller, [du] musst die ganze Zeit etwas machen“.

Das diskrete Führungsverfahren wurde hingegen von den Teilnehmern als „einfacher“ (P08), „körperlich weniger anstrengend“ (P09, P12), „praktikabler“ (P14), „entspannter und angenehmer“ (P07) beschrieben. Während ein Teil der Probanden die durchgängige Flugbewegung ohne Eingaben als Vorteil empfand (P06: „Ich weiß, was das Gerät in den nächsten 10 Sekunden machen wird.“, P14: „Einfacherer Modus für den Transit (…) kann mehr auf die Umgebung schauen.“), war sich ein anderer Teil nicht sicher, ob es das Fluggerät zuverlässig zum Anhalten bringen kann (P12: „`STOP` muss er erkennen (…), bin mir nicht sicher, ob das immer hinhaut.“). Insofern wurde der automatische Übergang in den Schwebeflug nach einem Zeitablauf als sinnvolle Absicherung erachtet (P13). P14 hat bei diesem Modus die Latenz „mehr gespürt beim Positionieren“. P07 beschrieb dieses Verfahren als „arbeitsintensiver, aber sicherer falls man beim Schauen, Steuern und Laufen hinfällt“.

Verbesserungsvorschläge und Schlussfolgerungen

Die durchschnittliche Systemlatenz von ca. 1,7 Sekunden wurde von einigen Probanden (P10, P11, P12) als Kritikpunkt aufgeführt, wodurch eine präzise Positionierung des UAVs vor allem innerhalb der Landezonen erschwert wurde. Auch wurden variable Fluggeschwindigkeiten in Abhängigkeit von der Flugrichtung (z.B. langsamere Bewegungen auf den Nutzer zu) sowie eine bessere Stabilisierung des Fluggerätes in Bodennähe gefordert (P13). P05 empfand das UAV-zentrierte Koordinatensystem als nicht intuitiv und schlug vor, das Rotationszentrum an die Position des Nutzers zu verlagern und das UAV stets in Richtung des Nutzers auszurichten. Um die Bedenken über unabsichtlich ausgelöste Flugbewegungen zu reduzieren, wurde zudem ein dediziertes Signal für einen zeitlich befristeten „Hör-mir-zu-Modus“ (P12) vorgeschlagen, in dem das UAV die nonverbalen Körpersignale analysiert und welcher auch durch ein entsprechendes Signal wieder beendet werden kann.

Zusammenfassend lässt sich in den Kommentaren ein Wunsch nach einem intelligenterem und situationsangepassten Verhalten des UAVs erkennen, welches einerseits über intelligente Funktionen wie einem „Follow-Me-Modus“ (P08, P10) für längere Flugabschnitte verfügt, gleichzeitig aber bei Bedarf auch in einen geführten Modus geschaltet werden kann (P15), der eine präzise Einweisung erlaubt. Aufgrund der geringeren subjektiven Beanspruchung und der Ähnlichkeit zu bestehenden Verfahren aus der Luftfahrt und dem KFZ-Wesen, scheint das kontinuierliche Führungsverfahren das geeignetere Werkzeug für die Positionierung des UAVs zu sein. Um die Positioniergenauigkeit weiter zu erhöhen, sollten die Bewegungsinkremente zudem variabel sein und beispielsweise an die Ausführungsgeschwindigkeit der Flugführungskommandos und die Bewegungsrichtung (zum Nutzer hin oder von ihm weg) gekoppelt werden.

6.4 Experiment C: nonverbaler Dialog

Wenn ein hoher Lärmpegel oder eine große Distanz eine verbale Kommunikation unmöglich machen, greifen Menschen oft zu nonverbalen Ausdrucksmitteln in Form von Handzeichen zurück. Wenn keine gemeinsame Sprache in Form eines nonverbalen Zeichensystems vorliegt, beschränken sich die Kommunikationsinhalte oftmals auf einfache, singuläre Informationseinheiten, die im jeweiligen situativen Kontext von Belang sind. Insofern stellt unter diesen Bedingungen ein darüber hinaus gehender Dialog bereits eine Herausforderung für die Kommunikation dar. Ersetzt man nun einen der Kommunikationspartner durch ein technisches System und stellt beiden Gesprächspartnern ein gemeinsames, nonverbales Vokabular zur Verfügung, ergeben sich jedoch neuartige Anwendungsmöglichkeiten nicht nur in der cHDI, sondern generell in der nonverbalen Interaktion mit technischen Systemen. Eine mögliche Anwendung stellt hierbei die visuelle Beauftragung eines UAVs dar, bei der ein Nutzer nonverbal seine Intention beschreibt, die vom Kommunikationssystem in eine Dienstanforderung übersetzt wird. Ein solcher Anwendungsfall wurde in Form eines nonverbalen Dialogs zwischen einer intensiv geschulten Nutzergruppe (Experten) und einem intelligenten UAV in Form des Experimentalsystems nachgestellt.

6.4.1 Untersuchungsinhalte

Diese Untersuchung wurde durchgeführt, um Antworten auf die Forschungsfragen aus dem Bereich der Wissenslücke 3 (WL3) zu finden und die Hypothese 3 (H3) zu überprüfen. Im Detail wurden folgende Fragestellungen untersucht.

H3 Einfluss der Dialogunterstützungsfunktionen auf den Dialogerfolg

Im Gegensatz zu klassischen, graphischen Benutzeroberflächen, bei denen die verfügbaren Eingabeoptionen für Nutzer direkt sichtbar sind (z.B. über Drop-Down-Menüs), sind diese Auswahlmöglichkeiten bei natürlichen Benutzerschnittstellen nicht so offensichtlich. Stattdessen sind sie hierbei auf eine mentale Vorstellung vom aktuellen Zustand des Systems und den in der jeweiligen Situation möglicherweise verfügbaren Optionen angewiesen. Eine Möglichkeit, Nutzer in diesen Situationen zu unterstützen, stellt die Visualisierung des gegenwärtigen Systemzustands über einen visuellen Rückkanal dar, wodurch Nutzer einen Einblick in die Erwartungshaltung des Systems erhalten. Gleichzeitig kann hierüber das System den Dialogverlauf aktiv beeinflussen. Im Dialogmanager des Dialogsystems wurde daher eine zu- und abschaltbare Kommunikationskomponente integriert, die die fehlenden Informationen der wahrscheinlichsten Nutzerintentionen über den visuellen Rückkanal in Form von Optionsvorschlägen nach außen führt (siehe Kapitel 5.3.1) und damit eine der zwei Dialogunterstützungsfunktionen aus der Fähigkeitslücke 3 darstellt. Die erste Frage ist daher, welchen Einfluss solche Optionsvorschläge des Systems auf den Dialogerfolg haben oder anders ausgedrückt: sollte das System den Nutzer während der Intentionsfindung aktiv begleiten oder sich so lange passiv verhalten, bis eine Intention mit hinreichend hoher Konfidenz erkannt wurde?

Die zweite Dialogunterstützungsfunktion, in Form einer kontextsensitiven Signalverarbeitung, ermöglicht es dem Dialogsystem, den Dialogverlauf während der Interaktion „mitzuverfolgen“ und die für den jeweiligen Kontext irrelevanten Körpersignale auszublenden. Gleichzeitig bietet dieser Mechanismus Nutzern mehr „Bewegungsfreiheit“ während der Interaktion, da unbewusst durchgeführte Bewegungen nicht sofort eine Reaktion des Systems auslösen. Im Umkehrschluss bedeutet das jedoch auch, dass durch dieses Verfahren der Eingaberaum der Nutzer eingeschränkt wird und somit nicht in jeder Gesprächssituation das gesamte Vokabular für die Interaktion zur Verfügung steht. Durch die Deaktivierung der Kontextsensitivität kann diese Einschränkung aufgehoben werden, verlangt von Nutzern jedoch eine höhere „nonverbale Disziplin“ bei der Interaktion. Als weitere Folge entfällt auch die Fähigkeit des Systems, mehrdeutige Signale auflösen zu können, sodass Alternativen in Form von eindeutigen Synonymen benötigt werden. Es stellt sich daher die Frage, in welchem Maße der Dialogerfolg durch die Kontextsensitivität des Systems beeinträchtigt wird.

WL3.1 Faktoren zur Begünstigung von Verständigungsproblemen bei nonverbalem Dialog

Ein idealer aufgabenzentrierter Mensch-Maschine-Dialog besteht aus so wenigen Interaktionen (Turns) wie möglich [359, S. 761]. In Dialogsituationen mit mehrdeutigen Signalen können jedoch schnell Missverständnisse entstehen und zu Kommunikationsproblemen führen, welche die Interaktion unnötig verlängern und sie dadurch ineffizient werden lassen. Ein weiterer wichtiger Untersuchungspunkt ist daher die Frage, welche Faktoren zu Missverständnissen beitragen.

WL3.2 Nonverbale Lösungsstrategien bei Verständigungsproblemen

In diesem Kontext erwächst auch die Frage, welches nonverbale Problemlösungsverhalten Nutzer in diesen Situationen zeigen und ob Verhaltensänderungen bei anhaltenden Verständigungsproblemen auftreten.

WL3.3 Bestimmung der Frustrationsgrenzen

Im Hinblick auf den möglichst optimalen Einsatz von Unterstützungsfunktionen ist es darüber hinaus wichtig zu wissen, wo die Frustrationsgrenzen liegen, d.h. nach welcher Anzahl an Eingaben (WL3.3.1) und nach welcher Zeitdauer (WL3.3.2) Nutzer den Dialog abbrechen.

6.4.2 Untersuchungsdesign

In einer kontrollierten Laborumgebung wurde eine Dialogsituation mit dem Experimentalsystem nachgestellt. Da der Schwerpunkt der Untersuchung auf den Dialogverhalten lag, wurde auf eine Interaktion im Realflug verzichtet, um eine höhere Reproduzierbarkeit zu erlauben bei gleichzeitig weniger äußerer Faktoren. Zudem wurde ein separater Monitor anstatt der integrierten LED-Matrix als visueller Rückkanal genutzt. Dieser bietet mehr Raum für die gleichzeitige Darstellung mehrerer Inhalte und verkürzt die Interaktionszeit, die sonst durch Scrolling-Vorgänge verlängert worden wäre.

Versuchsaufbau

Die Experimente wurden in einem Versuchslabor der Universität der Bundeswehr München in Neubiberg durchgeführt. Die Probanden standen dabei wie bei Experiment A (Kapitel 6.1) ca. drei Meter vom Experimentalsystem (ohne fliegende Komponente) entfernt und blickten auf einen 25“-Monitor, der auf Augenhöhe befestigt war (Abb. 105). Die Versuchsleitung befand sich dabei in einiger Entfernung links vom Probanden. Zusätzlich wurde rechts vom Probanden ein großer, mobiler Bildschirm aufgestellt, auf dem das verfügbare Vokabular mit den entsprechenden Darstellungen abgebildet war. Dadurch sollte der Einfluss der Gedächtnisleistung auf den Interaktionserfolg minimiert und allen Probanden dieselben Ausgangsvoraussetzungen ermöglicht werden.

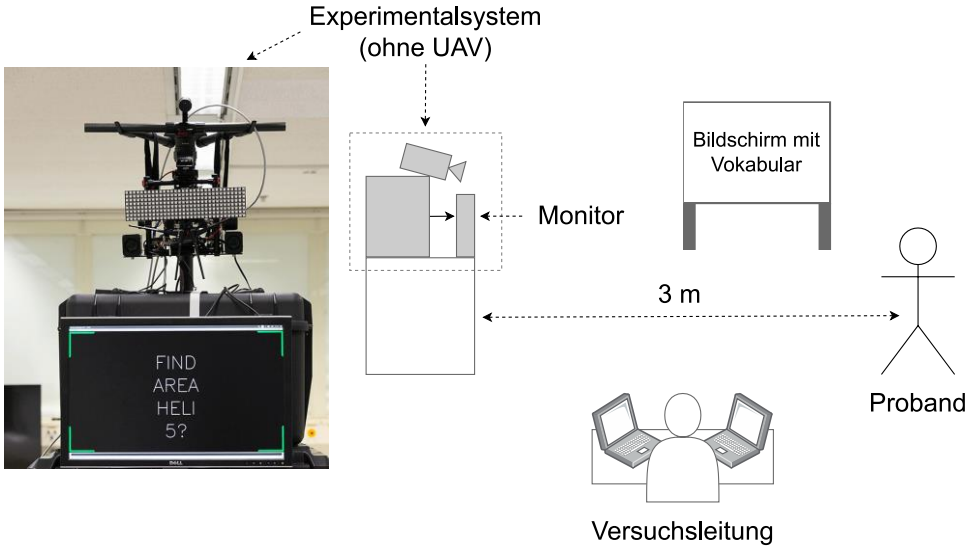


Abb. 105 Laboraufbau im Experiment C

Kommunikationsinhalte

Für die Untersuchung wurde der *Briefing Mode* des VisCom-Experimentalsystems genutzt, der über das größte Vokabular verfügt und eine Vielzahl von Diensten beinhaltet. Aus dieser Menge an Diensten wurden 12 Dienstbeschreibungen erstellt, die sich an typischen UAV-bezogenen Anwendungen (Verlegung, Transport, Suche, Dokumentation, Unterstützung, Parameterabfrage) orientieren. In Tab. 26 sind die dazugehörigen Dienstbeschreibungen dargestellt, die die Probanden im Laufe der Interaktion an das Experimentalsystem übermitteln sollten. Die im Wortschatz enthaltenen Vokabeln sind dabei in Großbuchstaben dargestellt. In Kleinbuchstaben geschriebene Füllwörter, die nur zum besseren Verständnis in der Dienstbeschreibung enthalten sind, mussten hingegen nicht übermittelt werden und waren auch nicht im Vokabular enthalten.

Tab. 26 enthält zudem Anmerkungen und Hervorhebungen, die für die Probanden nicht sichtbar waren. Mehrdeutige Signale sind mit einem hochgestellten „M“ markiert. Falls im Vokabular eine alternative Darstellung (Synonym) des Signals enthalten ist, ist das entsprechende Signal mit einem hochgestellten „S“ markiert. Diese Synonyme erlauben somit auch eine Übermittlung der Dienstbeschreibungen der Klasse A bei deaktivierter Kontextsensitivität des Systems, setzen jedoch eine akkurate Darstellung der Körpersignale voraus. Dienstbeschreibungen der Klasse B können aufgrund der enthaltenen mehrdeutigen Signale und der fehlenden Synonyme bei deaktivierter Kontextsensitivität nicht übermittelt werden und wurden daher für die Bestimmung der Frustrationsgrenze genutzt. Zur besseren Unterscheidung sind die Anweisungen der Klasse B zusätzlich grau hinterlegt. Für die Rückmeldesignale (YES, NO) standen immer visuelle Synonyme zur Verfügung (siehe semantischer Kontext „Feedback“ im Anhang A7).

Tab. 26 Übersicht der zu übermittelnden Dienstbeschreibungen inklusive Anmerkungen

Nr.	Dienstbeschreibung	Intention	Klasse
1	TRANSPORT BOX$^{M, S}$ on the LEFT to HOME	UAV soll Box (links vom Nutzer) zur Basis transportieren	A
2	FIND AREA$^{M, S}$ for HELICOPTER^{S} within 5^{S}	UAV soll geeigneten Landeplatz für einen Helikopter im Umkreis von 5 (generischen) Einheiten finden	A
3	LIFT BOX$^{M, S}$ on the RIGHT	UAV soll Box (rechts vom Nutzer) aufnehmen	A
4	SCAN^{S} TREE^{S} on the LEFT	UAV soll eine Rundumaufnahme des Baumes (links vom Nutzer) erzeugen	A
5	SCAN^{S} within RADIUS 2^{S}	UAV soll Umgebung im Umkreis von 2 Einheiten dokumentieren	A
6	FOLLOW ME^{M}	UAV soll Nutzer folgen	A
7	FLY to SECRET^{M} spot	UAV soll zum geheimen (vorher vereinbarten) Ort verlegen	B
8	SHOW HOME	UAV soll Nutzer den Weg zur Basis zeigen	A
9	LAND HERE	UAV soll in der Nähe des Nutzers landen	A
10	ASK parameter BATTERY^{M}	UAV soll seinen aktuellen Ladezustand mitteilen	B
11	ASK parameter HEIGHT^{M}	UAV soll die aktuelle Flughöhe mitteilen	B
12	ASK parameter HUMAN^{M}	UAV soll mitteilen, wie viele Menschen es gegenwärtig wahrnimmt	B

^{M}: mehrdeutiges Signal; ^{S}: Visuelles Synonym verfügbar (für YES/NO immer vorhanden);
gestrichelter Bereich: Anweisungen für Probanden (ohne Anmerkungen); **klein geschriebene Wörter**: müssen nicht übermittelt werden;
grau hinterlegt: bei deaktivierter Kontextsensitivität nicht an Experimentalsystem übermittelbar

Während des Experiments erhielten die Probanden die Dienstbeschreibungen nur in einer stark reduzierten Form (gestrichelter Bereich in Tab. 26) ohne zusätzliche Anmerkungen und Hervorhebungen.

Dialogbezogene Stolpersteine zur Förderung von Missverständnissen

Um Missverständnisse im Dialogverlauf zu fördern, wurden Mehrdeutigkeiten in das Vokabular eingebaut und Signalpaare gebildet, die jeweils dieselbe Darstellung nutzen, aber je nach Gesprächskontext unterschiedliche Signale transportieren (Tab. 27).

Tab. 27 Signalpaare mit identischen Darstellungen

Mehrdeutigkeit durch Synonyme auflösbar	Mehrdeutigkeit nicht auflösbar
BOX* – AREA*	BATTERY – FLY
	ME – HUMAN

*: dynamische Darstellung

Daneben wurde eine weitere Gruppe von Signalpaaren gebildet, deren Darstellungen sich auf den ersten Blick ähneln, im Detail jedoch – z.B. durch die Handform – unterscheiden (Tab. 28).

Tab. 28 Signalpaare mit minimal unterschiedlichen Darstellungen

Mehrdeutigkeit durch Synonyme auflösbar	Mehrdeutigkeit nicht auflösbar
AREA – HEIGHT	SECRET – SHOW

YES – BOX
YES/NO – LEFT/RIGHT

Eine weitere Hürde stellte die Deaktivierung der Kontextsensitivität des Experimentalsystems dar, wodurch eine höhere Sensitivität in Bezug auf nebenläufige und unbeabsichtigte Körpersignale erreicht wurde. Als Folge stiegen die Anforderungen an die Ausführungstreue (z.B. `ME/HUMAN` vs. `IGNORE`) und Ausführungsgeschwindigkeit bei der Produktion der Signale und erforderten auch eine höhere Disziplin[94] vom Nutzer während der Interaktion. Abweichungen von der optimalen Darstellung führten somit schneller zu Signalverwechslungen. Das war beispielsweise dann der Fall, wenn bei der Darstellung des Signals `RIGHT` der Daumen ungewollt ausgestreckt und somit fälschlicherweise das Signal YES vom System erkannt wurde. Empfing das System zudem singuläre Rückmeldungen, ohne dass zuvor eine Frage gestellt worden war, reagierte es mit störenden Zwischenfragen ohne erkennbaren Sinn (z.B. `CONFIRM YES?`).

Vortest

Vor dem Beginn der Untersuchungen wurde ein Probelauf mit einer Freiwilligen durchgeführt.

6.4.3 Stichprobenbeschreibung

An diesem Experiment nahmen 20 Probanden teil (zwei davon weiblich). Alle Teilnehmer verfügten über einen direkten Luftfahrtbezug durch Studium der Luft- und Raumfahrttechnik bzw. durch die berufliche Tätigkeit am Institut für Flugsysteme. Drei der Probanden verfügten zudem über eine militärische Ausbildung. Keiner der Teilnehmer hatte Erfahrungen mit Handzeichensystemen. Die vertretene Altersspanne reichte von 21 bis 55 Jahren, mit einem durchschnittlichen Alter von 30 Jahren (SD=7,7).

6.4.4 Datenerhebung

Die Daten wurden im Zeitraum vom 27.04.2021 bis 25.05.2021 in einem Versuchslabor der Universität der Bundeswehr München erhoben. Zu Beginn des Experiments wurden demographische Daten der Teilnehmer per Fragebogen erfasst. Anschließend wurde die Interaktion per Audio- und Videoaufzeichnung dokumentiert. Wichtige Parameter (Sensordaten sowie Klassifikationsergebnisse) des Experimentalsystems wurden ebenfalls aufgezeichnet. Eine Übersicht der ermittelten Daten befindet sich in Anhang A12.

6.4.5 Versuchsdurchführung

Beim Experiment C wurde das in Abb. 106 dargestellte Ablaufschema verwendet. Jeder Teilnehmer wurde zunächst basierend auf der Probandennummer einer Gruppe zugewiesen. Probanden mit gerader Nummer wurden der *Gruppe M* zugeordnet und erfuhren ein aktives Dialogverhalten des Systems (M = mit Optionsanzeige), während Probanden mit einer ungeraden Nummer der *Gruppe O* zugeordnet wurden und ein passives Dialogverhalten des Systems erfuhren (O = ohne Optionsanzeige).

Training

Nach dem Briefing wurde jedem Probanden das Vokabular und die Syntax nach einem festgelegten Trainingsablauf vorgestellt. Dabei wurde jeder der elf semantischen Kontexte separat betrachtet und die dazugehörigen nonverbalen Körpersignalen trainiert. Den Probanden stand es dabei frei, die korrekte Darstellung so lange zu üben, bis sie zuversichtlich waren, das entsprechende Signal zuverlässig

[94] Position und Verhalten der passiven Hand, Darstellungsdauer der Signale, Neutrale Pose, korrekte Handform.

übermitteln zu können. Die aufgezeichneten Darstellungen während dieser Phase des Experiments bildeten gleichzeitig die Datenbasis für die Evaluierung der Klassifikationsleistung des Experimentalsystems unter kontrollierten Bedingungen (siehe Auswertung in Kapitel 6.1). Anschließend wurde die Syntax für die Beschreibung der Dienstanforderungen anhand von elf Beispielen erläutert (siehe Syntax-Block in Anhang A7). Nach dem Training wurde die Syntax jedoch ausgeblendet, um den Einfluss der Systemrückmeldungen auf den Dialogverlauf besser untersuchen zu können.

Durchgang 1 (Between-Subjects-Design)

Im nächsten Schritt folgte der erste Durchgang mit aktivierter Kontextsensitivität des Experimentalsystems als Basisszenario. Jeder Proband zog jeweils in zufälliger Reihenfolge eine von zwölf Intentionen aus einem Briefumschlag und begann mit der Übermittlung der niedergeschriebenen Dienstbeschreibung. Die Probanden wurden instruiert, während der Interaktion laut mitzusprechen, um einerseits ihre Absichten gegenüber der Versuchsleitung deutlich zu machen (Ground Truth für die Auswertung) und andererseits einen Einblick in ihre Gedankengänge im Laufe der Interaktion zu ermöglichen. Die Übermittlung wurde als erfolgreich gewertet, wenn die vom System erstellte Zusammenfassung über die erkannte Intention deckungsgleich mit der gezogenen Dienstbeschreibung war und vom Probanden erfolgreich rückbestätigt wurde.

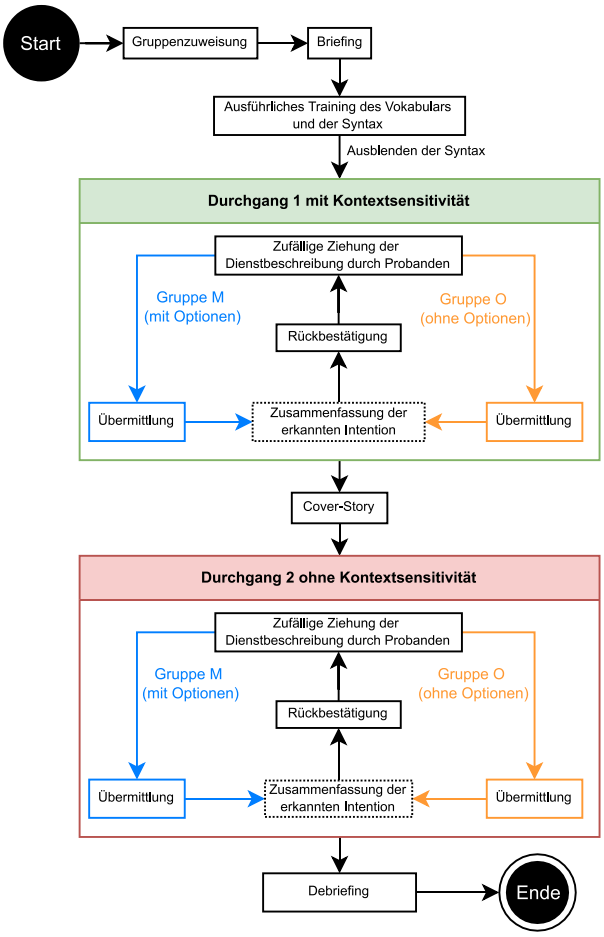


Abb. 106 Ablaufschema des Experiments C

Cover-Story und Durchgang 2 (Within-Subjects-Design)

Um die Probanden für einen zweiten Durchlauf zu motivieren, wurde eine Cover-Story genutzt, wonach eine Wiederholung des Durchgangs eine bessere Datengrundlage und damit eine höhere Validi-

tät der Ergebnisse liefern würde. Alle Probanden willigten zu einem zweiten Durchlauf ein und erwarteten dasselbe Systemverhalten. Im zweiten Durchlauf wurde jedoch die Kontextsensitivität des Systems deaktiviert und die Auswirkungen auf den Dialogverlauf beobachtet. Im Unterschied zum ersten Durchlauf, konnten nun als Folge die Intentionen der Klasse B prinzipbedingt nicht übermittelt werden, weil geeignete Synonyme für mehrdeutige Signale im Vokabular fehlten. Bei länger andauernden Interaktionsunterbrechungen wurden die Probanden von der Versuchsleitung gefragt, ob sie weiter probieren oder zur nächsten Dienstbeschreibung weitergehen möchten. Nach dem zweiten Durchgang fand ein ausführliches Debriefing mit den Probanden statt, bei dem der eigentliche Zweck des zweiten Durchlaufs offenbart und etwaige Fragen durch die Versuchsleitung beantwortet wurden.

6.4.6 Operationalisierung und Messmethoden

Die Mensch-UAV-Interaktion wurde mithilfe der aufgezeichneten Videodaten analysiert und das Verhalten der Probanden nach einem definierten Beobachtungssystem manuell transkribiert. Während der Interaktion wurde das vom Probanden beabsichtigte (laut ausgesprochene) Signal, das gesendete (dargestellte) Signal sowie das vom System erkannte Signal dokumentiert und daraus potentielle Fehlerursachen identifiziert und die Anzahl der Nutzereingaben sowie Systemausgaben berechnet. Für die Dokumentation der Ursachen von Kommunikationsproblemen wurde das Kodierschema in Tab. 29 genutzt.

Tab. 29 Verwendetes Kodierschema A zur Annotation der Ursachen von Kommunikationsproblemen

Ursache	Kodierung
Falsches Signal übermittelt	Fsi
Falsche Ausführung	Fa
Falscher Kontext	Fk
Falsche Syntax	Fsy
Systemausgabe übersehen	Aü
Rückbestätigung vergessen	Bv
Rückbestätigung ungewollt ausgelöst	Ba
Ungewollt übermitteltes Signal angenommen	Ea
Erwartet Anzeige der Zusammenfassung	Ezf
Erwartet Anzeige von Optionen	Eo
Zusammenfassung falsch	Fzf
Systemfehler	Sf

Die Reaktionen der Probanden zur Behebung des Problems wurden hingegen nach dem Kodierschema in Tab. 30 annotiert.

Tab. 30 Verwendetes Kodierschema B zur Annotation der Reaktionen nach Kommunikationsproblemen

Reaktion	Kodierung
Identische Wiederholung der Darstellung	Wi
Wiederholung mit langsamerer/schnellerer Darstellung	Wl/Ws
Wiederholung mit kleinerer/größerer Darstellung	Wk/Wg
Wiederholung mit veränderter Position der Arme/Hände	Wp
Wiederholung mit veränderter Handform	Wf
Wiederholung mit überdeutlicher/klopfender Darstellung	Wü
Veränderung der Körperhaltung	Kö
Übergang zum nächsten Signal	Nxt
Ablehnung der Systemausgabe	Abl
Test eines anderen Signals	Test
Test einer anderen Dienstbeschreibung	And
Wiederholung der letzten Systemausgabe	What
Bestätigung der Systemausgabe	Best
Revision der letzten Eingabe	Ignore

Zurücksetzen des Dialogs	Reset
Anforderung von Hilfe	Help
Abbruch des Dialogs	Abb

Die Dauer des Dialogs wurde von dem Zeitpunkt an gemessen, ab dem die erste, an das System gerichtete Körperbewegung des Probanden sichtbar war bis zu dem Zeitpunkt, an dem das System die Speicherung der Dienstanforderung bestätigt hat (Anzeige von „SAVED“ bzw. des angefragten Systemparameters) oder der Dialog vom Probanden abgebrochen wurde. Angelehnt an das Konzept der „Gebrauchstauglichkeit“ (Usability) [360] wurde der Dialogerfolg anhand der Effektivität (Anteil erfolgreicher Übermittlungen) und der Effizienz (bezogen auf Anzahl der Eingaben sowie die Dialogdauer) der Interaktion bewertet. Die Messung der subjektiven Zufriedenstellung (Satisfaction) erfolgte hierbei nicht. In Tab. 31 sind alle Untersuchungsinhalte und Messgrößen in der Übersicht dargestellt.

Tab. 31 Übersicht der Untersuchungsinhalte und Messgrößen zur Operationalisierung im Experiment C

Untersuchungsinhalt und Messgröße	Betrachtete Dienstklasse	Datenbasis (Durchgang)	Ausprägung
H3 Einfluss der Dialogunterstützungsfunktionen auf Dialogerfolg			
H3.1 Effektivität der Interaktion	A	1+2	0 … 100 %
H3.2 Effizienz der Interaktion (Nutzereingaben)	A	1+2	0 … x Eingaben
H3.3 Effizienz der Interaktion (Dialogdauer)	A	1+2	0 … x Sekunden
WL3 Verständigungsprobleme und Lösungsstrategien			
WL3.1 Verständigungsprobleme begünstigende Faktoren	A+B	1+2	Siehe Kodierschema A
WL3.2 Nonverbale Lösungsstrategien	A+B	1+2	Siehe Kodierschema B
WL3.3 Frustrationsgrenzen			
WL3.3.1 Anzahl der Eingaben bis Dialogabbruch	B	2	0 … x
WL3.3.2 Zeitdauer bis Dialogabbruch	B	2	0 … x Sekunden

Falls nicht anders angegeben, wurden alle metrisch skalierten Messgrößen zunächst mittels Shapiro-Wilk-Test (SW-Test) auf Normalverteilung untersucht[95] und anschließend die Mittelwerte der jeweiligen Gruppen über einen gepaarten t-Test (Normalverteilung) bzw. einen Wilcoxon-Test (keine Normalverteilung) auf statistisch signifikante Unterschiede geprüft. Bei ungepaarten Mittelwertvergleichen wurde der Mann-Whitney-U-Test (MWU-Test) verwendet.

6.4.7 Ergebnisse

Die nachfolgenden Abschnitte beschreiben die ermittelten Ergebnisse aus den 480 dokumentierten Interaktionen mit 17.118 annotierten Turns im Detail.

Einfluss der Dialogunterstützungsfunktionen auf die Effektivität der Interaktion (H3.1)

Die Effizienz der Interaktionen wurde anhand des Anteils der erfolgreich übermittelten Intention bestimmt (Abb. 107). Bei aktivierter Kontextsensitivität (KS) konnten im Durchschnitt 92,5 % (SD= 4,37) der Dienstbeschreibungen bei aktivem Systemverhalten und 96,25 % (SD=3,02) bei passivem Systemverhalten erfolgreich übermittelt werden. Bei deaktivierter KS sank die mittlere Effektivität auf 73,75 % (SD=11,08) bei der Gruppe mit aktivem Systemverhalten und auf 85 % (SD=8,74) bei der Gruppe mit passivem Systemverhalten. Die KS des Systems hatte innerhalb beider Probandengruppen

[95] Signifikanzniveau: 0,05.

einen statistisch signifikanten Einfluss mit hoher Effektstärke[96] auf die Effektivität der Interaktion ($p(Gruppe_M, Gruppe_O) < 0{,}001; r = 0{,}88$).

Der Vergleich beider Durchläufe (mit/ohne KS) mittels MWU-Test zeigte keinen statistisch signifikanten Einfluss des Systemverhaltens (aktiv/passiv) auf die Effektivität der Interaktion ($p(KS_{mit}) = 0{,}393$ bzw. $p(KS_{ohne}) = 0{,}190$).

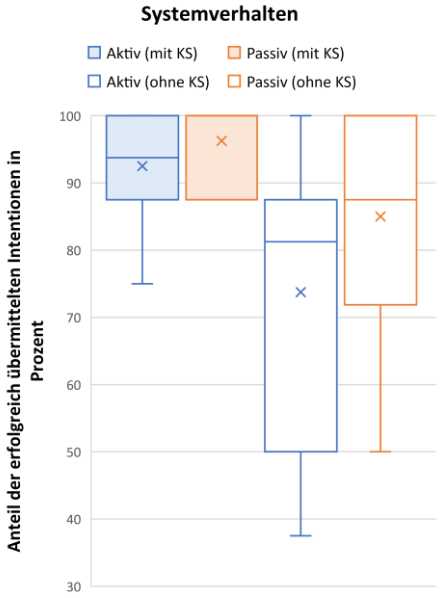


Abb. 107 Einfluss des Systemverhaltens auf die Effektivität der Interaktion (KS = Kontextsensitivität, Kreuze = Mittelwerte, Horizontale Striche = Median)

Einfluss der Dialogunterstützungsfunktionen auf die Effizienz der Interaktion (H3.2 und H3.3)

Die Effizienz der Interaktion wurde anhand der Gesamtanzahl der Nutzereingaben während der Übermittlung der Dienstbeschreibungen der Klasse A bestimmt. Die minimale Anzahl der Eingaben[97] lag hierbei bei 39. Betrachtet man nur diejenigen Probanden, die diese acht Dienstbeschreibungen erfolgreich übermitteln konnten, fällt auf, dass lediglich vier dieses Ziel bei deaktivierter KS erreichen konnten, wohingegen 12 Teilnehmer dies mit aktivierter KS schafften.

Tab. 32 Ermittelte Daten der Interaktionen, bei denen alle Dienstbeschreibungen der Klasse A erfolgreich übermittelt werden konnten

	Aktiv (mit KS)	Passiv (mit KS)	Aktiv (ohne KS)	Passiv (ohne KS)
Anzahl der Probanden	5	7	1	3
Anzahl der Eingaben				
Min	**49**	**49**	70	73
Max	70	75	70	186
Mittelwert	55,6	60,9	70	114,3
SD	8,4	9,4	-	62,3
Interaktionsdauer in Sekunden				
Min	203	**156**	392	255
Max	482	396	392	737
Mittelwert	289,4	273,9	392	428,7

[96] Effektstärke r nach Cohen [361, S. 156]: $r < 0{,}1\ (klein);\ 0{,}1 \leq r < 0{,}3\ (mittel); r > 0{,}5\ (stark)$
[97] Pro Dienstbeschreibung: Ansprache des UAVs + dienstbezogene Parameter (2-4) + Rückbestätigung.

SD	116,7	89,0	-	267,7

Im besten Fall benötigten Probanden 49 Eingaben (aktives SV mit KS), im schlechtesten Fall 186 Eingaben (passives SV ohne KS), um alle Kommunikationsziele zu erreichen. Der Einfluss der KS auf die Anzahl der Eingaben war hierbei stark und statistisch signifikant ($p(Gruppe_M) = 0{,}028; r = 0{,}90$ sowie $p(Gruppe_O) = 0{,}005; r = 0{,}87$), wohingegen der Einfluss des Systemverhaltens auf die Anzahl der Eingaben nicht signifikant war ($p(KS_{mit}) = 0{,}53$ bzw. $p(KS_{ohne}) = 0{,}50$).

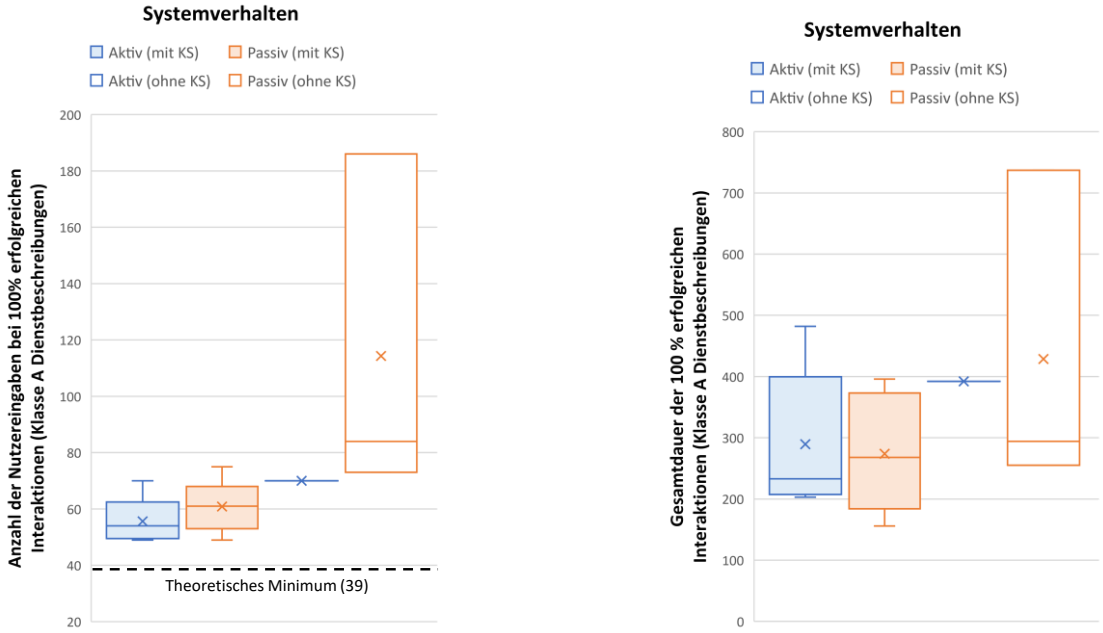


Abb. 108 Einfluss des Systemverhaltens auf die Anzahl der Nutzereingaben (links) und Gesamtdauer der Interaktionen in Sekunden (rechts) während der erfolgreichen Übermittlung aller Dienstbeschreibungen der Klasse A

Die kürzeste Übermittlung aller Dienstbeschreibungen erfolgte innerhalb von 156 Sekunden (passives SV mit KS), wohingegen die längste Übermittlung 737 Sekunden (passives SV ohne KS) dauerte (Abb. 108 rechts). Der Einfluss der KS war auch hierbei statistisch signifikant ($p(Gruppe_M) = 0{,}028; r = 0{,}90$ sowie $p(Gruppe_O) = 0{,}005; r = 0{,}89$), wohingegen der Einfluss des SV auf die Interaktionsdauer nicht nachgewiesen werden konnte ($p(KS_{mit}) = 1{,}0$ bzw. $p(KS_{ohne}) = 0{,}068$).

Verständigungsprobleme begünstigende Faktoren (WL3.1)

Vor allem im zweiten Durchlauf (ohne Kontextsensitivität des Systems) waren die Probanden vermehrt mit Kommunikationsproblemen konfrontiert (Abb. 109). Zu den drei häufigsten Ursachen zählten mehrdeutige Darstellungen (537), gefolgt von Ausführungsfehlern (291) der Probanden, bei denen die Darstellungen sichtbar von den Vorgaben abwichen (z.B. falsche Hand oder eine Hand statt zwei für die Signalisierung verwendet), sowie Systemfehler (126), bei denen das System einzelne Subsignale aufgriff und falsch interpretierte.

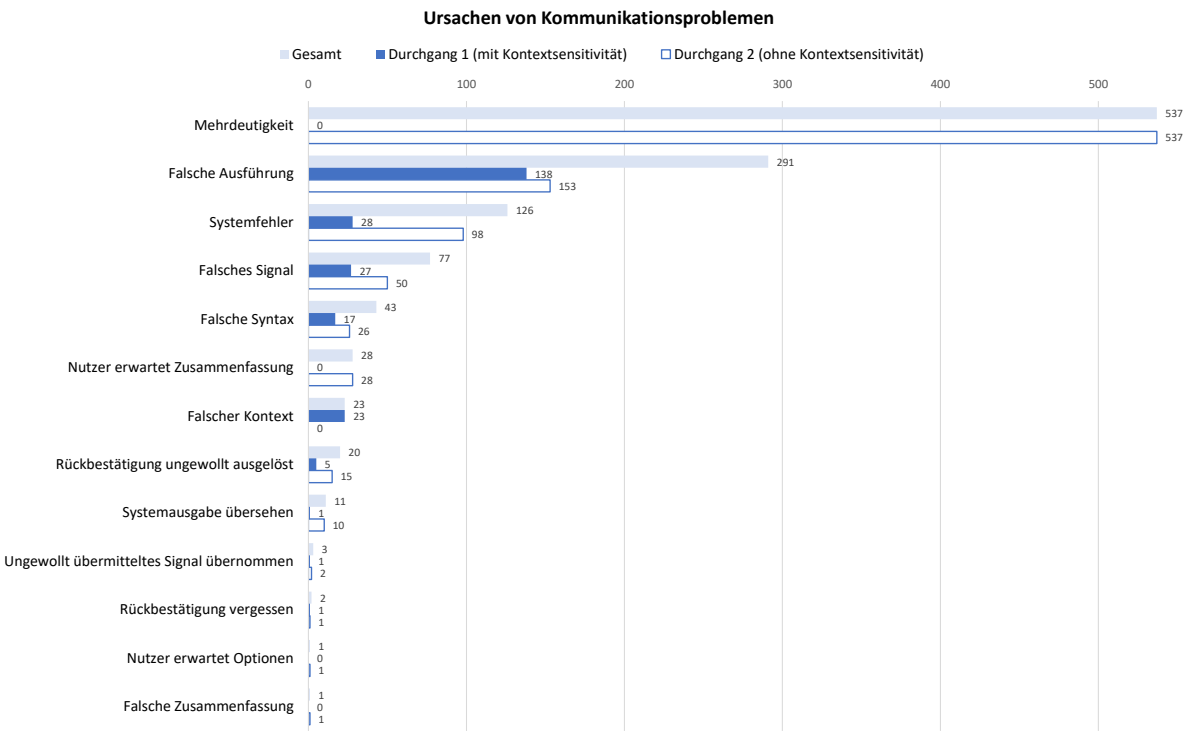


Abb. 109 Verteilung der Ursachen von Kommunikationsproblemen

Weitere Ursachen sind Signalverwechslungen (77), Veränderungen an der Syntax (43) und systemseitig nicht erstellte, aber vom Probanden erwartete Zusammenfassungen nach der Übermittlung (28). Insbesondere im ersten Durchlauf traten bei einigen Interaktionen (23) Diskrepanzen beim angenommenen (Proband) und dem vom System geschätzten Gesprächskontext auf. Unabsichtlich signalisierte Rückmeldungen (20) trugen darüber hinaus zu Kommunikationsproblemen vorrangig im zweiten Durchlauf bei (15), ebenso wie übersehene Systemrückmeldungen (11). Vergessene Rückbestätigungen (2) und erwartete, aber nicht angezeigte Optionen (1) sowie falsche Zusammenfassungen des Systems (1) resultierten in nur sehr wenigen Fällen in Missverständnissen.

Nonverbale Lösungsstrategien (WL3.2)

Die Reaktionen der Probanden auf die insgesamt 1.163 beobachteten Kommunikationsprobleme waren vielfältig (Abb. 110). Bei Erkennungsproblemen wurde die Darstellung zumeist auf identisches Weise (615) oder mit veränderter Position der Arme/Hände (193), einer modifizierten Handform (80), mit langsamer (48), schnellerer (14), kleinerer (20), größerer (43) oder überdeutlicher und klopfender Darstellung (N=34) wiederholt. Sofern verfügbar, wurde auch eine alternative Darstellung (134) gewählt, um das gewünschte Signal zu übermitteln. Nach mehreren erfolglosen Übermittlungsversuchen wurde manchmal auch direkt zum nächsten Signal übergegangen (30). In nur zwei Fällen konnte eine bewusste Veränderung der Körperhaltung zur Verbesserung der Kommunikation beobachtet werden.

Abb. 110 Verteilung der Reaktionen der Probanden nach Auftreten eines Kommunikationsproblems

Die Dialogsteuerungssignale zum Zurücksetzen der Dialoghistorie über das `RESET`-Signal (385) sowie die Revision der letzten Nutzereingabe über das `IGNORE`-Signal (270) gehörten zu den zweit- und dritthäufigsten Reaktionen. Die Anweisung an das System, die letzte Ausgabe zu wiederholen (`SAY–AGAIN`-Signal) wurde hingegen seltener genutzt (22). In 74 Fällen reagierten Probanden auf eine falsche oder unklare Systemantwort mit einer sichtbaren Ablehnung, bei 29 Fällen wurde eine uneindeutige Systemmeldung (z.B. `CONFIRM YES?`) hingegen abgenickt.

Einige Probanden führten während der Interaktionen Testläufe mit anderen Signalen (112) oder Dienstbeschreibungen (8) durch, um deren Funktionsfähigkeit zu überprüfen und die Ursache des Kommunikationsproblems einzugrenzen. In einigen wenigen Fällen (5) forderten Probanden aktiv Hilfe beim System an. Bei 56 Interaktionen brachen Probanden den Dialog vorzeitig ab.

Frustrationsgrenzen (WL3.3.1 und WL3.3.2)

Die Übermittlung der Dienstbeschreibungen der Klasse B war aufgrund der deaktivierten Kontextsensitivität im zweiten Durchgang prinzipbedingt nicht möglich. Nach durchschnittlich 29,5 Eingaben (SD=18,2) brachen Probanden in diesen Fällen die Interaktion ab. Das Konfidenzintervall (95 %) lag hierbei im Bereich zwischen 25,5 und 33,5 Eingaben. Abb. 111 zeigt die Verteilung der Nutzereingaben sortiert nach dem Median.

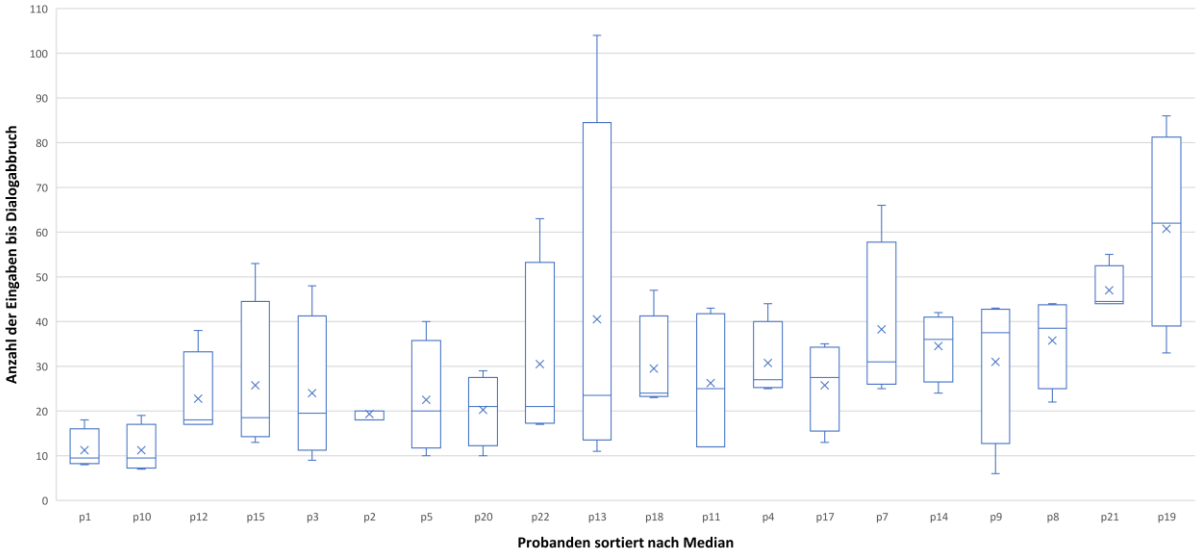


Abb. 111 Verteilung der Nutzereingaben pro Dienstbeschreibung bis zum Dialogabbruch

Bezogen auf die Interaktionsdauer wurden die Dialoge im Durchschnitt nach 161,4 Sekunden (SD=103,2) abgebrochen. Das Konfidenzintervall (95 %) lag hierbei im Bereich zwischen 138,7 und 184,2 Sekunden. Abb. 112 zeigt die Verteilung der Interaktionsdauern im Detail.

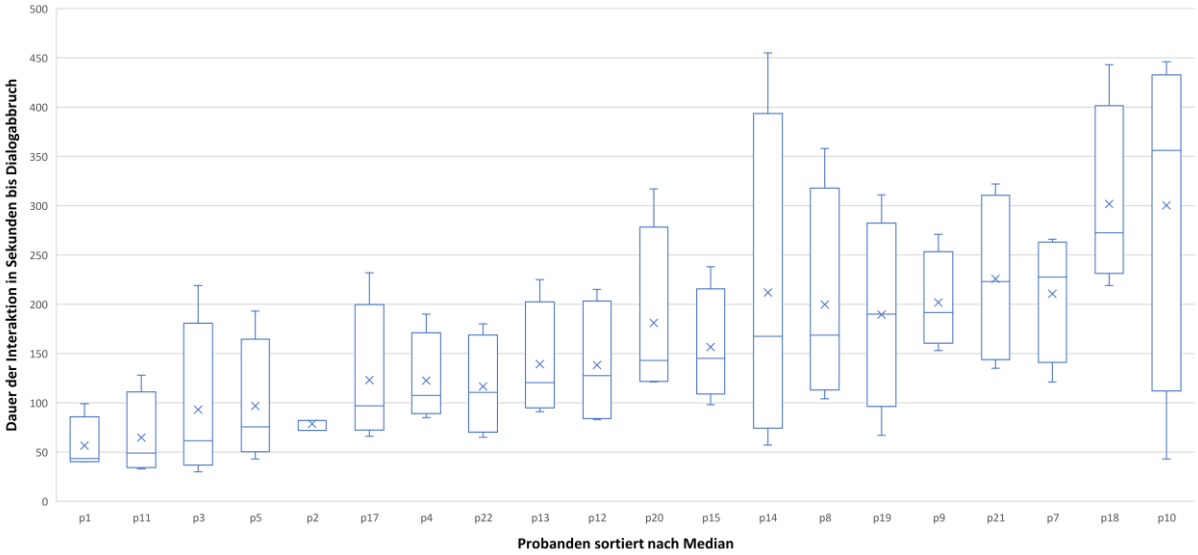


Abb. 112 Verteilung der Interaktionsdauer pro Dienstbeschreibung in Sekunden bis zum Dialogabbruch

Fehlerbetrachtung

Der Auswerteprozess wurde so weit wie möglich automatisiert, um Fehler bei der Datenverarbeitung zu minimieren. Es kann jedoch nicht ausgeschlossen werden, dass bei der manuellen Annotation der Beobachtungen Fehler eingebracht wurden. Um potentielle Fehlerquellen zu identifizieren, wurden Gegenmaßnahmen getroffen (Plausibilitätsüberprüfungen, Ausreißeranalyse).

6.4.8 Diskussion

Eine nonverbale Form der Interaktion zwischen Mensch und Maschine birgt Potential für die vielfältigsten Verständigungsprobleme. Während des Experiments konnten drei potentielle Fehlerquellen identifiziert werden, die auf das Systemverhalten, das Vokabular oder den Nutzer zurückgeführt werden können.

Systeminduzierte Kommunikationsprobleme

Trotz intensiver Vorabtests kamen während der Interaktionen einige Implementierungsfehler des Experimentalsystems zum Vorschein, die die Interaktion negativ beeinflussten. So zeigte das Experimentalsystem bei sehr vielen aufeinanderfolgenden, falschen Eingaben ohne eine zwischenzeitliche `RESET`-Anweisung keine Zusammenfassungen mehr an, auch wenn die Syntax der letzten Eingaben korrekt war. Auch die implementierte Eingabesperre von zwei Sekunden nach der Detektion eines Signals war zu gering gewählt, weil einige Probanden (P19) länger als notwendig in einer statischen Darstellung (Pose) eines Signals verharrten und somit Mehrfachdetektionen desselben Signals beim System auslösten. Zudem konnten Nutzer während einer Zusammenfassung das zuletzt übermittelte Signal nicht revidieren (`IGNORE`), sondern erhielten stattdessen eine Meldung (`NOTHING TO IGNORE`, die – nachdem sie verschwunden war – den Nutzer im Unklaren über den aktuellen Systemstatus ließ. Die einzige Möglichkeit, die Zusammenfassung zurückzuholen, war das `SAY-AGAIN`-Signal oder ein komplettes Zurücksetzen der Dialoghistorie mittels `RESET`.

Während der Interaktionen erfuhren Probanden entweder ein passives (Gruppe O) oder ein aktives (Gruppe M) Dialogverhalten des Systems. Der Nutzen des aktiven Dialogverhaltens (Gruppe M) konnte nicht eindeutig nachgewiesen werden. Einigen Probanden half die zusätzliche Anzeige der vom System erwarteten Eingabeoptionen bei der Dienstbeschreibung. So schaute beispielsweise P20 während der Übermittlung des Signals `BOX` zur Seite und nahm an, dass dieses vom System erfasst wurde, was jedoch nicht der Fall war. Durch die explizite Nachfrage des Systems nach der fehlenden Information im weiteren Verlauf des Dialogs, konnte der nutzerseitige Fehler abgefangen und das Kommunikationsziel dennoch erreicht werden. Wiederrum andere Probanden der Gruppe M deuteten das gelegentliche Fehlen der Optionsanzeige im zweiten Durchgang korrekterweise als einen Systemfehler und leiteten vorsichtshalber einen Neustart des Dialogs ein (P06: „Jetzt frag mich doch, was du haben willst. Ok, wir müssen resetten."). Dieses Verhalten war aber nicht durchgängig zu beobachten, da einige Probanden (P19, P20) die Systemhinweise zu ignorieren schienen und – unabhängig von den vorgeschlagenen Optionen – ihre Dienstbeschreibung formulierten. Probanden der Gruppe O bekamen keine Optionsvorschläge angezeigt und mussten daher öfter den Dialog neustarten, um den tatsächlichen Systemzustand mit dem vermuteten abzugleichen (P20: „Ich seh' jetzt halt nicht, was er grad macht.", P09: „Wer weiß, wo ich hier stecke – RESET"). Insgesamt liegt aber die Vermutung nahe, dass die permanente Anzeige der Optionen im Allgemeinen eher dazu beiträgt, den Nutzer zu verwirren, anstatt ihm zu helfen. Ein adaptives Dialogverhalten, welches sich an die jeweilige Gesprächssituation anpasst (stumm bis proaktiv), kann zweckdienlicher sein und die Vorteile beider Verfahren vereinen, um den Nutzer einerseits nicht unnötig mit Ausgaben zu verwirren und andererseits Hilfestellungen bei Interaktionsproblemen geben zu können. Dazu muss es jedoch in der Lage sein, Dialogkonflikte rechtzeitig zu erkennen und bei Bedarf proaktiv in Form eines Alternativvorschlages in den Dialog einzugreifen (vgl. Suchmaschinen bei fehlerhaften Eingaben: „Meinten Sie vielleicht das?").

Der Einfluss der Kontextsensitivität (KS) des Systems war hingegen nicht nur im Hinblick auf die Kommunikationseffektivität statistisch signifikant, sondern auch deutlich an den Kommentaren der Probanden im Debriefing erkennbar (P18: „Beim ersten Mal war es unglaublich akkurat und beim zweiten Mal hat er alles falsch erkannt und ist in den Kategorien gehüpft zum Teil.", P22: „Beim ersten Mal hat es recht gut funktioniert. Aber beim zweiten Mal (…) so viel missverstanden wie möglich."). Durch das Fehlen der KS hat das System zu jeder Zeit auf alle Bewegungen reagiert, die zu einer der Signalbeschreibungen im Vokabular passten, sodass auch die Syntaxtreue degradiert wurde (P17: „Wundert mich aber, dass der jetzt zwei Actions hintereinander erkennen kann.") und forderte eine höhere Ausführungstreue von den Probanden (P13: „Mehr Präzision bei der Ausführung gefragt."). Da diese jedoch nicht von jedem Probanden aufrechterhalten werden konnte, entstand eine Anhäufung

von Missverständnissen, die bei manchen Teilnehmern eine gewisse Frustration (P22: „Du willst schauen, wie lange Menschen mit so einem Ding umgehen können, wenn sie Probleme haben. Und das wird immer weniger.“, P09: „Ich habe keine Lust mehr, zu probieren.“), aber auch Euphorie (P05: „Yes, Endlich!“) bei erfolgreichen Übermittlungen im Laufe der Interaktion erkennen ließ.

Insbesondere im zweiten Durchgang (ohne KS) traten zudem vermehrt nutzerspezifische Klassifikationsprobleme in Form von Signalverwechslungen auf. So wurden mehrfach die Signale `TRANSPORT` als `RADIUS` (P13), `RADIUS` als `HELI` (P9) oder auch `RESET` als `HOME` (P10, P11, P29) vom System erkannt. Mögliche Ursachen dieses systemseitigen Fehlverhaltens können in der unterschiedlichen Ausführung, der Anatomie der Probanden, aber auch im Fehlen entsprechender Beispieldaten im Trainingsdatensatz der ML-gestützten Klassifizierungsmodule (Armstellung, Bewegungspfad, Handform) vermutet werden. Eine weitere Fehlerquelle des Experimentalsystems bildeten die minimalistisch formulierten textuellen Systemausgaben, die manchmal mehrdeutige Interpretationen zuließen (z.B. bei P20: „`WHAT/RADIUS?`“)

Vokabularinduzierte Kommunikationsprobleme

Sowohl die Wahl der Darstellungsform der nonverbalen Signale (Wie soll das Signal vom Nutzer dargestellt werden?) auf der Produzentenseite als auch die Signalbeschreibung auf der Signalerfassungsseite (Wie können die relevanten Merkmale des Signals erkannt und beschrieben werden?) bieten Potential für vielerlei Arten von Kommunikationsproblemen.

Einen maßgeblichen Beitrag zu kommunikativen Missverständnissen liefern Signale, deren Darstellungen sich zu sehr visuell und motorisch ähneln und die sich nur in einem visuellen Merkmal unterscheiden. Die Problematik war beispielsweise bei der Übermittlung von Richtungshinweisen (relevante Signalmerkmale: Armstellung und Handbewegung) im zweiten Durchgang ohne KS deutlich erkennbar. So war bei einigen Probanden beim seitlichen Ausstrecken des Armes der Daumen sichtbar und führte fälschlicherweise zur Erkennung eines Rückmeldungssignals `YES` (relevante Signalmerkmale: Armstellung, Handbewegung und Handform) anstatt der jeweiligen Richtung. Zu wenige Unterscheidungsmerkmale erschweren nicht nur die signaltechnische Klassifikation, sondern fördern auch Darstellungsverwechslungen beim Nutzer. Derlei Beispiele finden sich auch im verwendeten Vokabular:

- `HOME` (Hände bilden ein Dach über dem Kopf) und `RESET` (Hände überkreuzt über dem Kopf)
- `HUMAN` (Arme nebeneinander auf der Brust) und `IGNORE` (Arme überkreuzt auf der Brust)
- `HELI` (viele kleine, schnelle Rotationsbewegungen) und `RADIUS` (eine große, langsame Rotation)

Die Verwechslungsgefahr konnte mit dem verwendeten kontextsensitiven Signalübersetzungsverfahren zwar verringert werden, es ist jedoch nicht auszuschließen, dass bei Systemkonfigurationen mit weitaus mehr Kommunikationsinhalten zwei konkurrierende Kontexte im Laufe eines Dialogs gleichzeitig aktiv werden und so die Mehrdeutigkeit vom System nicht mehr aufgelöst werden kann. Daher müssen derlei Interferenzen bereits in der Designphase des Vokabulars vermieden werden. Hierfür werden jedoch geeignete Ähnlichkeitsmaße benötigt, welche das erfassungstechnische, motorische als auch kognitive Verwechslungspotential hinreichend genau beschreiben können.

Die nonverbale Kommunikationsfähigkeit der Probanden war unterschiedlich stark ausgeprägt, zudem traten im Laufe der Interaktion ermüdungsbedingte Schwankungen in der Darstellungsqualität auf (P16). Einige Probanden hatten zudem motorisch bedingte Schwierigkeiten, bestimmte Signale, die

ungewöhnliche Arm- oder Fingerstellungen beinhalteten (z.B. `CALL`-Signal mit gleichzeitig abgespreiztem Daumen und kleinen Finger), vorgabengetreu darzustellen. In Verbindung mit der Mehrdeutigkeit einzelner Signale, die ohne KS zu „Sackgassen“ im Dialogverlauf führten, war aus den Kommentaren der Teilnehmer ein klarer Wunsch nach alternativen Darstellungen erkennbar (P04: „Gibt leider nur ein Zeichen für TRANSPORT“, P17: „Für BATTERY kann ich ja keine Alternative angeben und FLY ist auch nur der eine Befehl.", P11: "Ich frage mich gerade, ob es eine Alternative für Höhe gibt, die gibt's aber nicht.“). Daher sollte beim Signaldesign ein weiteres Augenmerk auf ein redundantes Vokabular gelegt werden, bei dem für jedes Signal mindestens eine alternative Darstellungsart (z.B. mit und ohne Bewegungskomponente) verfügbar ist, um Handlungsoptionen bei nutzerspezifischen Darstellungsdefiziten oder systembedingten Erkennungsproblemen bieten zu können.

Nutzerinduzierte Kommunikationsprobleme

Als aktiver Teilnehmer der Interaktion kann natürlich auch der Mensch durch sein Verhalten Fehler in die Interaktion einbringen und zu Verständigungsproblemen beitragen. Basierend auf den Arbeiten von Rasmussen [362] und Reason [363] werden menschliche Fehler in komplexen Mensch-Maschine-Systemen allgemein als „nichtintentionale Handlungen mit unerwünschten Konsequenzen“ [364, S. 3] verstanden und können den zwei grundsätzlichen Fehlertypen „Ausführungsfehler“ (*execution failures*) und „Planungsfehler“ (*planning failures*) zugeordnet werden.

Ausführungsfehler treten vorrangig bei bereits bekannten Interaktionssituationen auf, bei denen Nutzer zwar die richtige Intention zur Lösung einer Situation haben, die entsprechende Handlung aber fehlerhaft ausführen (sog. Ausrutscher oder engl. *slips*). Solche Flüchtigkeits- oder Aufmerksamkeitsfehler konnten auch bei der nonverbalen Mensch-UAV-Interaktion beobachtet werden. Trotz intensiver Trainingsphase und des immer einsehbaren Vokabulars auf dem Großbildschirm, entstand ein nicht unerheblicher Anteil der Kommunikationsprobleme durch fehlerhafte Signaldarstellungen der Probanden. Dabei wurde während der Signalproduktion beispielsweise der Blick vom System abgewendet (P20), die Form der Hand nicht beachtet (Daumen ausgestreckt, P05, P13, P17), die passive Hand vernachlässigt (hinter dem Körper oder auf Bauchhöhe positioniert, P02, P06) oder die aktive Hand falsch im Raum positioniert (P14). Auch wurden Signale und deren Darstellungen an sich verwechselt (z.B. `FIND`, `SCAN` und `SHOW` bei P17, P18). Daneben wurden auch Systemausgaben übersehen (P17), Zahlen und Richtungen (P5) oder Zusammenfassungen vorschnell rückbestätigt, ohne den Inhalt mit der Aufgabenstellung abzugleichen (P1, P2, P12, P14).

Einen weiteren Ausführungsfehler stellen Versäumnisse (engl. *lapses*) dar, bei denen einzelne, eigentlich bekannte, Handlungsschritte ausgelassen oder vergessen werden. Derlei Fehler zeigten sich, wenn Probanden beispielsweise die Ansprache des UAVs zu Beginn einer Dienstbeschreibung (`YOU`) oder andere Teile der Syntax vergaßen (P03, P07). Die Neuformulierung einer Dienstbeschreibung ohne vorherige Übermittlung des `RESET`-Signals führte ebenfalls zu Verständigungsproblemen (P04). Ein Teil der Probanden fokussierte sich derart auf die korrekte Übermittlung eines fehlerhaften Signals, dass sie die eigentliche Aufgabenstellung während der Interaktion wieder vergaßen (P06: „YOU … ASK … Ich weiß schon selbst nicht mehr, was ich von ihm will.“, P21 nach mehreren, erfolglosen Versuchen, `SECRET` zu übermitteln: „Was muss ich eigentlich machen?“, P22 nach etwa einer Minute erfolglosem Dialog: „YOU ASK … was wollte ich fragen?“).

Planungs- oder Entscheidungsfehler (engl. *mistakes*) stellen eine weitere Kategorie von menschlichen Fehlern dar und sind vor allem dann anzutreffen, wenn Nutzer sich mit neuen, bisher unbekannten Situationen konfrontiert sehen und auf bekannte Verhaltensregeln zur Lösung einer Situation zurück-

greifen oder improvisieren müssen. Für die Mehrzahl der Teilnehmer verlief der erste Durchgang relativ problemfrei und stellte den Idealfall der Interaktion dar. Daher vermutete auch die Mehrheit der Probanden den Fehler im zweiten Durchlauf zunächst bei sich selbst, was sich auch in der hohen Anzahl an identischen Wiederholungen widerspiegelt (P13: „Mach‘ ich was an der Syntax falsch oder stimmt die Gestik nicht?“, P17: „Was hat er denn jetzt? Oder mach‘ ich was falsch?“). Erst nachdem sich die Verständigungsprobleme – welche während des ersten Durchlaufs in der Form nicht auftraten – häuften, wurden einige Teilnehmer skeptisch und hinterfragten die Systemkonfiguration (P03: „Das ist doch manipuliert.“). Die Reaktionen auf Verständigungsprobleme waren dabei vielfältig, führten jedoch nicht immer zum Erfolg.

Verhaltensmuster und Lösungsstrategien

Abhängig davon, wo Probanden die Problemursache vermuteten (bei sich selbst oder dem System), konnten mehrere, teils aufeinander aufbauende Verhaltensmuster beobachtet werden, bei denen Modifikationen an der Darstellung, der Syntax oder dem Dienstbeschreibungsverfahren vorgenommen wurden.

Nachdem die ersten Verständigungsprobleme auftauchten, vermuteten die meisten Probanden einen Fehler in der eigenen Darstellung und wiederholten daher zunächst das letzte Signal mehrmals. Nach einer Weile zeigten einige Probanden jedoch eine Verhaltensänderung. Während ein Teil stets dasselbe Signal immer wieder aufs Neue wiederholte[98] – teilweise unterbrochen durch Dialogsteuerungssignale wie `RESET` – fing der andere Teil an, einzelne Darstellungsparameter (Körperhaltung, Handform, Armstellung) zu verändern, um die Signalgüte zu optimieren. Bei einigen führten diese Änderungen zum Erfolg, bei anderen zu mechanisch anmutenden Darstellungen (P21) und bei manchen zu teils irrationalem Verhalten. P22 drückte beispielsweise seine Hand fester an die Brust (optisch nicht feststellbar) oder übermittelte das `RESET`-Signal bewusst mehrmals hintereinander (einmal hätte gereicht), um sicher zu gehen, dass keine „Signalreste“ im Dialogspeicher waren. P05 fing an, mit beiden Händen auf die Brust zu klopfen, nachdem die Übermittlung des Signals `HUMAN` mehrmals nach mehreren Versuchen erfolglos blieb. Wiederrum andere (P04, P09, P14, P16) wurden in eine Art „Spirale der Verschlimmbesserung“ hineingezogen, in der die vermeintliche Optimierung der Darstellung (extreme Verlangsamung oder übergroße Artikulation) in einer Verschlechterung der Erkennungsleistung des Systems (Fehldetektion bis gar keine Reaktion) resultierte und die Übermittlung des Signals praktisch unmöglich machte (P04: „Ich muss irgendwie die linke Hand verstecken, sonst merkt er es nicht.“). So führte P14 das `FOLLOW`-Signal mit jeder nicht erkannten Wiederholung zunehmend langsamer aus und konnte es letztendlich nicht übermitteln, da für eine zuverlässige Erkennung eine zügige Horizontalbewegung beider Hände auf Brusthöhe benötigt wurde.

Einige Teilnehmer kamen nach mehreren gescheiterten Kommunikationsversuchen zu dem Schluss, dass der Fehler beim System liegen muss und versuchten, die Problemursache durch systematische Tests näher einzugrenzen. Dazu testeten einige Probanden entweder andere Signale (P07, P09), um eventuelle Signalvertauschungen aufzudecken, oder andere Dienste (P09), um deren Funktionsfähigkeit zu überprüfen. Wiederrum andere nahmen Modifikationen an der Syntax vor (P15).

Ein weitere Lösungsstrategie stellte das Umschreiben des fehlerhaften Dienstes oder Signals durch andere ähnliche Dienste bzw. Signale dar. So versuchte P17 eine alternative Dienstbeschreibung (`TRANSPORT`) zu übermitteln, in der Hoffnung dasselbe Ziel (Verlegung des UAVs) zu erreichen. Teil-

[98] Die höchste, beobachtete Anzahl an identischen Wiederholungen einer Signaldarstellung innerhalb einer Dienstbeschreibung lag bei 62 (P19).

weise wurde auch versucht, das Erkennungsproblem des Systems durch sinnbildliche, visuelle Metaphern zu lösen. P16 ahmte dazu ein langsames Anheben zweier schwerer Tüten nach, um das nicht erkannte Signal `LIFT` zu umschreiben. Als letzte Maßnahmen probierten einige Probanden (P06, P09, P18) durch die Übermittlung des `CALL`-Signals, Hilfe beim System anzufordern. Abb. 113 zeigt eine Übersicht der beobachteten Problemlösungsstrategien bei nonverbalen Verständigungsproblemen und den entsprechenden Folgen, wenn systemseitig nicht rechtzeitig interveniert wird.

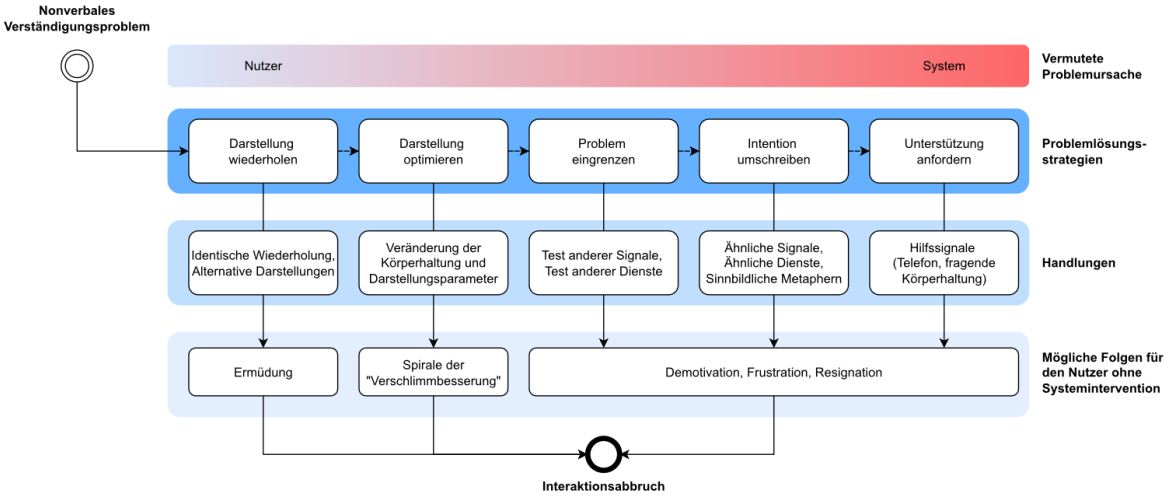


Abb. 113 Beobachtete Problemlösungsstrategien bei nonverbalen Verständigungsproblemen

Nonverbale Disziplin und Awareness über die ausgestrahlten Körpersignale

Das nonverbale Verhalten des Nutzers trägt maßgeblich zur Effektivität und Effizienz einer visuellen Interaktion bei. Während der Interaktion konnte bei einigen Probanden ein fehlendes Bewusstsein für die eigenen, ausgestrahlten Subsignale des Körpers (z.B. Nick-/Schüttelbewegungen des Kopfes während spontaner Blickbewegungen, Handformveränderungen) beobachtet werden, sodass es ihnen schwerfiel, über die Dauer der Interaktion eine „nonverbale Disziplin“ aufrecht zu erhalten. So freute sich beispielsweise P13 über ein korrekt erkanntes Signal, nickte jedoch dabei unbewusst mit dem Kopf und übermittelte damit ungewollt ein weiteres Signal (`YES`). P21 kratzte sich mit der rechten Hand am Bauch in der Nähe des anderen Handgelenks, schaute dabei in Richtung des Systems und stellte fest: „Jetzt sagt er irgendwas mit `TIME`, obwohl ich mich nicht bewegt hab'.“. Nur wenige Probanden waren sich der eigenen Darstellungsfehler bewusst (P09: „Ach ne, der gesamte Arm (...) Ich bin hier schon so in Erwartung, dass irgendwas nicht geht, dass ich gar nicht meine eigenen Fehler bemerke.“). Intensives Training und mehr Erfahrung im Umgang mit einer visuellen Benutzerschnittstelle können sicherlich dazu beitragen, das eigene nonverbale Verhalten zu schulen und dadurch die Effizienz der Interaktion zu verbessern. Systemseitig muss aber auch eine weitere Dimension in die Signalverarbeitung einfließen, nämlich die Gewichtung und Relevanz von konkurrierenden Subsignalen im situativen Kontext des Dialogs. Wenn der Nutzer (womöglich unbewusst) mit dem Kopf schüttelt, aber eine Daumen-Hoch-Geste macht, welches Signal soll das System dann verarbeiten?

7 Zusammenfassung und Ausblick

Die sichere und nahtlose Integration unbemannter Flugsysteme in bestehende menschenzentrierte Abläufe und Verfahren stellt die Luftfahrt vor viele Herausforderungen. Während in der zwischenmenschlichen Interaktion die nonverbale Verständigung per Handzeichen ein bewährtes Alternativverfahren bei technisch- oder umgebungsbedingten Kommunikationsproblemen bildet, stellt diese Form der Interaktion mit unbemannten Fluggeräten (UAV) eine relativ neue Disziplin dar. Trotz modernster Sensorik und zunehmender Automatisierung findet die Interaktion mit UAVs in Sichtweite primär über eine Funkverbindung statt, die nur über ein entsprechendes technisches Kommunikationsgerät hergestellt werden kann. Mit dem Verlust dieser Verbindung geht auch immer ein Verlust der Kontrolle über das fliegende System einher. Im Bereich der *collocated Human-Drone Interaction* (cHDI) wurden verschiedenste Anstrengungen unternommen, um eine natürliche, direkte Mensch-UAV-Interaktion ohne den Einsatz zusätzlicher Kommunikationsgeräte zu ermöglichen. Die nonverbale Interaktion über eine visuelle Schnittstelle bietet dabei im Vergleich zu anderen Modalitäten die meisten Vorteile. Dennoch wird das Potential dieser Schnittstelle nicht vollständig ausgeschöpft, sodass in bisherigen Arbeiten primär der *Interaktionsaspekt* und nicht der *Kommunikationsaspekt* behandelt wurde. Das Ziel dieser Arbeit war es somit, einen Beitrag zum Ausbau dieser kommunikativen Fähigkeiten zu leisten.

7.1 Beiträge dieser Arbeit zum Stand der Forschung

Die Kontributionen dieser Arbeit können zwei Bereichen zugeordnet werden. Im ingenieurstechnischen Bereich wurden mehrere Verfahren entwickelt, die die perzeptiven und kommunikativen Fähigkeiten von unbemannten Flugsystemen über eine visuelle Benutzerschnittstelle erweitern können:

- Entwurf und Umsetzung eines visuellen Kommunikationssystems zur direkten, nonverbalen Mensch-UAV-Interaktion und dessen Funktionsnachweis in unterschiedlichen Nutzungskontexten
- Modulares Verfahren zur Erzeugung und Klassifizierung unterschiedlicher nonverbaler Darstellungsformen ohne die Notwendigkeit eines erneuten Trainings des Klassifizierungssystems
- Kontextsensitives Signaldekodierungsverfahren zur Übersetzung mehrdeutiger Körpersignale mittels eines regelbasierten Inferenzalgorithmus
- Transfer des Intent-Matching-Verfahrens aus dem Bereich der natürlichen Sprachverarbeitung auf den Bildverarbeitungsbereich zur Erkennung von Nutzerintentionen

Darüber hinaus konnten im wissenschaftlichen Bereich neue Erkenntnisse im Bereich der nonverbalen Mensch-UAV-Interaktion gewonnen werden durch die:

- Untersuchung relevanter nonverbaler Kommunikationsmittel von Nutzern ohne Systemerfahrung während der Interaktion mit einem technischen System im Kontext einer simulierten Notlage
- Untersuchung des Raumnutzungs- und Interaktionsverhaltens von geschulten Nutzern während der kollokierten Führung eines unbemannten Flugsystems aus der 15 kg Klasse über eine visuelle Schnittstelle

- Untersuchung des nonverbalen Problemlösungsverhaltens von intensiv geschulten Nutzern bei Verständigungsproblemen während der visuellen Übermittlung von Auftragsbeschreibungen an ein technisches Gerät

Im Anhand A13 befindet sich eine chronologische Auflistung der Veröffentlichungen des Autors, welche in den jeweiligen Detailkapiteln referenziert sind.

7.2 Bewertung der entwickelten Verfahren zur Adressierung der Fähigkeitslücken und Überprüfung der aufgestellten Hypothesen

In dieser Arbeit wurden drei identifizierte Fähigkeitslücken adressiert, die die Erkennung von Nutzerintentionen auf der Grundlage von nonverbalen Eingaben, die Erkennung generischer Körpersignale als auch die Verbesserung der Kommunikationsleistung in einem nonverbalen Dialog durch geeignete Dialogunterstützungsfunktionen betreffen.

7.2.1 Fähigkeitslücke 1: Visuelles Kommunikationssystem zur Intentionserkennung

Im Bereich der cHDI sind visuelle Schnittstellen üblicherweise an klassische Gestenerkennungssysteme gekoppelt, mit denen singuläre Nutzereingaben erfasst und auf automatisierte Systemfunktionen abgebildet werden (z.B. Positionsveränderung, Start, Landung, Aufnahme). Für einen darüber hinaus gehenden wechselseitigen Informationsaustausch im Sinne einer Mensch-UAV-Kommunikation werden hingegen erweiterte Funktionalitäten benötigt, um aus einer ganzen Sequenz von Eingaben die Intention des Nutzers ableiten zu können. Um diese Fähigkeitslücke zu adressieren wurde ein interdisziplinäres Konzept für ein bidirektionales, visuelles Mensch-UAV-Kommunikationssystem (VisCom) auf der Grundlage eines *„UAV-as-a-Service"*-Ansatzes entwickelt und die vom UAV angebotenen Dienste mit Methoden aus der natürlichen Sprachverarbeitung als Nutzerintentionen modelliert. Für die nonverbale Beschreibung dieser Nutzerintentionen wurde eine symbolbasierte Mensch-UAV-Verkehrssprache entwickelt, die die systematische Kodierung von Informationen in nonverbale Körpersignale erlaubt. Um diese Informationen über eine bildgebende Sensorik verarbeiten zu können, wurde ein *Motion-To-Symbol*-Ansatz vorgeschlagen, der die erfassten Körpermerkmale auf relevante Subsignale untersucht und diese klassifizierten Beobachtungen über einen kontextsensitiven Inferenzalgorithmus in eindeutige lexikalische Einheiten auflöst, die von einem aufgabenbasiertes Dialogsystem interpretiert werden.

Hypothese 1:

Das entwickelte und implementiere Systemkonzept ermöglicht die Untersuchung verschiedener Nutzungssituationen und Kommunikationsabsichten.

Das auf einem Baukastenprinzip basierende Signalkodierungs- und Dekodierungskonzept (Kapitel 5.2.11) ermöglichte nicht nur eine systematische Erzeugung des für die jeweilige Nutzungssituation notwendige Vokabular, sondern auch eine Anpassung der Signalgüte an das nonverbale Ausdrucksvermögen unterschiedlicher Nutzergruppen. Darüber hinaus konnten mittels des NLP basierten Intentionsbestimmungsverfahrens (Kapitel 5.3.2) unterschiedliche Kommunikationsabsichten durch entsprechende Konfigurationsdateien abgebildet werden. Für die Durchführung der Experimente A bis C wurde stets dasselbe Experimentalsystem in unterschiedlichen Konfigurationen genutzt (vgl. Interview Mode, Control Mode, Briefing Mode in Kapitel 5.5). Insofern kann die Hypothese 1 bestätigt werden.

7.2.2 Fähigkeitslücke 2: Erkennung generischer Nutzereingaben

In der neueren Literatur wird für die Klassifizierung nonverbaler Äußerungen vermehrt auf maschinelle Lernverfahren gesetzt, bei denen alle erfassten Körpermerkmale gleichzeitig und gleichwertig in die Klassifikation einer Darstellung einfließen. Im Unterschied zu diesem holistischen Ansatz wurde bei dieser Arbeit die Klassifizierung auf die Merkmalebene verlagert und auf Grundlage der Phonologie der Deutschen Gebärdensprache geeignete Merkmalbeschreiber entwickelt (u.a. zur Klassifizierung der Kopfbewegung, Stellung der Arme, Bewegungspfad der Hände, Form der Hände). Dadurch konnte eine Vielzahl von generischen und spezifischen Nutzereingaben modelliert werden, ohne das gesamte Klassifikationssystem neu trainieren zu müssen. Als generische Nutzereingaben wurde der Interaktionswunsch, eine zustimmende oder Ablehnende Rückmeldung als auch die Beschreibung eines körperlichen Problems definiert.

Für die Erkennung eines Interaktionswunsches wurde das visuelle Merkmal „Gestikulationsstärke" (Kapitel 5.2.5), d.h. die Magnitude der Armbewegungen, mit dem Mundbild (Kapitel 5.2.6) als Nutzereingabe abgebildet. Um zustimmende oder ablehnende Nutzerrückmeldungen erkennen zu können, wurden Nick- und Schüttelbewegungen des Kopfes (Kapitel 5.2.4) sowie bestimmte Handformen (vgl. Daumen hoch/Daumen runter in Kapitel 5.2.10) in das Vokabular aufgenommen. Um auch die Referenzierung von Körperteilen durch Berührungen während der Interaktion (z.B. zur Problembeschreibung) zu ermöglichen, wurde die Annäherung der Hände an andere Körperteile (Kapitel 5.2.9) ebenfalls als nonverbales Signal in das Vokabular aufgenommen.

Hypothese 2:

Die im Systemkonzept vorgesehenen Verfahren erlauben die Erfassung und Erkennung generischer Körpersignale, um auch eine rudimentäre Kommunikation mit Nutzern ohne Systemerfahrung zu ermöglichen.

Im Kontext einer simulierten alpinen Notsituation wurde die Eignung der umgesetzten Verfahren zur Erfassung generischer Körpersignale von Nutzern ohne Systemerfahrung überprüft. In den meisten Fällen konnten die für den simulierten Anwendungsfall relevanten Nutzereingaben detektiert und damit das Kommunikationsziel (Interaktionsaufnahme, Problembeschreibung, Absetzen eines Notrufs) erreicht werden. In einigen Fällen stieß jedoch der zweidimensionale Ansatz an seine Grenzen, weil insbesondere an das UAV gerichtete herbeiholende Bewegungen in der frontalen Ebene aufgrund der fehlenden Tiefeninformation sensorisch nicht erfasst und damit auch nicht als solche klassifiziert werden konnten. Auch die Erkennung von Zeigegesten zur Referenzierung problematischer Körperteile ohne deren Berührung konnte vom Systemkonzept nicht abgedeckt werden, weil die dreidimensionale Ausrichtung der Finger oder der Hand nicht betrachtet wurden. Daher kann die Hypothese 2 nur zum Teil bestätigt werden.

7.2.3 Fähigkeitslücke 3: Dialogunterstützungsfunktionen zur Verbesserung der nonverbalen Mensch-UAV-Kommunikation

Im Experimentalsystem wurden zwei Dialogunterstützungsfunktionen implementiert, die sowohl dem System als auch dem Nutzer während einer dialogähnlichen Kommunikation helfen sollten, das Kommunikationsziel effektiver zu erreichen. Eine dieser Funktionen stellt die *Kontextsensitivität* (KS) dar, die es dem System erlaubt, im jeweiligen Nutzungs- und Dialogkontext relevante von irrelevanten Körpersignalen zu unterscheiden (Kapitel 4.6). Die zweite Dialogunterstützungsfunktion bezieht sich auf das *Systemverhalten* (aktiv, passiv) während des Dialogs (Kapitel 5.3.1). Bei einem aktiven System-

verhalten reagiert das System auf jede Nutzereingabe, zeigt die möglichen Optionen an und ermöglicht dem Nutzer einen besseren Einblick in den gegenwärtigen Systemzustand. Bei einem passiven Verhalten reagiert das System erst, wenn es eine vollständige Dienstbeschreibung erkannt hat.

Hypothese 3:

Die im Systemkonzept vorgeschlagenen Dialogunterstützungsfunktionen helfen Nutzern während eines nonverbalen Dialogs mit einem technischen System bei der Erreichung des Kommunikationsziels.

Überraschenderweise konnten Probanden die meisten Kommunikationsziele erreichen, wenn die KS aktiviert und das Systemverhalten *passiv* war (96,25 %). Die Aktivierung beider Dialogunterstützungsfunktionen, d.h. KS und aktives Systemverhalten, führte zu einer geringeren Effektivität (92,5 %). Ohne KS lag die mittlere Effektivität bei 85 %, bei der Deaktivierung beider Dialogunterstützungsfunktionen sank sie sogar auf 73,75 %.

Die kontextsensitive Signalauswertung leistet somit einen wichtigen Beitrag zur Effektivitätssteigerung während der nonverbalen Kommunikation, da nutzerinduzierte (Darstellungsfehler) als auch systeminduzierte (Klassifizierungsfehler) Verständigungsprobleme besser abgefangen werden können. Die Optionsanzeige führte in Kombination mit der KS zwar zu einer Verringerung der Effektivität, im Vergleich zu gar keiner Unterstützung konnte sie jedoch die Effektivität verbessern. Insofern kann die Hypothese 3 bestätigt werden.

7.3 Beantwortung der identifizierten Wissenslücken

In dieser Arbeit wurden drei Wissenslücken identifiziert, die die nonverbalen Kommunikationssignale von Nutzern ohne Systemerfahrung, die kollokierte UAV-Führung über eine visuelle Schnittstelle sowie die Ursachen und das Problemlösungsverhalten von Nutzern bei nonverbalen Verständigungsproblemen betreffen.

7.3.1 Wissenslücke 1: Nonverbale Kommunikationssignale von Nutzern ohne Systemerfahrung im Kontext einer simulierten Notsituation

Bei Nutzern ohne jegliches Vorwissen über die Systemfunktionen konnte eine große Bandbreite an nonverbalen Kommunikationssignalen beobachtet werden. Für die Signalisierung eines Interaktionswunsches wurde in der Mehrzahl der Fälle erwartungsgemäß auf ein großräumiges einarmiges oder zweiarmiges Winken zurückgegriffen mit tatsächlicher (Rufen) oder angedeuteter Mundbeteiligung. Als visuelle Bestätigung auf eine vom System formulierte Frage setzte etwa ein Drittel der Probanden auf die Daumen-Hoch-Geste, während ein weiteres Drittel dieselbe Darstellung wählte wie für den Interaktionswunsch, jedoch ergänzt durch ein Kopfnicken oder eine verbale Bestätigung. Der Rest nutzte Kombinationen aus Kopfnicken, seitwärts gerichteten Winkbewegungen und der Daumen-Hoch-Geste. Lediglich ein Proband nutzte das offizielle Notsignal (Y-Pose) zur positiven Rückmeldung. Um auf ein medizinisches Problem hinzuweisen wurde auf das entsprechende Körperteil ein- oder zweihändig gezeigt, geklopft oder gestrichen. Teilweise wurde das problematische Körperteil auch umgriffen oder Mischformen für die Beschreibung gewählt.

Insgesamt waren die produzierten Signale nicht immer eindeutig und klar voneinander trennbar, mehrfach wurden dieselben Hand- und Armbewegungen für unterschiedliche Kommunikationsabsichten (Interaktionswunsch, Rückmeldung, Problembeschreibung) genutzt und unterschieden sich nur in den Subsignalen (Kopfbewegungen, verbale Rückmeldungen), was die Notwendigkeit eines multimodalen, kontextsensitiven Klassifizierungsansatzes verdeutlicht, um relevante Subsignale

(Kopfnicken) von irrelevanten (Winkbewegungen mit den Armen) im situativen Kontext unterscheiden zu können. Auch konnte ein Übergang von einer gemischt nonverbalen/verbalen Interaktionsform zu einer rein verbalen Interaktionsform beobachtet werden, sodass auch hier eine gewisse Notwendigkeit erwächst, diesen Kommunikationskanal in die Signalbewertung (z.B. durch maschinelle Verfahren zum Lippenlesen) mit einfließen zu lassen.

7.3.2 Wissenslücke 2: Kollokierte UAV-Führung über eine visuelle Schnittstelle

Im Nutzungskontext einer gestischen Mensch-UAV-Einweisung wurde in einem Realflugversuch mit einem hochautomatisierten Multikopter aus der 15 kg Klasse die kollokierte UAV-Führung über eine visuelle Schnittstelle untersucht.

Welches nonverbale Flugführungsverfahren eignet sich für die sichere Fernführung eines Multikoptersystems über eine visuelle Schnittstelle (WL 2.1)?

Im Experimentalsystem wurden zwei Flugführungsverfahren umgesetzt. Das erste *kontinuierliche Flugführungsverfahren* (KF-Verfahren) orientierte sich am Verfahren aus der Luftfahrt, bei dem das Luftfahrzeug die Flugbewegung so lange ausführt, wie die entsprechende Anweisung vom Einweiser produziert wird (d.h. die Probanden mussten *kontinuierlich* Eingaben produzieren). Sobald die Signalproduktion stoppt, unterbricht auch das Fluggerät seine Bewegung und geht in den Schwebeflug über. Beim zweiten *diskreten Flugführungsverfahren* (DF-Verfahren) wurde lediglich ein Start- und Stoppsignal vom Nutzer benötigt, dazwischen setzte das Fluggerät die Flugbewegung in die gezeigte Richtung automatisch fort (d.h. die Probanden mussten nur zu *diskreten* Zeitpunkten Eingaben produzieren).

Beide Flugführungsverfahren wurden sowohl mit subjektiven (Beanspruchung, Kontrollierbarkeit, Vorhersagbarkeit des UAV-Verhaltens) als auch objektiven Messmethoden (Anzahl Nutzereingaben, Flugdauer, Landegenauigkeit) miteinander verglichen. Bezüglich der Kontrollierbarkeit und Vorhersagbarkeit ergab sich kein eindeutiges Meinungsbild. Während ein Teil der Probanden die einfache Handhabung des KF-Verfahrens bevorzugte und die körperliche höhere Anstrengung kritisierte, gefiel einem anderen Teil die Vorhersagbarkeit und missfiel die wahrnehmbare Latenz des DF-Verfahrens. Die Auswertung des NASA TLX Fragebogens lieferte jedoch eine statistisch signifikant höhere subjektive Beanspruchung beim DF-Verfahren von ca. 24 % im Vergleich zum KF-Verfahren. Im objektiven Vergleich wurden beim DF-Verfahren erwartungsgemäß weniger Eingaben produziert, in Bezug zur Flugdauer und der Landegenauigkeit konnten jedoch keine signifikanten Unterschiede festgestellt werden.

Aufgrund der geringeren subjektiven Belastung und der Ähnlichkeit zu bestehenden Verfahren aus der Luftfahrt, erscheint ein kontinuierliches Führungsverfahren mit einer inkrementellen Flugbewegung des UAVs das geeignetere Werkzeug für die Positionierung eines Fluggerätes über einen visuelle Schnittstelle zu sein. Um die Positioniergenauigkeit weiter zu erhöhen, sollten die Bewegungsinkremente zudem variabel sein und beispielsweise an die Ausführungsgeschwindigkeit der Flugführungskommandos gekoppelt werden. Insgesamt konnte ein Wunsch nach einem intelligenterem und situationsangepassten Verhalten des UAVs beobachtet werden, welches einerseits über intelligente Funktionen wie einem automatischen Folgemodus für längere Flugabschnitte verfügt, gleichzeitig aber bei Bedarf auch in einen geführten Modus geschaltet werden kann, der eine präzise Positionierung des UAV erlaubt.

Welches Raumnutzungsverhalten zeigen Menschen während der kollokierten Interaktion mit einem hochautomatisierten, schweren Multikopter (WL 2.2)?

Bei der kollokierten Interaktion konnten drei Führungsarten in Bezug zur Raumnutzung beobachtet werden, die zudem in unterschiedlichen Flugphasen bevorzugt genutzt wurden. Die Mehrzahl der Probanden verwendete im Abflug die Pull-Methode (Heranziehen des UAVs), bei längeren Transitstrecken die Parallel-Methode (seitliches Führen des UAVs) und in der Landephase die Push-Methode (Wegdrücken des UAVs an die Landeposition).

Welchen Einfluss hat dabei die Erfahrung des Nutzers im Umgang mit UAVs auf die Interaktionsdistanz (WL 2.3)?

Nach der Landung des UAVs bestand zwischen den Probanden und dem Fluggerät eine Interaktionsdistanz von 4,7 bis 16,1 Metern. Zwar konnte eine Verkürzung der Interaktionsdistanz mit zunehmender UAV-Erfahrung beobachtet werden, jedoch konnte dies nicht statistisch nachgewiesen werden.

7.3.3 Wissenslücke 3: Verständigungsprobleme bei nonverbalen Dialogen mit technischen Systemen

In einer statischen Laborumgebung wurde die nonverbaler Mensch-UAV-Kommunikation untersucht. Hierfür sollten Probanden in zufälliger Reihenfolge vorgegebene Auftragsbeschreibungen in einem dialogähnlichen Vorgang nonverbal an ein technisches System übermitteln. Durch die Deaktivierung dialogunterstützender Hilfsfunktionen und den Einbau nicht übermittelbarer Auftragsbeschreibungen wurden Verständigungsprobleme provoziert, die Lösungsstrategien der Probanden zur Auflösung der dieser untersucht und die Frustrationsgrenzen ermittelt.

Welche Faktoren begünstigen Verständigungsprobleme bei einem nonverbalen Dialog mit einem technischen System (WL 3.1)?

Bei der nonverbalen Interaktion mit einem technischen Dialogpartner konnten drei potentielle Fehlerquellen identifiziert werden: das technische System, das verwendete Vokabular und das Verhalten des Nutzers selbst. Mit Klassifizierungsfehlern, Signalverwechslungen oder zu späten und zu häufigen Ausgaben kann das System den Nutzer bei der Interaktion verwirren und damit Missverständnisse begünstigen. Einen maßgeblichen Beitrag zu kommunikativen Missverständnissen liefert auch das verwendete Vokabular. Darstellungen sich visuell und motorisch zu sehr ähneln und sich nur in einem Parameter unterscheiden (z.B. der Handform), begünstigen Verwechslungen beim Nutzer bei der Zuordnung der Signale zu den korrekten Darstellungen. Als aktiver Teilnehmer der Interaktion kann natürlich auch der Mensch durch sein Verhalten Fehler in die Interaktion einbringen und zu Verständigungsproblemen beitragen. Dabei konnten Ausführungsfehler (falsche Darstellungen), Versäumnisse (vergessene Signale) sowie Planung- und Entscheidungsfehler (Syntaxveränderungen) beobachtet werden.

Die nonverbale Kommunikationsfähigkeit der Probanden war unterschiedlich stark ausgeprägt, zudem traten im Laufe der Interaktion ermüdungsbedingte Schwankungen in der Darstellungsqualität auf. Einige Probanden hatten zudem motorisch bedingte Schwierigkeiten, bestimmte Signale, die ungewöhnliche Arm- oder Fingerstellungen beinhalteten vorgabengetreu darzustellen. In Verbindung mit der Mehrdeutigkeit einzelner Signale war aus den Kommentaren der Teilnehmer ein klarer Wunsch nach alternativen Darstellungen erkennbar. Daher sollte beim Signaldesign ein weiteres Augenmerk auf ein redundantes Vokabular gelegt werden, bei dem für jedes Signal mindestens eine alternative Darstellungsart (z.B. mit und ohne Bewegungskomponente) verfügbar ist, um Handlungsop-

tionen bei nutzerspezifischen Darstellungsdefiziten oder systembedingten Erkennungsproblemen bieten zu können. Darüber hinaus ist es wichtig, dass geeignete Dialogsteuerungsfunktionen (z.B. letzte Eingabe ignorieren, letzte Ausgabe wiederholen, Dialog zurücksetzen) ebenfalls im Vokabular enthalten sind und zuverlässig, d.h. ohne Verwechslungsgefahr mit anderen Eingaben, vom Nutzer abgerufen werden können, um den Dialogverlauf beeinflussen und etwaige Kommunikationsfehler korrigieren zu können.

Welche nonverbalen Lösungsstrategien wenden Menschen an, wenn es Probleme bei der Verständigung über eine visuelle Schnittstelle gibt (WL 3.2)?

Abhängig davon, wo Probanden die Problemursache vermuteten (bei sich selbst oder dem technischen System), konnten fünf, teils aufeinander aufbauende Verhaltensmuster identifiziert werden, bei denen Modifikationen an der Darstellung, der Syntax oder dem Dienstbeschreibungsverfahren vorgenommen wurden. Das am häufigsten beobachtete und nach als erstes gezeigte Verhalten nach einer nicht erkannten Eingabe stellte die exakte Wiederholung der letzten Darstellung dar. Darauffolgend begannen einige Probanden eine mehr oder minder systematischen Optimierung der Darstellungsparameter, in der Erwartung, die Darstellungsgüte zu verbessern. Bei einigen führten diese Änderungen zum Erfolg, bei anderen zu mechanisch anmutenden Darstellungen und bei manchen zu teils irrationalem Verhalten. Aufgrund von Überoptimierung geriet ein Teil der Probanden in eine Art „Spirale der Verschlimmbesserung“, in der die vermeintliche Optimierung der Darstellung (extreme Verlangsamung oder übergroße Artikulation) in einer Verschlechterung der Erkennungsleistung des Systems (Fehldetektion bis gar keine Reaktion) resultierte und die Übermittlung des Signals praktisch unmöglich machte. Einige Teilnehmer kamen nach mehreren gescheiterten Kommunikationsversuchen zu dem Schluss, dass der Fehler beim System liegen muss und versuchten, die Problemursache durch systematische Tests näher einzugrenzen. Dazu testeten einige Probanden entweder andere Signale, um eventuelle Signalvertauschungen aufzudecken, oder andere Dienste, um deren Funktionsfähigkeit zu überprüfen. Ein weitere Lösungsstrategie stellte das Umschreiben des fehlerhaften Dienstes oder Signals durch andere ähnliche Dienste bzw. Signale dar. Als letzten Ausweg versuchten einige Probanden durch die Übermittlung eines geeigneten Signals aus dem Vokabular oder durch eine fragende Körperhaltung Hilfe oder Unterstützung beim System anzufragen.

Wo liegt die Frustrationsgrenze von Menschen bei einem nonverbalen Dialog bei anhaltenden Verständigungsproblemen (WL 3.3)?

Bei anhaltenden Verständigungsproblemen brachen die Probanden der betrachteten Stichprobe den nonverbalen Dialog mit dem technischen System nach durchschnittlich 29,5 Eingaben bzw. 161,4 Sekunden ab. Um negative Dialogverläufe zu vermeiden, werden daher zusätzliche systemseitige Unterstützungsfunktionen benötigt, die Dialogkonflikte rechtzeitig erkennen, um bei Bedarf proaktiv – beispielsweise in Form eines Alternativvorschlags – in den Dialog eingreifen zu können.

7.4 Ausblick und Anknüpfungspunkte für weitere Forschung

In dieser Arbeit wurden mehrere Anwendungen aus der unbemannten Luftfahrt vorgestellt, die von einer visuellen Benutzerschnittstelle profitieren können. Um eine genauere Abschätzung des Kommunikationsziels zu erreichen, kann eine dreidimensionale Modellierung des Nutzers sowie eine detailliertere Analyse manueller (z.B. Ausrichtung der Finger) und nonmanueller Kommunikationsmittel (z.B. Mimik und Blickbewegung) erforderlich sein. Darüber hinaus bietet die psycho-physiologische und signaltechnische Optimierung des nonverbalen Vokabulars als auch die Verwendung alternativer Verfahren zur Übermittlung der Systemantworten viel Raum für weitergehende Forschung. Neben

den Fragestellungen, die die perzeptiven und kommunikativen Fähigkeiten des technischen Systems betreffen, sollte der Blick auch in Richtung des Nutzers erfolgen. Die Untersuchungen in dieser Arbeit wurden größtenteils mit Mitarbeitern und Studenten aus dem Bereich der Luft- und Raumfahrt durchgeführt, sodass von einer gewissen Technikaffinität ausgegangen werden muss. Für eine tiefergehende Evaluierung dieser natürlichen Interaktionsform sollten jedoch auch andere, nichttechnische Nutzergruppen miteinbezogen und deren nonverbales Verhalten während der Interaktion untersucht werden.

Danksagung

„Der Weg ist immer besser als die schönste Herberge.“
– Miguel de Cervantes

Diese Arbeit entstand im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter an der Professur für Luftfahrttechnik des Instituts für Flugsysteme der Universität der Bundeswehr München. Während dieser fordernden, aber durchgängig spannenden Reise wurde ich von vielen außergewöhnlichen Menschen begleitet, ohne die diese Arbeit nicht möglich gewesen wäre und die ich daher nicht unerwähnt lassen möchte.

Meinem Doktorvater und Betreuer dieser Arbeit, Herrn Professor Peter Stütz, danke ich herzlich für die durchgehende wissenschaftliche und methodische Unterstützung sowie die eingeräumten Freiheiten, durch die ich erst meine Ideen in die Tat umsetzen und mich dadurch fachlich wie persönlich weiterentwickeln konnte. Herrn Professor Andreas Wendemuth danke ich für das Interesse an meiner Arbeit und die Übernahme des Zweitgutachtens. Herrn Professor Axel Schulte danke ich für die Leitung und Organisation der Promotionskommission. Ein großer Dank gilt auch Melanie Sabella, die zu jeder Zeit eine verlässliche Ansprechpartnerin für alle Belange des täglichen Institutslebens war und die Professuren zusammengehalten hat.

„Debug-Dennis“ und Marc danke ich für den Support bei allen Softwareproblemen. Euer Talent, Fehlerquellen in unbekanntem Code innerhalb kürzester Zeit aufzuspüren, hat mir so manch frustrierende Fehlersuche erspart. Markus danke ich für sein hervorragendes EPM-Tool, ohne dem ich mein Experimentalsystem hätte niemals so komfortabel betreiben können. Sten, Georg, Franziska, Jean, Christian, Denis, Markus, Julian, Adrian, Sebastian, beiden Simons und Mike danke ich für die motivierenden Gespräche, inspirierenden Diskussionen und gemeinsamen Erlebnisse, welche die tägliche Arbeit am Institut nicht als solche erschienen ließen.

Ein besonderer Dank geht an Alex und Peter, die mir nicht nur bei der Herstellung der mechanischen Komponenten geholfen, sondern mir auch das Fliegen verschiedenster UAVs beigebracht haben und mir mit ihrem nahezu unendlichen Erfahrungsschatz stets mit Rat und Tat zur Seite standen. Melanie, Michael, Jessica und Marvin danke ich für die Unterstützung bei der Dokumentation und Durchführung der Flugversuche. Ich möchte mich auch bei allen Probanden bedanken, die mir einen Teil ihrer wertvollen Lebenszeit geschenkt und freiwillig an meinen Experimenten in fensterlosen Laborräumen ohne Tageslicht als auch auf dem Testgelände bei sengender Hitze teilgenommen haben.

Mein größter Dank gilt meiner guten Seele Alice, meiner Familie und meinen Freunden, die mich während dieser langen Reise mit viel Geduld moralisch unterstützt und mir in den intensiven Phasen den Rücken freigehalten haben, damit ich mein Ziel nicht aus den Augen verliere.

Alexander Schelle,
Unterhaching, 08.01.2023

Literaturverzeichnis

[1] M. D. Hauser, N. Chomsky und W. T. Fitch, „The faculty of language: what is it, who has it, and how did it evolve?“, Science (New York, N.Y.), Jg. 298, Nr. 5598, S. 1569–1579, 2002, doi: 10.1126/science.298.5598.1569.

[2] J. R. Cauchard, M. Khamis, J. Garcia, M. Kljun und A. M. Brock, „Toward a roadmap for human-drone interaction“, *interactions*, Jg. 28, Nr. 2, S. 76–81, 2021, doi: 10.1145/3447889.

[3] SESAR Joint Undertaking, „European ATM Master Plan“, Brüssel, 15. Dez. 2015. [Online]. Verfügbar unter: https://www.sesarju.eu/sites/default/files/documents/reports/European%20ATM%20Master%20Plan%20Drone%20roadmap.pdf. Zugriff am: 11. September 2018.

[4] SESAR Joint Undertaking, „European Drones Outlook Study: Unlocking the value for Europe“, 2016. [Online]. Verfügbar unter: http://www.sesarju.eu/sites/default/files/documents/reports/European_Drones_Outlook_Study_2016.pdf. Zugriff am: 29. Oktober 2020.

[5] S. Barua, M. S. Shafkat Tanjim, A. N. Oishi, S. C. Das, M. A. Basar und S. A. Rafi, „Design and Implementation of Fire Extinguishing Ball Thrower Quadcopter“ in *2020 IEEE Region 10 Symposium (TENSYMP)*, Dhaka, Bangladesh, 2020, S. 1404–1407, doi: 10.1109/TENSYMP50017.2020.9230808.

[6] DRK-Kreisverband Bad Doberan e.V., *Einsatz von Drohnen in der Wasserrettung*. [Online]. Verfügbar unter: https://www.drk-dbr.de/der-kreisverband/spalte-25/nachrichten/meldungen.html (Zugriff am: 29. Oktober 2020).

[7] MATTERNET, *Matternet Launches Drone Delivery Operations at Labor Berlin in Germany*. Mountain View, CA, USA, 2020. [Online]. Verfügbar unter: https://mttr.net/images/Matternet_Press_Release-Labor_Berlin_Drone_Program-2020.11.17.pdf

[8] L. A. Young, „Aerobots as a ubiquitous part of society“, AHS Vertical Lift Aircraft Design Conference, San Francisco, CA, 2006. [Online]. Verfügbar unter: https://rotorcraft.arc.nasa.gov/Publications/files/Aerobots_n_Society_Final.pdf. Zugriff am: 3. Juni 2022.

[9] A. O. Abioye, S. D. Prior, G. T. Thomas, P. Saddington und S. D. Ramchurn, „Multimodal Human Aerobotic Interaction“ in *Advances in Business Information Systems and Analytics, Smart Technology Applications in Business Environments*, M. Tavana et al., Hg., IGI Global, 2017, S. 39–62, doi: 10.4018/978-1-5225-2492-2.ch003.

[10] M. Orsag, C. Korpela, P. Oh und S. Bogdan, Hg., *Aerial Manipulation*. Cham: Springer, 2018. [Online]. Verfügbar unter: http://www.springer.com/

[11] M. ki, J. cha und H. Lyu, „Detect and avoid system based on multi sensor fusion for UAV“ in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju, 2018, S. 1107–1109, doi: 10.1109/ICTC.2018.8539587.

[12] M. P. Owen, A. Panken, R. Moss, L. Alvarez und C. Leeper, „ACAS Xu: Integrated Collision Avoidance and Detect and Avoid Capability for UAS“ in *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*, San Diego, CA, USA, 2019, S. 1–10, doi: 10.1109/DASC43569.2019.9081758.

[13] A. Chakrabarty, C. A. Ippolito, J. Baculi, K. S. Krishnakumar und S. Hening, „Vehicle to Vehicle (V2V) communication for Collision avoidance for Multi-copters flying in UTM –TCL4“ in *AIAA Scitech 2019 Forum*, San Diego, California, 2019, doi: 10.2514/6.2019-0690.

[14] J. Zhao, Y. Chen und Y. Gong, „Study of Connectivity Probability of Vehicle-to-Vehicle and Vehicle-to-Infrastructure Communication Systems“ in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, Nanjing, China, 2016, S. 1–4, doi: 10.1109/VTCSpring.2016.7504493.

[15] P. Belanovic, D. Valerio, A. Paier, T. Zemen, F. Ricciato und C. F. Mecklenbrauker, „On Wireless Links for Vehicle-to-Infrastructure Communications“, *IEEE Trans. Veh. Technol.*, Jg. 59, Nr. 1, S. 269–282, 2010, doi: 10.1109/TVT.2009.2029119.

[16] A. Schelle, F. Völk, R. T. Schwarz, A. Knopp und P. Stütz, „Evaluation of a Multi-Mode-Transceiver for Enhanced UAV Visibility and Connectivity in Mixed ATM/UTM Contexts“, *Drones*, Jg. 6, Nr. 4, 2022, doi: 10.3390/drones6040080.

[17] W. S. Ng und E. Sharlin, „Collocated interaction with flying robots“ in *2011 RO-MAN: The 20th IEEE International Symposium on Robot and Human Interactive Communication*, Atlanta, GA, USA, 2011, S. 143–149, doi: 10.1109/ROMAN.2011.6005280.

[18] A. Wojciechowska, J. Frey, S. Sass, R. Shafir und J. R. Cauchard, „Collocated Human-Drone Interaction: Methodology and Approach Strategy“ in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Daegu, Korea (South), 2019, S. 172–181, doi: 10.1109/HRI.2019.8673127.

[19] D. Crystal, *Die Cambridge-Enzyklopädie der Sprache*. Frankfurt, New York: Campus, 1993. [Online]. Verfügbar unter: http://www.bsz-bw.de/depot/media/3400000/3421000/3421308/94_0053.html

[20] F. F. Mueller und M. Muirhead, „Jogging with a Quadcopter“ in *CHI '15: CHI Conference on Human Factors in Computing Systems*, Seoul Republic of Korea, 2015, S. 2023–2032, doi: 10.1145/2702123.2702472.

[21] Niederländisches Verteidigungsministerium, *NH90-maritieme gevechtshelikopter.* [Online]. Verfügbar unter: https://www.defensie.nl/binaries/large/content/gallery/defensie/content-afbeeldingen/onderwerpen/materieel/vliegtuigen-en-helikopters/nh90/d140502dv1302.jpg (Zugriff am: 23. April 2022).

[22] M. Ray, Hg., *Joseph-Michel and Jacques-Étienne Montgolfier.* Encyclopedia Britannica, 2008. [Online]. Verfügbar unter: https://www.britannica.com/biography/Montgolfier-brothers

[23] L. Brooke-Holland, „Unmanned Aerial Vehicles (drones): an introduction" SN06493, 5. Dez. 2012. [Online]. Verfügbar unter: https://www.files.ethz.ch/isn/157096/SN06493.pdf. Zugriff am: 5. Dezember 2021.

[24] A. C. Watts, V. G. Ambrosia und E. A. Hinkley, „Unmanned Aircraft Systems in Remote Sensing and Scientific Research: Classification and Considerations of Use", *Remote Sensing*, Jg. 4, Nr. 6, S. 1671–1692, 2012, doi: 10.3390/rs4061671.

[25] R. Weibel und R. J. Hansman, „Safety Considerations for Operation of Different Classes of UAVs in the NAS" in *AIAA 3rd "Unmanned Unlimited" Technical Conference, Workshop and Exhibit*, Chicago, Illinois, 2004, doi: 10.2514/6.2004-6421.

[26] N. Homainejad und C. Rizos, „Application of Multiple Categories of Unmanned Aircraft Systems (UAS) in Different Airspaces for Bushfire Monitoring and Response", *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XL-1/W4, S. 55–60, 2015, doi: 10.5194/isprsarchives-XL-1-W4-55-2015.

[27] S. Lange, „Flugroboter statt bemannter Militärflugzeuge?". Studie, Stiftung Wissenschaft und Politik, Deutsches Institut für Internationale Politik und Sicherheit, Berlin, 2003.

[28] M. d. F. Bento, „Unmanned aerial vehicles: an overview", *Inside GNSS*, Jg. 3, Nr. 1, S. 54–61, 2008.

[29] B. Vergouw, H. Nagel, G. Bondt und B. Custers, „Drone Technology: Types, Payloads, Applications, Frequency Spectrum Issues and Future Developments" in *Information Technology and Law Series, The Future of Drone Use*, B. Custers, Hg., The Hague: T.M.C. Asser Press, 2016, S. 21–45, doi: 10.1007/978-94-6265-132-6_2.

[30] M. Hassanalian und A. Abdelkefi, „Classifications, applications, and design challenges of drones: A review", *Progress in Aerospace Sciences*, Jg. 91, S. 99–131, 2017, doi: 10.1016/j.paerosci.2017.04.003.

[31] U. Ozdemir *et al.*, „Design of a Commercial Hybrid VTOL UAV System", *J Intell Robot Syst*, Jg. 74, 1-2, S. 371–393, 2014, doi: 10.1007/s10846-013-9900-0.

[32] Y. M. Chukewad, J. James, A. Singh und S. Fuller, „RoboFly: An Insect-Sized Robot With Simplified Fabrication That Is Capable of Flight, Ground, and Water Surface Locomotion", *IEEE Transactions on Robotics*, S. 1–16, 2021, doi: 10.1109/TRO.2021.3075374.

[33] *Leittechnik - Prozeßautomatisierung - Automatisierung mit Prozeßrechensystemen, Begriffe,* DIN V 19233:1998-07, Berlin, 1998.

[34] G. Beavers und H. Hexmoor, „Types and Limits of Agent Autonomy" in *Lecture Notes in Computer Science, Agents and Computational Autonomy*, G. Goos et al., Hg., Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, S. 95–102, doi: 10.1007/978-3-540-25928-2_8.

[35] E. Sholes, „Evolution of a UAV Autonomy Classification Taxonomy" in *2007 IEEE Aerospace Conference*, Big Sky, MT, USA, 2007, S. 1–16, doi: 10.1109/AERO.2007.352738.

[36] K. P. Valavanis, Hg., *Advances in unmanned aerial vehicles: State of the art and the road to autonomy.* Berlin: Springer, 2010.

[37] S. Balasubramanian, Y. M. Chukewad, J. M. James, G. L. Barrows und S. B. Fuller, „An Insect-Sized Robot That Uses a Custom-Built Onboard Camera and a Neural Network to Classify and Respond to Visual Input" in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, Enschede, 2018, S. 1297–1302, doi: 10.1109/BIOROB.2018.8488007.

[38] J. Burgués, V. Hernández, A. J. Lilienthal und S. Marco, „Smelling Nano Aerial Vehicle for Gas Source Localization and Mapping", *Sensors (Basel, Switzerland)*, Jg. 19, Nr. 3, 2019, doi: 10.3390/s19030478.

[39] AeroVironment, *Nano Hummingbird.* [Online]. Verfügbar unter: https://robots.ieee.org/robots/hummingbird/Photos/SD/hummingbird-photo4-full.jpg (Zugriff am: 3. Dezember 2021).

[40] „Durchführungsverordnung (EU) 2019/947 der Kommission vom 24. Mai 2019 über die Vorschriften und Verfahren für den Betrieb unbemannter Luftfahrzeuge: C/2019/3824", *Amtsblatt der Europäischen Union*, OJ L 152, S. 45–71, 11. Juni 2019. [Online]. Verfügbar unter: http://data.europa.eu/eli/reg_impl/2019/947/oj

[41] FLIR Systems, Inc., *Black Hornet PRS Datasheet.* [Online]. Verfügbar unter: https://flir.netx.net/file/asset/14121/original/attachment (Zugriff am: 3. Dezember 2021).

[42] Pexels GmbH, *7582146.* [Online]. Verfügbar unter: https://www.pexels.com/de-de/foto/fliegen-blauer-himmel-dji-drohne-kamera-7582146/ (Zugriff am: 15. März 2022).

[43] S. Wilke, *KZO.* [Online]. Verfügbar unter: https://www.bundeswehr.de/resource/image/112158/landscape_ratio16x9/1800/1013/57f1633ca239729bfd2d90d0ab8fdad8/Ik/bild-truppengattung-heeresaufklaerungstruppe15.jpg (Zugriff am: 16. August 2021).

[44] Airbus Defence and Space, *Zephyr HAPS.* [Online]. Verfügbar unter: https://mediacentre.airbus.com/mediacentre/media?mediaTitle=title_Zephyr+HAPS&mediaId=522841 (Zugriff am: 3. Dezember 2021).

[45] K. Kleinert, *Die Drohne Heron 1.* [Online]. Verfügbar unter: https://www.bundeswehr.de/de/ausruestung-technik-bundeswehr/luftsysteme-bundeswehr/drohne-luna (Zugriff am: 7. Dezember 2021).

[46] Prodrone Co. Ltd., *PD6B-AW-ARM: Flying Robot "DRORIA".* [Online]. Verfügbar unter: https://www.youtube.com/watch?v=tAgl2tzX6qI (Zugriff am: 7. Dezember 2021).

[47] T. Viglianco, *180816-F-ZH169-1222.* [Online]. Verfügbar unter: https://www.dvidshub.net/image/4657288/global-hawk-makes-historic-first-landing-red-flag-alaska (Zugriff am: 15. März 2022).

[48] „Unmanned aircraft systems: (UAS)“, Montréal, ICAO circular AN 190, 2011.

[49] J. Barnard, „Small UAV Command, Control and Communication Issues“, 2007. [Online]. Verfügbar unter: http://www.barnardmicrosystems.com/media/presentations/IET_UAV_C2_Barnard_DEC_2007.pdf

[50] J. Neumann, *Die Minidrohne Black Hornet in Aktion.* [Online]. Verfügbar unter: https://www.bundeswehr.de/resource/image/137398/landscape_ratio16x9/1000/563/1f82bcc3b1528caf0b36fd15b6bfd2f9/2FD3AD3276F4E4710A588058CCB28232/black-hornet-bildschirm.jpg (Zugriff am: 16. August 2021).

[51] L. Venetsky, M. Husni und M. Yager, „Gesture Recognition For UCAV-N Flight Deck Operations: Problem Definition Final Report“, Naval Air Systems Command, Lakehurst, NJ, 23. Jan. 2003. [Online]. Verfügbar unter: https://apps.dtic.mil/sti/pdfs/ADA422629.pdf. Zugriff am: 23. September 2021.

[52] U.S. Navy, *X-47B UCAS B-Roll.* [Online]. Verfügbar unter: https://www.youtube.com/watch?v=pCAe81aq6k0 (Zugriff am: 22. September 2021).

[53] U.S. Navy, *X-47B Completes First Carrier-based Arrested Landing.* [Online]. Verfügbar unter: https://www.youtube.com/watch?v=Rc2k6G8LuqY&t=147s (Zugriff am: 22. September 2021).

[54] A. K. Lampton, J. R. Gray und J. P. Miller, „Formal Evaluation of IMU-based Gesture Recognition for UAS Aircraft Carrier Deck Handling“ in *2018 AIAA Information Systems-AIAA Infotech @ Aerospace*, Kissimmee, Florida, 2018, doi: 10.2514/6.2018-0075.

[55] General Atomics Aeronautical Systems, Inc., *MQ-25 Deck Handling System Demonstration Reel.* [Online]. Verfügbar unter: https://www.youtube.com/watch?v=gDodQeJP5YA (Zugriff am: 10. Februar 2021).

[56] T. J. Ortega, *Photo Release -- X-47B Unmanned Aircraft Demonstrates the First Autonomous Aerial Refueling*. Patuxent River, 2015. [Online]. Verfügbar unter: https://news.northropgrumman.com/news/releases/photo-release-x-47b-unmanned-aircraft-demonstrates-the-first-autonomous-aerial-refueling

[57] Naval Air Systems Command, *Fueling the Future: MQ-25 first to conduct unmanned aerial tanking*. Patuxent River, 2021. [Online]. Verfügbar unter: https://www.navair.navy.mil/news/Fueling-Future-MQ-25-first-conduct-unmanned-aerial-tanking/Fri-06042021-0828

[58] K. Samson, *Kaman Unveils New Medium Lift UAV: the KARGO UAV Is a New Solution for Expeditionary Logisitics.* Bloomfield, Connecticut, 2021. [Online]. Verfügbar unter: https://www.kaman.com/news/kaman-unveils-new-medium-lift-uav-kargo-uav-new-solution-expeditionary-logisitics

[59] Verband Unbemannte Luftfahrt, Hg., „Analyse des deutschen Drohnenmarktes“, 12. Feb. 2019. [Online]. Verfügbar unter: https://www.verband-unbemannte-luftfahrt.de/wp-content/uploads/2019/03/VUL-Markststudie_Deutsch_final.pdf. Zugriff am: 26. Januar 2021.

[60] Verband Unbemannte Luftfahrt, Hg., „Analyse des deutschen Drohnenmarktes“, März 2021. [Online]. Verfügbar unter: https://www.verband-unbemannte-luftfahrt.de/wp-content/uploads/2021/03/VUL-Markststudie_2021_DE_f.pdf. Zugriff am: 15. März 2021.

[61] G. Von Schoenebeck, *Neue Schwerlast-Drohne Griff 300 fliegt mit 225 kg.* [Online]. Verfügbar unter: https://www.ingenieur.de/technik/fachbereiche/luftfahrt/neue-schwerlast-drohne-griff-300-fliegt-225-kg/ (Zugriff am: 2. Juli 2021).

[62] The Boeing Company, *Cargo Air Vehicle Completes First Outdoor Flight*, 2019. [Online]. Verfügbar unter: https://www.boeing.com/features/2019/05/cav-first-flight-05-19.page

[63] Volocopter GmbH, *Volocopter Expands into Utility Drone Business with Heavy-Lift VoloDrone*, 2019. [Online]. Verfügbar unter: https://www.volocopter.com/newsroom/volocopter-expands-into-utility-drone-business-with-heavy-lift-volodrone/

[64] M. E. Lussier, J. M. Bradley und C. Detweiler, „Extending Endurance of Multicopters: The Current State-of-the-Art“ in *AIAA Scitech 2019 Forum*, 2019, S. 1790.

[65] DLR, „Concept for Urban Airspace Integration: DLR U-Space Blueprint“, Wessling, Germany, 2017.

[66] Nesta, „Use cases for drones in your city: Introductory pack“, Nesta, 6. Dez. 2017. [Online]. Verfügbar unter: https://www.science-practice.com/teams/good-problems/projects/flying-high/flying-high-intro-pack.pdf. Zugriff am: 30. September 2021.

[67] Sunflower Labs Inc., *Sunflower.* [Online]. Verfügbar unter: https://www.sunflower-labs.com/ (Zugriff am: 30. Oktober 2020).

[68] M. Fritzsche, *Cleaning Drones: Let the robots clean for you!* [Online]. Verfügbar unter: http://window-cleaning-drones.com (Zugriff am: 6. Dezember 2021).

[69] J. Kral, *Debatte um Sylvester 2021: Drohnenshow statt Feuerwerk? (Update).* [Online]. Verfügbar unter: https://invidis.de/2020/12/debatte-um-sylvester-2021-drohnenshow-statt-feuerwerk/ (Zugriff am: 6. Dezember 2021).

[70] A. Williams, *The Drones Were Ready for This Moment.* [Online]. Verfügbar unter: https://www.nytimes.com/2020/05/23/style/drones-coronavirus.html (Zugriff am: 25. November 2021).

[71] K. Goeckel, *Drohnen sollen bayerische Polizei aus der Luft unterstützen.* [Online]. Verfügbar unter: https://www.br.de/nachrichten/bayern/drohnen-sollen-bayerische-polizei-aus-der-luft-unterstuetzen.

[72] R. A. Jaber, M. S. Sikder, R. A. Hossain, K. F. N. Malia und M. A. Rahman, „Unmanned Aerial Vehicle for Cleaning and Firefighting Purposes“ in *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, DHAKA, Bangladesh, 2021, S. 673–677, doi: 10.1109/ICREST51555.2021.9331147.

[73] J. Wang, *The First Fire Drill for High-Rise Fire Fighting Drones Was Held in Dazu, Chongqing.* [Online]. Verfügbar unter: https://www.ichongqing.info/2020/01/22/the-first-fire-drill-for-high-rise-fire-fighting-drones-was-held-in-dazu-chongqing/ (Zugriff am: 11. August 2021).

[74] L. St George, *Drone Technology proving to be pivotal for Shark Alerts*, 2020. [Online]. Verfügbar unter: https://ripper-corp.com/2020/08/drone-technology-proving-to-be-pivotal-for-shark-alers

[75] A. Schweikert, *Fast jeder zehnte Bauer setzt auf Drohnen.* Berlin, 2018. [Online]. Verfügbar unter: https://www.bitkom.org/Presse/Presseinformation/Fast-jeder-zehnte-Bauer-setzt-auf-Drohnen

[76] E. P. Fortes, „Seed Plant Drone for Reforestation“, *The Graduate Review*, Jg. 2, S. 13–26, 2017. [Online]. Verfügbar unter: https://vc.bridgew.edu/grad_rev/vol2/iss1/7

[77] UPS, *UPS Flight Forward adds innovative new aircraft, enhancing capabilities and network sustainability.* Atlanta, GA, 2021. [Online]. Verfügbar unter: https://about.ups.com/us/en/newsroom/press-releases/innovation-driven/ups-flight-forward-adds-new-aircraft.html

[78] FedEx, *Wing drone deliveries take flight with FedEx®.* Memphis, Tennessee, 2019. [Online]. Verfügbar unter: https://www.fedex.com/en-us/sustainability/wing-drones-transport-fedex-deliveries-directly-to-homes.html

[79] Wing Aviation LLC, *Wing's new model for drone delivery makes service possible in major U.S. metros*, 2021. [Online]. Verfügbar unter: https://blog.wing.com/2021/10/wings-new-model-for-drone-delivery.html

[80] Deutsche Post AG, *DHL Express launches its first regular fully-automated and intelligent urban drone delivery service.* Bonn/Guangzhou. [Online]. Verfügbar unter: https://www.dhl.com/global-en/home/press/press-archive/2019/dhl-express-launches-its-first-regular-fully-automated-and-intelligent-urban-drone-delivery-service.html

[81] Publications Office of the European Union, Hg., „Consolidated report on SESAR U-space research and innovation results“, SESAR Joint Undertaking, Luxembourg, 2020.

[82] V. Lappas *et al.*, „EuroDRONE, A European UTM Testbed for U-Space“ in *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*, Athens, Greece, 2020, S. 1766–1774, doi: 10.1109/ICUAS48674.2020.9214020.

[83] H. Treeck, *Volocopter absolviert ersten pilotierten Flugtaxi-Testflug in Südkorea.* 11.11.2021. [Online]. Verfügbar unter: https://www.volocopter.com/de/newsroom/volocopter-absolviert-ersten-pilotierten-flugtaxi-testflug-in-suedkorea/

[84] Lilium, *Stuttgart soll Knotenpunkt für Liliums regionales, emissionsfreies Flugnetzwerk werden.* München, 2021. [Online]. Verfügbar unter: https://lilium.com/newsroom-detail/stuttgart-to-become-regional-hub-for-liliums-planned-german-network?file=files/redaktion/newsroom/releases/Stuttgart/Lilium_Press_Release_Stuttgart_Airport_DE_1.pdf

[85] C. Barrado *et al.*, „U-Space Concept of Operations: A Key Enabler for Opening Airspace to Emerging Low-Altitude Operations“, *Aerospace*, Jg. 7, Nr. 3, S. 24, 2020, doi: 10.3390/aerospace7030024.

[86] J. A. Besada, A. M. Bernardos, L. Bergesio, D. Vaquero, I. Campana und J. R. Casar, „Drones-as-a-service: A management architecture to provide mission planning, resource brokerage and operation support for fleets of drones“ in *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Kyoto, Japan, 2019, S. 931–936, doi: 10.1109/PERCOMW.2019.8730838.

[87] DFS Aviation Services GmbH, *Unleash the power of remote tower control*, 2019. [Online]. Verfügbar unter: https://www.dfs.de/dfs_homepage/en/Press/Publications/RTC_Update%202019_web.pdf

[88] NATS Press Office, *London City is first major airport controlled by remote digital tower*, 2021. [Online]. Verfügbar unter: https://www.nats.aero/news/london-city-is-first-major-airport-controlled-by-remote-digital-tower/

[89] Frequentis AG, „White paper: Introduction to remote virtual tower“, 2016. [Online]. Verfügbar unter: https://www.frequentis.com/sites/default/files/support/2018-02/RVT_whitepaper.pdf. Zugriff am: 11. August 2021.

[90] R. Inam *et al.*, „Feasibility assessment to realise vehicle teleoperation using cellular networks“ in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Rio de Janeiro, Brazil, 2016, S. 2254–2260, doi: 10.1109/ITSC.2016.7795920.

[91] Bundesnetzagentur, Hg., „Frequenzplan“, Okt. 2019. [Online]. Verfügbar unter: https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Telekommunikation/Unternehmen_Institutionen/Frequenzen/20191202_Frequenzplan.pdf. Zugriff am: 30. November 2020.

[92] A. L. Drozd, „Spectrum-secure communications for autonomous UAS/UAV platforms“ in *2015 IEEE Military Communications Conference (MILCOM)*, Tampa, FL, 26.10.2015 - 28.10.2015, doi: 10.1109/MILCOM36929.2015.

[93] uAvionix, *microLink UAX-90045-01.* [Online]. Verfügbar unter: https://uavionix.com/products/microlink/#specs (Zugriff am: 9. Dezember 2021).

[94] Air Accident Investigation Unit, „Safety Investigation Report: ACCIDENT TO A UAV NAVIGATION ATLANTIC /OCULUS AT DIZY-LE-GROS, FRANCE ON 29 FEBRUARY 2016“ AAIU-2016-AII-01, 17. März 2017. [Online]. Verfügbar unter: https://www.skybrary.aero/bookshelf/books/4107.pdf. Zugriff am: 3. Januar 2020.

[95] M. Strohmeier, M. Schäfer, R. Pinheiro, V. Lenders und I. Martinovic, „On Perception and Reality in Wireless Air Traffic Communications Security“, *IEEE Trans. Intell. Transport. Syst.*, S. 1–20, 2016, doi: 10.1109/TITS.2016.2612584.

[96] Y. R. Yngvason und V. Le Strat, „UAS Safety Risk Portfolio and Analysis“, EASA, Safety Intelligence and Performance SM1.1, 1. Okt. 2016. [Online]. Verfügbar unter: https://www.easa.europa.eu/sites/default/files/dfu/UAS%20Safety%20Analysis.pdf. Zugriff am: 21. September 2021.

[97] M. Kugler-Kruse, *Die Entwicklung visueller Zeichensysteme: Von der Geste zur Gebärdensprache.* Zugl.: Bochum, Univ., Diss., 1988. Bochum: Brockmeyer, 1988.

[98] A. Kendon, „How gestures can become like words“ in *Cross-cultural perspectives in nonverbal communication*, Ashland, OH, US: Hogrefe & Huber Publishers, 1988, S. 131–141.

[99] C. Papaspyrou, A. v. Meyenn, M. Matthaei und B. Herrmann, *Grammatik der deutschen Gebärdensprache aus der Sicht gehörloser Fachleute.* Seedorf: Signum, 2008.

[100] E. Broszinsky-Schwabe, „Nonverbale Kommunikation“ in *Interkulturelle Kommunikation*, E. Broszinsky-Schwabe, Hg., Wiesbaden: VS Verlag für Sozialwissenschaften, 2011, S. 117–139, doi: 10.1007/978-3-531-92764-0_6.

[101] K. Emmorey, *Language, cognition, and the brain: Insights from sign language research.* Mahwah, N.J: Lawrence Erlbaum Associates, 2002. [Online]. Verfügbar unter: http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10084535

[102] K. Bühler, *Ausdruckstheorie,* 2. Aufl. Stuttgart: Fischer, 1968.

[103] C. Müller, „Wie Gesten bedeuten Eine kognitiv-linguistische und sequenzanalytische Perspektive“, *Sprache und Literatur*, Jg. 41, Nr. 1, S. 37–68, 2010, doi: 10.1163/25890859-041-01-90000004.

[104] U. Bellugi und S. Fischer, „A comparison of sign language and spoken language“, *Cognition*, Jg. 1, 2-3, S. 173–200, 1972.

[105] J. Lindemann, „Vergleich der Produktionsgeschwindigkeit von Laut- und Gebärdensprache“ in *Studientexte zur Sprachkommunikation*, Bd. 64, *Elektronische Sprachsignalverarbeitung 2012: Tagungsband der 23. Konferenz, Cottbus, 29. - 31. August 2012*, Dresden: TUDpress, 2012, S. 236–243. [Online]. Verfügbar unter: https://www.essv.de/paper.php?id=161

[106] P. Boyes Braem, *Einführung in die Gebärdensprache und ihre Erforschung,* 2. Aufl. Hamburg: Signum-Verl., 1992.

[107] K. Koch *et al.,* „How much the eye tells the brain“, *Current biology : CB*, Jg. 16, Nr. 14, S. 1428–1434, 2006, doi: 10.1016/j.cub.2006.05.056.

[108] P. Watzlawick, J. H. Beavin und D. D. Jackson, *Menschliche Kommunikation: Formen, Störungen, Paradoxien,* 1. Aufl. s.l.: Hogrefe Verlag Bern (ehemals Hans Huber), 2016. [Online]. Verfügbar unter: http://ebooks.ciando.com/book/index.cfm?bok_id/2240200

[109] D. McNeill, *Hand and mind:: What gestures reveal about thought.* University of Chicago Press, 1992.

[110] P. Ekman und W. V. Friesen, „The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding“, *Semiotica*, Jg. 1, Nr. 1, S. 49–98, 1969, doi: 10.1515/semi.1969.1.1.49.

[111] P. Ekman, „Emotional and conversational nonverbal signals“ in *Philosophical studies series*, v. 99, *Language, knowledge, and representation: Proceedings of the Sixth International Colloquium on Cognitive Science (ICCS-99)*, J. M. Larrazabal, Hg., Dordrecht: Kluwer Academic Publ, 2004, S. 39–50.

[112] A. Kendon, „Gesture and speech: How they interact“ in *Sage annual reviews of communication research*, Bd. 11, *Nonverbal interaction*, J. M. Wiemann, Hg., 1 Aufl., Beverly Hills, Calif.: Sage, 1983, S. 13–45.

[113] S. Teßendorf, „4. Emblems, quotable gestures, or conventionalized body movements“ in *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1*, C. Müller et al., Hg., Berlin, Boston: De Gruyter, 2013, doi: 10.1515/9783110261318.82.

[114] D. Morris, *Der Mensch, mit dem wir leben: Ein Handbuch unseres Verhaltens ; mit 403 Farbfotos, Zeichnungen und Diagrammen,* 1. Aufl. München: Droemer Knaur, 1978.

[115] M. Meissner und S. B. Philpott, „The Sign Language of Sawmill Workers in British Columbia“, *Sign Language Studies*, Nr. 9, S. 291–308, 1975. [Online]. Verfügbar unter: https://search.proquest.com/docview/1297873110?accountid=12746

[116] *Mobile and Locomotive Cranes: Safety Standard for Cableways, Cranes, Derricks, Hoists, Hooks, Jacks and Slings*, B30.5-2004, The American Society of Mechanical Engineers, Sep. 2004. [Online]. Verfügbar unter: https://law.resource.org/pub/us/cfr/ibr/002/asme.b30.5.2004.pdf

[117] Logistikkommando der Bundeswehr, „Ausbildungshilfe KfWBw Nr. 2: Kraftfahrbetrieb Einweiserzeichen“, 2016.

[118] *Convention on International Civil Aviation, Annex 2: Rules of the Air,* ICAO, Jul. 2005. [Online]. Verfügbar unter: http://www.icao.int/Meetings/anconf12/Document%20Archive/an02_cons[1].pdf

[119] *Visual Signals: TC 3-21.60*, 2017. [Online]. Verfügbar unter: https://armypubs.army.mil/e-pubs/DR_pubs/DR_a/pdf/web/ARN2747_TC%203-21x60%20FINAL%20WEB.pdf

[120] U.S. Navy, „Aircraft Signals NATOPS Manual: NAVAIR 00-80T-113“, Washington, DC, 1. Dez. 2001.

[121] Headquarters, Department of the Army, „Shipboard Operartions Field Manual Number 1-564: Appendix A: Aircraft Handling Signals“, Washington, DC, 29. Juni 1997.

[122] Army Publishing Directorate, „Techniques for Forward Arming and Refueling Points (ATP 3-04.17)“, 18. Mai 2021. [Online]. Verfügbar unter: https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN32371-ATP_3-04.17-001-WEB-3.pdf. Zugriff am: 13. Juni 2022.

[123] *STANAG 3117: Aircraft Marshalling Signals,* North Atlantic Treaty Organization (NATO), Sep. 2014.

[124] *Common Hand Signals for Recreational Scuba Diving,* World Recreational Scuba Training Council, Dez. 2005. [Online]. Verfügbar unter: https://wrstc.com/downloads/12%20-%20Common%20Hand%20Signals.pdf

[125] Bundeswehr, „Handzeichen im Tauchdienst“. LT 052/7/097.

[126] J. Prosser und H. V. Grey, *Cave Diving Communications,* 1. Aufl. Branford, Florida: The Cave Diving Section of the National Speleogological Society, Inc., 1990. [Online]. Verfügbar unter: http://www.swiss-cave-diving.ch/PDF-dateien/cavediving-signals.pdf

[127] *Apollo Recovery Operational Procedures Manual: MSC-01856,* National Aeronautics and Space Administration (NASA), Jun. 1971.

[128] P. Morgan-Dimmick, *EVA Hand Signals: MDRS Crew 141.* [Online]. Verfügbar unter: https://mdrs141.weebly.com/blog/eva-hand-signals (Zugriff am: 16. September 2020).

[129] ATI Avionics Inc., *Model 901 Signal Light Gun.* [Online]. Verfügbar unter: https://atiavionics.com/signal-light-gun (Zugriff am: 18. August 2021).

[130] Corpo nazionale soccorso alpino e speleologico, *Rescue Sign.* [Online]. Verfügbar unter: https://www.cnsas.it/wp-content/uploads/2015/07/Rescue-Sign-2.jpg (Zugriff am: 2. Mai 2022).

[131] *Field Manual No. 21-60: Visual Signals: FM 21-60*, 1987.

[132] Eurocontrol, „AVAIR Safety Bulletin No 20: Summer seasons and full years 2013-2017“, Sep. 2019. [Online]. Verfügbar unter: https://www.eurocontrol.int/sites/default/files/2019-09/evair-bulletin-20.pdf. Zugriff am: 30. August 2021.

[133] R. Petrova, „Die internationalen Unterschiede der Gebärdensprache – ein Interview mit Hochschuldozentin Liona Paulus“, Hochschule Fresenius online plus GmbH, 14. Dez. 2017. [Online]. Verfügbar unter: https://www.hs-fresenius.de/blog/wissen/die-internationalen-unterschiede-der-gebaerdensprache-ein-interview-mit-hochschuldozentin-liona-paulus/. Zugriff am: 9. November 2020.

[134] *Gesetz zur Gleichstellung von Menschen mit Behinderungen (Behindertengleichstellungsgesetz): BGG*, 2002.

[135] B. Sequeira Gerardo, „Deutsche Gebärdensprache in Ostbelgien anerkannt“, *Taubenschlag*, 19. März 2019, 2019. [Online]. Verfügbar unter: https://www.taubenschlag.de/2019/03/deutsche-gebaerdensprache-in-ostbelgien-anerkannt/. Zugriff am: 11. September 2020.

[136] B. Sequeira Gerardo, „Gebärdensprache in Luxemburg anerkannt“, *Taubenschlag*, 30. Juli 2018, 2018. [Online]. Verfügbar unter: https://www.taubenschlag.de/2018/07/gebaerdensprache-in-luxemburg-anerkannt/. Zugriff am: 9. November 2020.

[137] R. Battison, *Lexical borrowing in American sign language*. Linstok Press, 1978.

[138] D. C. Lillo-Martin und J. Gajewski, „One grammar or two? Sign Languages and the Nature of Human Language“, *Wiley interdisciplinary reviews. Cognitive science*, Jg. 5, Nr. 4, S. 387–401, 2014, doi: 10.1002/wcs.1297.

[139] G. Albeaino, M. Gheisari und R. R. A. Issa, „Human-Drone Interaction (HDI): Opportunities and Considerations in Construction“ in *Automation and Robotics in the Architecture, Engineering, and Construction Industry*, H. Jebelli, M. Habibnezhad, S. Shayesteh, S. Asadi und S. Lee, Hg., Cham: Springer International Publishing, 2022, S. 111–142, doi: 10.1007/978-3-030-77163-8_6.

[140] M. Shafiee, Z. Zhou, L. Mei, F. Dinmohammadi, J. Karama und D. Flynn, „Unmanned Aerial Drones for Inspection of Offshore Wind Turbines: A Mission-Critical Failure Analysis“, *Robotics*, Jg. 10, Nr. 1, S. 26, 2021, doi: 10.3390/robotics10010026.

[141] A. Alharam, E. Almansoori, W. Elmadeny und H. Alnoiami, „Real Time AI-Based Pipeline Inspection using Drone for Oil and Gas Industries in Bahrain“ in *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)*, Sakheer, Bahrain, 2020, S. 1–5, doi: 10.1109/3ICT51146.2020.9312021.

[142] F. Stephan, N. Reinsperger, M. Grünthal, D. Paulicke und P. Jahn, „Human drone interaction in delivery of medical supplies: A scoping review of experimental studies“, *PloS one*, Jg. 17, Nr. 4, e0267664, 2022, doi: 10.1371/journal.pone.0267664.

[143] D. Cawthorne und A. R.-V. Wynsberghe, „From HealthDrone to FrugalDrone: Value-Sensitive Design of a Blood Sample Transportation Drone“ in *2019 IEEE International Symposium on Technology and Society (ISTAS)*, Medford, MA, USA, 2019, S. 1–7, doi: 10.1109/ISTAS48451.2019.8938019.

[144] B. Wang *et al.*, „Current technologies and challenges of applying fuel cell hybrid propulsion systems in unmanned aerial vehicles“, *Progress in Aerospace Sciences*, Jg. 116, S. 100620, 2020, doi: 10.1016/j.paerosci.2020.100620.

[145] X. Zhou *et al.*, „Swarm of micro flying robots in the wild“, *Science robotics*, Jg. 7, Nr. 66, eabm5954, 2022, doi: 10.1126/scirobotics.abm5954.

[146] M. Petrlik, T. Baca, D. Hert, M. Vrba, T. Krajnik und M. Saska, „A Robust UAV System for Operations in a Constrained Environment“, *IEEE Robotics and Automation Letters*, Jg. 5, Nr. 2, S. 2169–2176, 2020, doi: 10.1109/LRA.2020.2970980.

[147] H. Baek und J. Lim, „Design of Future UAV-Relay Tactical Data Link for Reliable UAV Control and Situational Awareness“, *IEEE Commun. Mag.*, Jg. 56, Nr. 10, S. 144–150, 2018, doi: 10.1109/MCOM.2018.1700259.

[148] J. M. Teixeira, R. Ferreira, M. Santos und V. Teichrieb, „Teleoperation Using Google Glass and AR, Drone for Structural Inspection“ in *2014 XVI Symposium on Virtual and Augmented Reality (SVR)*, Piata Salvador, Bahia, Brazil, 2014, S. 28–36, doi: 10.1109/SVR.2014.42.

[149] M. Obaid, W. Johal und O. Mubin, „Domestic Drones“ in *HAI '20: 8th International Conference on Human-Agent Interaction*, Virtual Event USA, 2020, S. 196–203, doi: 10.1145/3406499.3415076.

[150] P. van de Voorde, S. Gautama, A. Momont, C. M. Ionescu, P. de Paepe und N. Fraeyman, „The drone ambulance A-UAS: golden bullet or just a blank?“, *Resuscitation*, Jg. 116, S. 46–48, 2017, doi: 10.1016/j.resuscitation.2017.04.037.

[151] S. Mayer, L. Lischke und P. W. Woźniak, „Drones for Search and Rescue“ in *1st International Workshop on Human-Drone Interaction*, 2019.

[152] M. Xiong, D. Zeng, H. Yao und Y. Li, „A Crowd Simulation Based UAV Control Architecture for Industrial Disaster Evacuation“ in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, Nanjing, China, 2016, S. 1–5, doi: 10.1109/VTCSpring.2016.7504062.

[153] A. Alhindi *et al.*, „Emergency Planning for UAV-Controlled Crowd Evacuations“, *Applied Sciences*, Jg. 11, Nr. 19, S. 9009, 2021, doi: 10.3390/app11199009.

[154] T. Uzakov, T. P. Nascimento und M. Saska, „UAV Vision-Based Nonlinear Formation Control Applied to Inspection of Electrical Power Lines“ in *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*, Athens, Greece, 2020, S. 1301–1308, doi: 10.1109/ICUAS48674.2020.9213967.

[155] M. Monajjemi, S. Mohaimenianour und R. Vaughan, „UAV, come to me: End-to-end, multi-scale situated HRI with an uninstrumented human and a distant UAV“ in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, 2016, S. 4410–4417, doi: 10.1109/IROS.2016.7759649.

[156] J. R. Cauchard, K. Y. Zhai und J. A. Landay, „Drone & me: an exploration into natural human-drone interaction“ in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, S. 361–365.

[157] M. A. Goodrich und A. C. Schultz, „Human-Robot Interaction: A Survey“, *FNT in Human-Computer Interaction*, Jg. 1, Nr. 3, S. 203–275, 2007, doi: 10.1561/1100000005.

[158] D. Tezza und M. Andujar, „The State-of-the-Art of Human–Drone Interaction: A Survey“, *IEEE Access*, Jg. 7, S. 167438–167454, 2019, doi: 10.1109/ACCESS.2019.2953900.

[159] G. Xiang, A. Hardy, M. Rajeh und L. Venuthurupalli, „Design of the life-ring drone delivery system for rip current rescue“ in *2016 Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA, USA, 2016, S. 181–186, doi: 10.1109/SIEDS.2016.7489295.

[160] A. Romanowski *et al.*, „Towards Supporting Remote Cheering during Running Races with Drone Technology“ in *CHI '17: CHI Conference on Human Factors in Computing Systems*, Denver Colorado USA, 2017, S. 2867–2874, doi: 10.1145/3027063.3053218.

[161] J. Theiß, I. Albert, N. Burkard und M. Herrlich, „Towards Using Drones as Personal Spatial Search Assistants“ in *MuC '21: Mensch und Computer 2021*, Ingolstadt Germany, 2021, S. 180–188, doi: 10.1145/3473856.3473877.

[162] M. Avila Soto, M. Funk, M. Hoppe, R. Boldt, K. Wolf und N. Henze, „DroneNavigator“ in *ASSETS '17: The 19th International ACM SIGACCESS Conference on Computers and Accessibility*, Baltimore Maryland USA, 2017, S. 300–304, doi: 10.1145/3132525.3132556.

[163] A. Gomes, C. Rubens, S. Braley und R. Vertegaal, „BitDrones: Towards Using 3D Nanocopter Displays as Interactive Self-Levitating Programmable Matter“ in *CHI'16: CHI Conference on Human Factors in Computing Systems*, San Jose California USA, 2016, S. 770–780, doi: 10.1145/2858036.2858519.

[164] J. La Delfa, M. A. Baytas, R. Patibanda, H. Ngari, R. A. Khot und F. '. Mueller, „Drone Chi: Somaesthetic Human-Drone Interaction“ in *CHI '20: CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA, 2020, S. 1–13, doi: 10.1145/3313831.3376786.

[165] M. A. Baytaş, D. Çay, Y. Zhang, M. Obaid, A. E. Yantaç und M. Fjeld, „The Design of Social Drones“ in *CHI '19: CHI Conference on Human Factors in Computing Systems*, Glasgow Scotland Uk, 2019, S. 1–13, doi: 10.1145/3290605.3300480.

[166] M. A. Baytaş, D. Çay, Y. Zhang, M. Obaid, A. E. Yantaç und M. Fjeld, *Consolidating Empirical Research Findings on Social Drone Design: What are social drones?* [Online]. Verfügbar unter: https://github.com/socialdrones/designreview (Zugriff am: 25. November 2021).

[167] A. Yeh *et al.*, „Exploring Proxemics for Human-Drone Interaction“ in *Proceedings of the 5th International Conference on Human Agent Interaction*, 2017, S. 81–88, doi: 10.1145/3125739.3125773.

[168] S. Biundo-Stephan und A. Wendemuth, Hg., *Companion technology: A paradigm shift in human-technology interaction*. Cham: Springer, 2017.

[169] S. Biundo und A. Wendemuth, „Von kognitiven technischen Systemen zu Companion-Systemen“, *Künstl Intell*, Jg. 24, Nr. 4, S. 335–339, 2010, doi: 10.1007/s13218-010-0056-9.

[170] C. F. Liew, „Towards Human-Robot Interaction in Flying Robots : A User Accompanying Model and A Sensing Interface“. Dissertation, University of Tokyo, Tokyo, 2016.

[171] *Ergonomie der Mensch-System-Interaktion - Teil 110: Grundsätze der Dialoggestaltung,* DIN EN ISO 9241-110:2008-09, Berlin, 2008.

[172] M. Funk, „Human-Drone Interaction: Let's Get Ready for Flying User Interfaces!“, *interactions*, Jg. 25, Nr. 3, S. 78–81, 2018, doi: 10.1145/3194317.

[173] G. Jones, N. Berthouze, R. Bielski und S. Julier, „Towards a situated, multimodal interface for multiple UAV control“ in *2010 IEEE International Conference on Robotics and Automation (ICRA 2010)*, Anchorage, AK, 2010, S. 1739–1744, doi: 10.1109/ROBOT.2010.5509960.

[174] O. Jokisch, I. Siegert, M. Maruschke, T. Strutz und A. Ronzhin, „Don’t Talk to Noisy Drones – Acoustic Interaction with Unmanned Aerial Vehicles“ in *Lecture Notes in Computer Science, Speech and Computer*, A. A. Salah, A. Karpov und R. Potapova, Hg., Cham: Springer International Publishing, 2019, S. 180–190, doi: 10.1007/978-3-030-26061-3_19.

[175] O. A. Abioye, D. S. Prior, P. Saddington und D. S. Ramchurn, „Effects of Varying Noise Levels and Lighting Levels on Multimodal Speech and Visual Gesture Interaction with Aerobots“, *Applied Sciences*, Jg. 9, Nr. 10, 2019, Art. no. 2066. [Online]. Verfügbar unter: http://www.mdpi.com/2076-3417/9/10/2066

[176] É. Lombard, „Le signe de l'élévation de la voix“, *Annales des Maladies de L’Oreille et du Larynx*, Jg. 37, S. 101–119, 1911.

[177] A. Torija, R. H. Self und J. L. Lawrence, „Psychoacoustic Characterisation of a Small Fixed-pitch Quadcopter“ in *Noise control for a better environment: Proceedings of INTERNOISE 2019 : the 48th International Congress and Exhibition on Noise Control Engineering*, A. Calvo-Manzano, Hg., Madrid: Sociedad Española de Acústica, 2019.

[178] B. Schäffer, R. Pieren, K. Heutschi, J. M. Wunderli und S. Becker, „Drone Noise Emission Characteristics and Noise Effects on Humans-A Systematic Review“, *International journal of environmental research and public health*, Jg. 18, Nr. 11, 2021, doi: 10.3390/ijerph18115940.

[179] M. Khamis, A. Kienle, F. Alt und A. Bulling, „GazeDrone“ in *MobiSys '18: The 16th Annual International Conference on Mobile Systems, Applications, and Services*, Munich Germany, 2018, S. 66–71, doi: 10.1145/3213526.3213539.

[180] B. Gromov, J. Guzzi, L. M. Gambardella und A. Giusti, „Intuitive 3D Control of a Quadrotor in User Proximity with Pointing Gestures“ in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France, 31.05.2020 - 31.08.2020, S. 5964–5971, doi: 10.1109/ICRA40945.2020.9196654.

[181] B. Sanders, D. Vincenzi und Y. Shen, „Investigation of Gesture Based UAV Control“ in *Advances in Intelligent Systems and Computing, Advances in Human Factors in Robots and Unmanned Systems*, J. Chen, Hg., Cham: Springer International Publishing, 2018, S. 205–215, doi: 10.1007/978-3-319-60384-1_20.

[182] E. Peshkova, M. Hitz, D. Ahlstrom, R. W. Alexandrowicz und A. Kopper, „Exploring intuitiveness of metaphor-based gestures for UAV navigation“ in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Lisbon, 2017, S. 175–182, doi: 10.1109/ROMAN.2017.8172298.

[183] M. Bhat, G. Mahto, S. Kesaria, V. Femandes und K. Arya, „Real-time gesture control UAV with a low resource framework“ in *2021 International Symposium of Asian Control Association on Intelligent Robotics and Industrial Automation (IRIA)*, Goa, India, 2021, S. 19–24, doi: 10.1109/IRIA53009.2021.9588681.

[184] S. Mohaimenianpour und R. Vaughan, „Hands and Faces, Fast: Mono-Camera User Detection Robust Enough to Directly Control a UAV in Flight“ in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, S. 5224–5231, doi: 10.1109/IROS.2018.8593709.

[185] E. Tsykunov *et al.*, „DroneStick: Flying Joystick as a Novel Type of Interface“, 18. Okt. 2021. [Online]. Verfügbar unter: http://arxiv.org/pdf/2110.09041v1.

[186] A. O. Abioye, S. D. Prior, G. T. Thomas, P. Saddington und S. D. Ramchurn, „The Multimodal Speech and Visual Gesture (mSVG) Control Model for a Practical Patrol, Search, and Rescue Aerobot“ in *Lecture Notes in Computer Science, Towards Autonomous Robotic Systems*, M. Giuliani, T. Assaf und M. E. Giannaccini, Hg., Cham: Springer International Publishing, 2018, S. 423–437, doi: 10.1007/978-3-319-96728-8_36.

[187] A. Schelle und P. Stütz, „Visual Communication with UAV: Use Cases and Achievements“ in *Communications in Computer and Information Science, Computer Analysis of Images and Patterns*, M. Vento et al., Hg., Cham: Springer International Publishing, 2019, S. 120–128, doi: 10.1007/978-3-030-29930-9_12.

[188] J. Cacace, A. Finzi und V. Lippiello, „Multimodal interaction with multiple co-located drones in search and rescue missions“, *arXiv preprint arXiv:1605.07316*, 2016.

[189] E. Peshkova, M. Hitz und D. Ahlström, „Exploring User-Defined Gestures and Voice Commands to Control an Unmanned Aerial Vehicle“ in *Intelligent Technologies for Interactive Entertainment: 8th International Conference, INTETAIN 2016, Utrecht, The Netherlands, June 28–30, 2016, Revised Selected Papers*, R. Poppe, J.-J. Meyer, R. Veltkamp und M. Dastani, Hg., Cham: Springer International Publishing, 2017, S. 47–62, doi: 10.1007/978-3-319-49616-0_5.

[190] C. F. Graumann, „Interaktion und Kommunikation“ in *Handbuch der Psychologie, Bd. 7. Sozialpsychologie. Halbbd. 2. Forschungsberichte*, C. F. Graumann, Hg., Göttingen: Hogrefe, 1972, S. 1109–1262.

[191] F. W. Nerdinger, „Interaktion und Kommunikation“ in *Springer-Lehrbuch, Arbeits- und Organisationspsychologie*, F. W. Nerdinger, G. Blickle und N. Schaper, Hg., Berlin, Heidelberg: Springer Berlin Heidelberg, 2019, S. 63–80, doi: 10.1007/978-3-662-56666-4_5.

[192] O. Lavrynenko, G. Konakhovych und D. Bakhtiiarov, Hg., *Method of voice control functions of the UAV*. 2016 4th International Conference on Methods and Systems of Navigation and Motion Control (MSNMC), 2016.

[193] Z. Marinov, S. Vasileva, Q. Wang, C. Seibold, J. Zhang und R. Stiefelhagen, „Pose2Drone: A Skeleton-Pose-based Framework for Human-Drone Interaction“, 27. Mai 2021. [Online]. Verfügbar unter: http://arxiv.org/pdf/2105.13204v2.

[194] T. Sun, S. Nie, D.-Y. Yeung und S. Shen, „Gesture-based piloting of an aerial robot using monocular vision“ in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, Singapore, 2017, S. 5913–5920, doi: 10.1109/ICRA.2017.7989696.

[195] A. C. S. Medeiros, P. Ratsamee, J. Orlosky, Y. Uranishi, M. Higashida und H. Takemura, „3D pointing gestures as target selection tools: guiding monocular UAVs during window selection in an outdoor environment“, *Robomech J*, Jg. 8, Nr. 1, 2021, doi: 10.1186/s40648-021-00200-w.

[196] V. M. Monajjemi, J. Wawerla, R. Vaughan und G. Mori, „HRI in the sky:: Creating and commanding teams of UAVs with a vision-mediated gestural interface“ in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, 2013, S. 617–623.

[197] F. Patrona, I. Mademlis und I. Pitas, „An Overview of Hand Gesture Languages for Autonomous UAV Handling“ in *2021 Aerial Robotic Systems Physically Interacting with the Environment (AIRPHARO)*, Biograd na Moru, Croatia, 2021, S. 1–7, doi: 10.1109/AIRPHARO52252.2021.9571027.

[198] DJI, „MAVIC AIR User Manual: SmartCapture Mode“, 2018. [Online]. Verfügbar unter: https://dl.djicdn.com/downloads/Mavic%20Air/Mavic%20Air%20User%20Manual%20v1.0.pdf. Zugriff am: 20. März 2019.

[199] R. C. Taylor, D. R. High und J. P. Thompson, “Apparatuses and methods for gesture-controlled unmanned aerial vehicles,” US 2018/0251219 A1. USA 15/913,544, Sep 6, 2018.

[200] Y. Song, D. Demirdjian und R. Davis, „Continuous body and hand gesture recognition for natural human-computer interaction“, *ACM Transactions on Interactive Intelligent Systems (TiiS)*, Jg. 2, Nr. 1, S. 1–28, 2012, doi: 10.1145/2133366.2133371.

[201] A. Schelle und P. Stütz, „Modelling Visual Communication with UAS“ in *Lecture Notes in Computer Science, Modelling and Simulation for Autonomous Systems*, J. Hodicky, Hg., Cham: Springer International Publishing, 2016, S. 81–98, doi: 10.1007/978-3-319-47605-6_7.

[202] A. Schelle und P. Stütz, „Gestural Transmission of Tasking Information to an Airborne UAV“ in *Lecture Notes in Computer Science, Human Interface and the Management of Information. Interaction, Visualization, and Analytics*, S. Yamamoto und H. Mori, Hg., Cham: Springer International Publishing, 2018, S. 318–335, doi: 10.1007/978-3-319-92043-6_27.

[203] A. Sorokowska *et al.*, „Preferred Interpersonal Distances: A Global Comparison“, *Journal of Cross-Cultural Psychology*, Jg. 48, Nr. 4, S. 577–592, 2017, doi: 10.1177/0022022117698039.

[204] E. T. Hall, *The hidden dimension: Man's use of space in public and private*. London: Bodley Head, 1969.

[205] B. A. Duncan und R. R. Murphy, „Comfortable approach distance with small Unmanned Aerial Vehicles“ in *2013 IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Gyeongju, 2013, S. 786–792, doi: 10.1109/ROMAN.2013.6628409.

[206] P. Abtahi, D. Y. Zhao, J. L. E. und J. A. Landay, „Drone Near Me“, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Jg. 1, Nr. 3, S. 1–8, 2017, doi: 10.1145/3130899.

[207] J. Han, D. Moore und I. Bae, „Exploring the Social Proxemics of Human-Drone Interaction“ in *International Journal of Advanced Smart Convergence*, 2019.

[208] M. Lieser, U. Schwanecke und J. Berdux, „Evaluating Distances in Tactile Human-Drone Interaction“ in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, Vancouver, BC, Canada, 2021, S. 1275–1282, doi: 10.1109/RO-MAN50785.2021.9515313.

[209] W. Mao, Z. Zhang, L. Qiu, J. He, Y. Cui und S. Yun, „Indoor Follow Me Drone“ in *MobiSys'17: The 15th Annual International Conference on Mobile Systems, Applications, and Services*, 2017, S. 345–358, doi: 10.1145/3081333.3081362.

[210] A. Colley, L. Virtanen, P. Knierim und J. Häkkilä, „Investigating drone motion as pedestrian guidance“ in *MUM 2017: The 16th International Conference on Mobile and Ubiquitous Multimedia*, Stuttgart Germany, 2017, S. 143–150, doi: 10.1145/3152832.3152837.

[211] J. Auda, M. Weigel, J. R. Cauchard und S. Schneegass, „Understanding Drone Landing on the Human Body“ in *MobileHCI '21: 23rd International Conference on Mobile Human-Computer Interaction*, Toulouse & Virtual France, 2021, S. 1–13, doi: 10.1145/3447526.3472031.

[212] C. F. Liew und T. Yairi, „Companion Unmanned Aerial Vehicles: A Survey“, 14. Jan. 2020. [Online]. Verfügbar unter: http://arxiv.org/pdf/2001.04637v1.

[213] K. Pfeil, S. L. Koh und J. LaViola, „Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles“ in *Proceedings of the 2013 international conference on Intelligent user interfaces*, Santa Monica, California, USA, 2013, S. 257–266, doi: 10.1145/2449396.2449429.

[214] J. Nagi, A. Giusti, G. A. Di Caro und L. M. Gambardella, „HRI in the sky: controlling UAVs using face poses and hand gestures“ in *HRI*, 2014, S. 252–253.

[215] M. Obaid, O. Mubin, C. A. Basedow, A. A. Ünlüer, M. J. Bergström und M. Fjeld, „A Drone Agent to Support a Clean Environment“ in *HAI 2015: The Third International Conference on Human-Agent Interaction*, Daegu Kyungpook Republic of Korea, 2015, S. 55–61, doi: 10.1145/2814940.2814947.

[216] M. Monajjemi, J. Bruce, S. A. Sadat, J. Wawerla und R. Vaughan, „UAV, do you see me?: Establishing mutual attention between an uninstrumented human and an outdoor UAV in flight“ in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, Hamburg, Germany, 2015, S. 3614–3620, doi: 10.1109/IROS.2015.7353882.

[217] J. R. Cauchard *et al.*, „Drone.io: A Gestural and Visual Interface for Human-Drone Interaction“ in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Daegu, Korea (South), 2019, S. 153–162, doi: 10.1109/HRI.2019.8673011.

[218] D. Palossi *et al.*, „Fully Onboard AI-powered Human-Drone Pose Estimation on Ultra-low Power Autonomous Flying Nano-UAVs“, 19. März 2021. [Online]. Verfügbar unter: http://arxiv.org/pdf/2103.10873v1.

[219] A. Schelle und P. Stütz, „Visual Communication with UAS: Estimating Parameters for Gestural Transmission of Task Descriptions“ in *Lecture Notes in Networks and Systems, Advances in Human Factors in Robots, Unmanned Systems and Cybersecurity*, M. Zallio, C. Raymundo Ibañez und J. H. Hernandez, Hg., Cham: Springer International Publishing, 2021, S. 16–24, doi: 10.1007/978-3-030-79997-7_3.

[220] J. Triesch und C. von der Malsburg, „Robotic Gesture Recognition by Cue Combination“ in *Informatik aktuell, Informatik '98*, W. Brauer, J. Dassow und R. Kruse, Hg., Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, S. 223–232, doi: 10.1007/978-3-642-72283-7_22.

[221] J. Shotton *et al.*, „Real-time human pose recognition in parts from single depth images“, *Communications of the ACM*, Jg. 56, Nr. 1, S. 116–124, 2013.

[222] M. G. Jacob, J. P. Wachs und R. A. Packer, „Hand-gesture-based sterile interface for the operating room using contextual cues for the navigation of radiological images“, *Journal of the American Medical Informatics Association : JAMIA*, Jg. 20, e1, e183-6, 2013, doi: 10.1136/amiajnl-2012-001212.

[223] T. Naseer, J. Sturm und D. Cremers, „Followme: Person following and gesture recognition with a quadrocopter“ in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, S. 624–630.

[224] Y. W. Kuan, N. O. Ee und L. S. Wei, „Comparative Study of Intel R200, Kinect v2, and Primesense RGB-D Sensors Performance Outdoors“, *IEEE Sensors J.*, Jg. 19, Nr. 19, S. 8741–8750, 2019, doi: 10.1109/JSEN.2019.2920976.

[225] S. Achar, J. R. Bartels, W. L. '. Whittaker, K. N. Kutulakos und S. G. Narasimhan, „Epipolar time-of-flight imaging“, *0730-0301*, Jg. 36, Nr. 4, S. 1–8, 2017, doi: 10.1145/3072959.3073686.

[226] L. E. Ortiz, E. V. Cabrera und L. M. Goncalves, „Depth Data Error Modeling of the ZED 3D Vision Sensor from Stereolabs“ in *ELCVIA: electronic letters on computer vision and image analysis*, 2018, S. 1–15.

[227] O. Alon, S. Rabinovich, C. Fyodorov und J. R. Cauchard, „First Step toward Gestural Recognition in Harsh Environments“, *Sensors (Basel, Switzerland)*, Jg. 21, Nr. 12, 2021, doi: 10.3390/s21123997.

[228] W. Gong *et al.*, „Human Pose Estimation from Monocular Images: A Comprehensive Survey“, *Sensors (Basel, Switzerland)*, Jg. 16, Nr. 12, 2016, doi: 10.3390/s16121966.

[229] Y. Chen, Y. Tian und M. He, „Monocular human pose estimation: A survey of deep learning-based methods“, *Computer Vision and Image Understanding*, Jg. 192, 2020, doi: 10.1016/j.cviu.2019.102897.

[230] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers und J. Davis, „SCAPE“ in *ACM SIGGRAPH 2005 Papers*, Los Angeles, California, 2005, S. 408, doi: 10.1145/1186822.1073207.

[231] P. F. Felzenszwalb und D. P. Huttenlocher, „Pictorial Structures for Object Recognition“, *International Journal of Computer Vision*, Jg. 61, Nr. 1, S. 55–79, 2005, doi: 10.1023/B:VISI.0000042934.15159.49.

[232] S. Zuffi, O. Freifeld und M. J. Black, „From Pictorial Structures to deformable structures“ in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, 2012, S. 3546–3553, doi: 10.1109/CVPR.2012.6248098.

[233] A. Mittal, L. Zhao und L. S. Davis, „Human body pose estimation using silhouette shape analysis“ in *IEEE Conference on Advanced Video and Signal Based Surveillance. AVSS 2003*, Miami, FL, USA, 2003, S. 263–270, doi: 10.1109/AVSS.2003.1217930.

[234] N. Dalal und B. Triggs, „Histograms of Oriented Gradients for Human Detection“ in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, San Diego, CA, USA, 2005, S. 886–893, doi: 10.1109/CVPR.2005.177.

[235] J. Redmon, S. Divvala, R. Girshick und A. Farhadi, „You Only Look Once: Unified, Real-Time Object Detection“ in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[236] T. Baltrusaitis, P. Robinson und L.-P. Morency, „Constrained Local Neural Fields for Robust Facial Landmark Detection in the Wild“ in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, 2013.

[237] T. Ahonen, A. Hadid und M. Pietikäinen, „Face description with local binary patterns: application to face recognition“, *IEEE Trans. Pattern Anal. Mach. Intell.*, Jg. 28, Nr. 12, S. 2037–2041, 2006, doi: 10.1109/TPAMI.2006.244.

[238] G. Costante, E. Bellocchio, P. Valigi und E. Ricci, „Personalizing vision-based gestural interfaces for HRI with UAVs: a transfer learning approach“ in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, S. 3319–3326, doi: 10.1109/IROS.2014.6943024.

[239] E. W. Taylor und S. J. Lent, „BBC Test Card No. 61 (Flesh Tone Reference): Colorimetric and other Optical Considerations“, *J SMPTE*, Jg. 87, Nr. 2, S. 76–78, 1978, doi: 10.5594/J17432.

[240] A. Schelle und P. Stütz, „Visual Communication with UAS: Recognizing Gestures from an Airborne Platform“ in *Lecture Notes in Computer Science, Virtual, Augmented and Mixed Reality*, S. Lackey und J. Chen, Hg., Cham: Springer International Publishing, 2017, S. 173–184, doi: 10.1007/978-3-319-57987-0_14.

[241] S. Handrich, O. Rashid und A. Al-Hamadi, „Non-intrusive Gesture Recognition in Real Companion Environments“ in *Cognitive technologies, Companion technology: A paradigm shift in human-technology interaction*, S. Biundo-Stephan und A. Wendemuth, Hg., Cham: Springer, 2017, S. 321–343, doi: 10.1007/978-3-319-43665-4_16.

[242] T. L. Munea, Y. Z. Jembre, H. T. Weldegebriel, L. Chen, C. Huang und C. Yang, „The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation“, *IEEE Access*, Jg. 8, S. 133330–133348, 2020, doi: 10.1109/ACCESS.2020.3010248.

[243] S. Johnson und M. Everingham, „Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation“ in *Proceedings of the British Machine Vision Conference*, 2010, doi: 10.5244/C.24.12.

[244] T.-Y. Lin *et al.*, „Microsoft COCO: Common Objects in Context“ in *Lecture Notes in Computer Science, Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele und T. Tuytelaars, Hg., Cham: Springer International Publishing, 2014, S. 740–755, doi: 10.1007/978-3-319-10602-1_48.

[245] M. Andriluka, L. Pishchulin, P. Gehler und B. Schiele, „2D Human Pose Estimation: New Benchmark and State of the Art Analysis“ in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, S. 3686–3693, doi: 10.1109/CVPR.2014.471.

[246] C. Ionescu, D. Papava, V. Olaru und C. Sminchisescu, „Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments“, *IEEE transactions on pattern analysis and machine intelligence*, Jg. 36, Nr. 7, S. 1325–1339, 2014, doi: 10.1109/TPAMI.2013.248.

[247] A. Toshev und C. Szegedy, „DeepPose: Human Pose Estimation via Deep Neural Networks“, S. 1653–1660, 2014, doi: 10.1109/CVPR.2014.214.

[248] S.-E. Wei, V. Ramakrishna, T. Kanade und Y. Sheikh, „Convolutional pose machines“ in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, S. 4724–4732.

[249] Z. Cao, T. Simon, S.-E. Wei und Y. Sheikh, „Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields“ in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[250] H. W. Kuhn, „The Hungarian method for the assignment problem“, *Naval Research Logistics*, Jg. 2, 1-2, S. 83–97, 1955, doi: 10.1002/nav.3800020109.

[251] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei und Y. Sheikh, „OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields“, 18. Dez. 2018. [Online]. Verfügbar unter: http://arxiv.org/pdf/1812.08008v1.

[252] D. Groos, H. Ramampiaro und E. af Ihlen, „EfficientPose: Scalable single-person pose estimation“, *Appl Intell*, 2020, doi: 10.1007/s10489-020-01918-7.

[253] J. Li, C. Wang, H. Zhu, Y. Mao, H.-S. Fang und C. Lu, „CrowdPose: Efficient Crowded Scenes Pose Estimation and a New Benchmark“ in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[254] S. Kreiss, L. Bertoni und A. Alahi, „PifPaf: Composite Fields for Human Pose Estimation“, 15. März 2019. [Online]. Verfügbar unter: http://arxiv.org/pdf/1903.06593v2.

[255] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang und M. Grundmann, „BlazePose: On-device Real-time Body Pose tracking“, 18. Juni 2020. [Online]. Verfügbar unter: http://arxiv.org/pdf/2006.10204v1.

[256] C. Liu und T. Szirányi, „Real-Time Human Detection and Gesture Recognition for On-Board UAV Rescue“, *Sensors*, Jg. 21, Nr. 6, S. 2180, 2021, doi: 10.3390/s21062180.

[257] T. Mantecón, C. R. Del Blanco, F. Jaureguizar und N. García, „New generation of human machine interfaces for controlling UAV through depth-based gesture recognition“ in *SPIE Defense + Security*, Baltimore, Maryland, USA, 2014, 90840C, doi: 10.1117/12.2053244.

[258] M. A. Kassab, M. Ahmed, A. Maher und B. Zhang, „Real-Time Human-UAV Interaction: New Dataset and Two Novel Gesture-Based Interacting Systems“, *IEEE Access*, Jg. 8, S. 195030–195045, 2020, doi: 10.1109/ACCESS.2020.3033157.

[259] O. Kopuklu, A. Gunduz, N. Kose und G. Rigoll, „Online Dynamic Hand Gesture Recognition Including Efficiency Analysis“, *IEEE Trans. Biom. Behav. Identity Sci.*, Jg. 2, Nr. 2, S. 85–97, 2020, doi: 10.1109/TBIOM.2020.2968216.

[260] T. Baudel und M. Beaudouin-Lafon, „Charade: remote control of objects using free-hand gestures“, *Commun. ACM*, Jg. 36, Nr. 7, S. 28–35, 1993, doi: 10.1145/159544.159562.

[261] S. Escalera, V. Athitsos und I. Guyon, „Challenges in multimodal gesture recognition“ in *Journal of Machine Learning Research*, 2016, S. 1–54.

[262] T. Hachaj und M. R. Ogiela, „Rule-based approach to recognizing human body poses and gestures in real time“, *Multimedia Systems*, Jg. 20, Nr. 1, S. 81–99, 2013, doi: 10.1007/s00530-013-0332-2.

[263] J.-C. Lementec und P. Bajcsy, „Recognition of arm gestures using multiple orientation sensors: gesture classification“ in *The 7th International IEEE Conference on Intelligent Transportation Systems*, Washington, WA, USA, 3-6 Oct. 2004, S. 965–970, doi: 10.1109/ITSC.2004.1399037.

[264] L. Venetsky und J. W. Tieman, “Robotic gesture recognition system,” US7606411 B2. United States of America 11/586,750, Okt 20, 2009.

[265] L.-P. Morency, A. Quattoni und T. Darrell, „Latent-Dynamic Discriminative Models for Continuous Gesture Recognition“ in *Saliency detection: A spectral residual approach*, 2007, S. 1–8, doi: 10.1109/CVPR.2007.383299.

[266] C. Choi, J. Ahn und H. Byun, „Visual recognition of aircraft marshalling signals using gesture phase analysis“ in *IEEE Intelligent Vehicles Symposium*, 2008, S. 853–858, doi: 10.1109/IVS.2008.4621186.

[267] L. Bergroth, H. Hakonen und T. Raita, „A survey of longest common subsequence algorithms“ in *SPIRE'2000 - String Processing and Information Retrieval*, A Curuna, Spain, 2000, S. 39–48, doi: 10.1109/SPIRE.2000.878178.

[268] R. P. Higgins, “Method, Apparatus and Computer Program Product for Recognizing a Gesture,” US 8,194,925 B2. USA, Jun 5, 2012.

[269] C. Kawatsu *et al.*, „Gesture recognition for robotic control using deep learning“ in *NDIA Ground Vehicle Systems Engineering and Technology Symposium*, 2017.

[270] K. Natarajan, T.-H. D. Nguyen und M. Mete, „Hand Gesture Controlled Drones: An Open Source Library“ in *1st International Conference on Data Intelligence and Security (ICDIS)*, South Padre Island, TX, 2018, S. 168–175, doi: 10.1109/ICDIS.2018.00035.

[271] L. Zhou, C. Du, Z. Sun, T. L. Lam und Y. Xu, „Long-Range Hand Gesture Recognition via Attention-based SSD Network“ in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, 2021, S. 1832–1838, doi: 10.1109/ICRA48506.2021.9561189.

[272] A. Kendon, *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press, 2004.

[273] J. Urakami, „Developing and testing a human-based gesture vocabulary for tabletop systems“, *Human factors*, Jg. 54, Nr. 4, S. 636–653, 2012, doi: 10.1177/0018720811433052.

[274] Y. Song, D. Demirdjian und R. Davis, „Tracking body and hands for gesture recognition: NATOPS aircraft handling signals database“ in *Gesture Recognition (FG 2011)*, Santa Barbara, CA, USA, 2011, S. 500–506, doi: 10.1109/FG.2011.5771448.

[275] T. Starner, J. Weaver und A. Pentland, „Real-time american sign language recognition using desk and wearable computer based video“, *IEEE Trans. Pattern Anal. Mach. Intell.*, Jg. 20, Nr. 12, S. 1371–1375, 1998.

[276] V. E. Balas *et al., Hg., Indian Sign Language Recognition Using Optimized Neural Networks: Information Technology and Intelligent Transportation Systems*. Springer International Publishing: Springer International Publishing, 2017.

[277] O. Koller, J. Forster und H. Ney, „Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers“, *Computer Vision and Image Understanding*, Jg. 141, S. 108–125, 2015.

[278] U. von Agris, J. Zieren, U. Canzler, B. Bauer und K.-F. Kraiss, „Recent developments in visual sign language recognition“, *Univ Access Inf Soc*, Jg. 6, Nr. 4, S. 323–362, 2008, doi: 10.1007/s10209-007-0104-x.

[279] N. Gio, R. Brisco und T. Vuletic, „CONTROL OF A DRONE WITH BODY GESTURES“, *Proc. Des. Soc.*, Jg. 1, S. 761–770, 2021, doi: 10.1017/pds.2021.76.

[280] F. Taralle, A. Paljic, S. Manitsaris, J. Grenier und C. Guettier, „A Consensual and Non-ambiguous Set of Gestures to Interact with UAV in Infantrymen“ in *CHI '15: CHI Conference on Human Factors in Computing Systems*, Seoul Republic of Korea, 2015, S. 797–803, doi: 10.1145/2702613.2702971.

[281] W. Wang, X. Qin, C. Zheng, H. Wang, J. Li und N. Junlong, „Comfort Evaluation Model for Gesture Motion based on Mechanical Energy Expenditure and Efficiency“, *Procedia CIRP*, Jg. 78, S. 13–18, 2018, doi: 10.1016/j.procir.2018.08.323.

[282] M. Obaid, F. Kistler, M. Häring, R. Bühling und E. André, „A Framework for User-Defined Body Gestures to Control a Humanoid Robot“, *International Journal of Social Robotics*, Jg. 6, Nr. 3, S. 383–396, 2014, doi: 10.1007/s12369-014-0233-3.

[283] W. B. Rouse und N. M. Morris, „On looking into the black box: Prospects and limits in the search for mental models: Technical Report 85-2“, 1985.

[284] E. Peshkova, M. Hitz und B. Kaufmann, „Natural Interaction Techniques for an Unmanned Aerial Vehicle System“, *IEEE Pervasive Comput.*, Jg. 16, Nr. 1, S. 34–42, 2017, doi: 10.1109/MPRV.2017.3.

[285] A. G. Perera, Y. W. Law, J. Chahl, L. Leal-Taixé und S. Roth, „UAV-GESTURE: A Dataset for UAV Control and Gesture Recognition“ in *Image Processing, Computer Vision, Pattern Recognition, and Graphics*, Bd. 11130, *Computer Vision - ECCV 2018 Workshops: Munich, Germany, September 8-14, 2018, Proceedings, Part II*, L. Leal-Taixé und S. Roth, Hg., Cham: Springer International Publishing, 2019, S. 117–128.

[286] C. Schuldt, I. Laptev und B. Caputo, „Recognizing human actions: a local SVM approach“ in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*, Cambridge, UK, 2004, 32-36 Vol.3, doi: 10.1109/ICPR.2004.1334462.

[287] Z. Lin, Z. Jiang und L. S. Davis, „Recognizing actions by shape-motion prototype trees“ in *2009 IEEE 12th International Conference on Computer Vision (ICCV)*, Kyoto, 2009, S. 444–451, doi: 10.1109/ICCV.2009.5459184.

[288] Y. Song, D. Demirdjian und R. Davis, „Tracking Body and Hands For Gesture Recognition: NATOPS Aircraft Handling Signals Database“ in *Proceedings of the 9th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2011)*, Santa Barbara, CA, 2011.

[289] I. Guyon, V. Athitsos, P. Jangyodsuk, B. Hamner und H. J. Escalante, „ChaLearn gesture challenge: Design and first results“ in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, Providence, RI, USA, 2012, S. 1–6, doi: 10.1109/CVPRW.2012.6239178.

[290] M. Sharma, D. Hildebrandt, G. Newman, J. E. Young und R. Eskicioglu, „Communicating affect via flight path Exploring use of the Laban Effort System for designing affective locomotion paths“ in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Tokyo, Japan, 2013, S. 293–300, doi: 10.1109/HRI.2013.6483602.

[291] D. Szafir, B. Mutlu und T. Fong, „Communication of intent in assistive free flyers“ in *HRI'14: ACM/IEEE International Conference on Human-Robot Interaction*, Bielefeld Germany, 2014, S. 358–365, doi: 10.1145/2559636.2559672.

[292] W. Jensen, S. Hansen und H. Knoche, „Knowing You, Seeing Me“ in *the 2018 CHI Conference*, Montreal QC, Canada, 2018, S. 1–12, doi: 10.1145/3173574.3173939.

[293] J. R. Cauchard, K. Y. Zhai, M. Spadafora und J. A. Landay, „Emotion encoding in Human-Drone Interaction“ in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Christchurch, New Zealand, 2016, S. 263–270, doi: 10.1109/HRI.2016.7451761.

[294] D. Szafir, B. Mutlu und T. Fong, „Communicating Directionality in Flying Robots“ in *HRI '15: ACM/IEEE International Conference on Human-Robot Interaction*, Portland Oregon USA, 2015, S. 19–26, doi: 10.1145/2696454.2696475.

[295] S. Schneegass, F. Alt, J. Scheible, A. Schmidt und H. Su, „Midair displays“ in *CHI '14: CHI Conference on Human Factors in Computing Systems*, Toronto Ontario Canada, 2014, S. 2035–2040, doi: 10.1145/2559206.2581190.

[296] K. Beomshik und Yongin-si, “Flying Display Device,” US 9,891,885 B2. USA.

[297] Revell Control, *Texter Helicopter [23976].* [Online]. Verfügbar unter: https://www.youtube.com/watch?v=kg4h-scQuc8 (Zugriff am: 9. Februar 2021).

[298] M. Benadik, *RGB-Blades.* [Online]. Verfügbar unter: http://www.nightmagicblades.com/wp-content/gallery/new-rgb-blades/img_3548-copy.jpg (Zugriff am: 13. Mai 2019).

[299] A. M. Brock *et al.*, „FlyMap“ in *PerDis '18: The International Symposium on Pervasive Displays*, 2018, S. 1–9, doi: 10.1145/3205873.3205877.

[300] J. Scheible und M. Funk, „In-situ-displaydrone“ in *PerDis '16: The International Symposium on Pervasive Displays*, Oulu Finland, 2016, S. 251–252, doi: 10.1145/2914920.2940334.

[301] „Spaxels, Pixels in Space - A Novel Mode of Spatial Display“ in *International Conference on Signal Processing and Multimedia Applications*, Rome, Italy, 2012, S. 19–24, doi: 10.5220/0004126400190024.

[302] Guinness World Records Limited, *Most remote controlled multirotors/drones airborne simultaneously*. Shenzhen, 2021. [Online]. Verfügbar unter: https://www.guinnessworldrecords.com/world-records/373319-most unmanned-aerial vehicles uavs-airborne simultaneously 5-kg-or-less

[303] J. P. Wachs, M. Kölsch, H. Stern und Y. Edan, „Vision-based hand-gesture applications“, *Commun. ACM*, Jg. 54, Nr. 2, S. 60–71, 2011, doi: 10.1145/1897816.1897838.

[304] G. Sperling, M. Landy, Y. Cohen und M. Pavel, „Intelligible encoding of ASL image sequences at extremely low information rates“, *Computer Vision, Graphics, and Image Processing*, Jg. 31, Nr. 3, S. 335–391, 1985, doi: 10.1016/0734-189X(85)90034-9.

[305] G. Fischer, „User Modeling in Human–Computer Interaction“, *User Modeling and User-Adapted Interaction*, Jg. 11, Nr. 1, S. 65–86, 2001, doi: 10.1023/A:1011145532042.

[306] A. Couture-Beil, R. T. Vaughan und G. Mori, „Selecting and commanding individual robots in a vision-based multi-robot system“ in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Osaka, Japan, 2010, S. 355–356, doi: 10.1109/HRI.2010.5453167.

[307] A. Godil, P. Grother und S. Ressler, „Human identification from body shape“ in *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, 2003, S. 386–392.

[308] J. Wu, J. Konrad und P. Ishwar, „Dynamic time warping for gesture-based user identification and authentication with Kinect“ in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, 2013, S. 2371–2375, doi: 10.1109/ICASSP.2013.6638079.

[309] Y. Kortli, M. Jridi, A. A. Falou und M. Atri, „Face Recognition Systems: A Survey“, *Sensors (Basel, Switzerland)*, Jg. 20, Nr. 2, 2020, doi: 10.3390/s20020342.

[310] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas und M. J. Marín-Jiménez, „Automatic generation and detection of highly reliable fiducial markers under occlusion“, *Pattern Recognition*, Jg. 47, Nr. 6, S. 2280–2292, 2014, doi: 10.1016/j.patcog.2014.01.005.

[311] Q. Jia, W. D. Xu, J. H. Hu, J. Liu, X. Yang und L. Y. Zhu, „Design and evaluation of digital camouflage pattern by spot combination“, *Multimedia Tools and Applications*, Jg. 79, 29-30, S. 22047–22064, 2020, doi: 10.1007/s11042-020-09002-5.

[312] W. Hubig und D. Weber, „Techniken zur Leistungsanalyse von Kommunikationssystemen — Kurzfassung —“ in *Informatik-Fachberichte, Messung, Modellierung und Bewertung von Rechensystemen und Netzen*, W. Brauer, G. Stiege und J. S. Lie, Hg., Berlin, Heidelberg: Springer Berlin Heidelberg, 1989, S. 255–257, doi: 10.1007/978-3-642-75079-3_18.

[313] P. Scholz, *Softwareentwicklung eingebetteter Systeme*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[314] E. Yang und M. C. Dorneich, „The Effect of Time Delay on Emotion, Arousal, and Satisfaction in Human-Robot Interaction“, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Jg. 59, Nr. 1, S. 443–447, 2015, doi: 10.1177/1541931215591094.

[315] R. B. Miller, „Response time in man-computer conversational transactions“ in *the December 9-11, 1968, fall joint computer conference, part I*, San Francisco, California, 1968, S. 267, doi: 10.1145/1476589.1476628.

[316] W. Wesemann, U. Schiefer und M. Bach, „Neue DIN-Normen zur Sehschärfebestimmung“ (ger), *Der Ophthalmologe : Zeitschrift der Deutschen Ophthalmologischen Gesellschaft*, Jg. 107, Nr. 9, S. 821–826, 2010, doi: 10.1007/s00347-010-2228-2.

[317] H. Kaufmann, *Strabismus*. Stuttgart: Georg Thieme Verlag, 2004.

[318] J. Nielsen, *Usability engineering*. Boston: Acad. Press, 1993.

[319] D. Rempel, M. J. Camilleri und D. L. Lee, „The Design of Hand Gestures for Human-Computer Interaction: Lessons from Sign Language Interpreters“, *International journal of human-computer studies*, Jg. 72, 10-11, S. 728–735, 2015, doi: 10.1016/j.ijhcs.2014.05.003.

[320] H. I. STERN, J. P. WACHS und Y. Edan, „Designing hand gesture vocabularies for natural interaction by combining psycho-physiological and recognition factors“, *Int. J. Semantic Computing*, Jg. 02, Nr. 01, S. 137–160, 2008, doi: 10.1142/S1793351X08000385.

[321] S. Rath und J. Y. J. Chow, „Air Taxi Skyport Location Problem for Airport Access“, 1. Apr. 2019. [Online]. Verfügbar unter: http://arxiv.org/pdf/1904.01497v4.

[322] J. Yapp, R. Seker und R. Babiceanu, „UAV as a service: Enabling on-demand access and on-the-fly re-tasking of multi-tenant UAVs using cloud services“ in *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, Sacramento, CA, USA, 2016, S. 1–8, doi: 10.1109/DASC.2016.7778007.

[323] H. König, *Protocol Engineering: Prinzip, Beschreibung und Entwicklung von Kommunikationsprotokollen,* 1. Aufl. Stuttgart: Teubner, 2003.

[324] C. E. Shannon, „A Mathematical Theory of Communication“, *Bell System Technical Journal*, Jg. 27, Nr. 3, S. 379–423, 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x.

[325] *ISO/IEC 7498-1:1994: Information technology — Open Systems Interconnection — Basic Reference Model: The Basic Model,* International Organization for Standardization, Berlin, Nov. 1994.

[326] R. Sharp, *Principles of Protocol Design*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[327] J. Nagi, A. Giusti, L. M. Gambardella und G. A. Di Caro, „Human-swarm interaction using spatial gestures“ in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, S. 3834–3841, doi: 10.1109/IROS.2014.6943101.

[328] M. Herczeg, *Software-Ergonomie: Theorien, Modelle und Kriterien für gebrauchstaugliche interaktive Computersysteme,* 4. Aufl. Berlin, Boston: De Gruyter Oldenbourg, 2018. [Online]. Verfügbar unter: http://www.degruyter.com/search?f_0=isbnissn&q_0=9783110446852&searchTitles=true

[329] F. M. Noori, B. Wallace, M. Z. Uddin und J. Torresen, „A Robust Human Activity Recognition Approach Using OpenPose, Motion Features, and Deep Recurrent Neural Network“ in *Image Analysis*, Cham, 2019, S. 299–310.

[330] P. Collett, *Ich sehe was, was du nicht sagst: So deuten Sie die Gesten der anderen - und wissen, was diese wirklich denken,* 1. Aufl. Bergisch Gladbach: Ehrenwirth, 2004.

[331] A. Coucke *et al.,* „Snips Voice Platform: an embedded Spoken Language Understanding system for private-by-design voice interfaces“, 25. Mai 2018. [Online]. Verfügbar unter: http://arxiv.org/pdf/1805.10190v3.

[332] C. A. Ferguson, „‘Foreigner Talk’ as the Name of a Simplified Register“, *International Journal of the Sociology of Language*, Jg. 1981, Nr. 28, 1981, doi: 10.1515/ijsl.1981.28.9.

[333] K. Müller, *Lernen im Dialog: Gestaltlinguistische Aspekte des Zweitspracherwerbs*. Tübingen: Gunter Narr Verlag, 2000.

[334] J. Roche, „Variation in Xenolects (Foreigner Talk)“, *Sociolinguistica*, Nr. 12, S. 117–139, 1998, doi: 10.5282/ubm/epub.14134.

[335] E. Loper und S. Bird, „NLTK: The Natural Language Toolkit“, 17. Mai 2002. [Online]. Verfügbar unter: http://arxiv.org/pdf/cs/0205028v1.

[336] M. Quigley *et al.*, „ROS: An open-source Robot Operating System“ in *ICRA Workshop on Open Source Software*, 2009.

[337] *OpenPose: Real-time multi-person keypoint detection library for body, face and hands estimation*. https://github.com/CMU-Perceptual-Computing-Lab/openpose. Perceptual Computing Laboratory, Carnegie Mellon University, 2018. [Online]. Verfügbar unter: https://github.com/CMU-Perceptual-Computing-Lab/openpose

[338] C. Szegedy *et al.*, „Intriguing properties of neural networks“, 21. Dez. 2013. [Online]. Verfügbar unter: http://arxiv.org/pdf/1312.6199v4.

[339] I. J. Goodfellow, J. Shlens und C. Szegedy, „Explaining and Harnessing Adversarial Examples“, 20. Dez. 2014. [Online]. Verfügbar unter: http://arxiv.org/pdf/1412.6572v3.

[340] K. Xu *et al.*, „Adversarial T-Shirt! Evading Person Detectors in a Physical World“ in *Lecture Notes in Computer Science, Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox und J.-M. Frahm, Hg., Cham: Springer International Publishing, 2020, S. 665–681, doi: 10.1007/978-3-030-58558-7_39.

[341] J. P. Vox und F. Wallhoff, Hg., *Preprocessing and Normalization of 3D-Skeleton-Data for Human Motion Recognition*. IEEE: IEEE, 2018.

[342] G. Bradski, „The OpenCV Library“, *Dr. Dobb’s Journal of Software Tools*, 2000.

[343] C.-C. Chang und C.-J. Lin, „LIBSVM“, *ACM Trans. Intell. Syst. Technol.*, Jg. 2, Nr. 3, S. 1–27, 2011, doi: 10.1145/1961189.1961199.

[344] L. Wang, X. Ding, C. Fang, C. Liu und K. Wang, „Eye blink detection based on eye contour extraction“ in *IS&T/SPIE Electronic Imaging*, San Jose, CA, 2009, 72450R, doi: 10.1117/12.804916.

[345] D. E. King, „Dlib-ml:: A Machine Learning Toolkit“, *Journal of Machine Learning Research*, Nr. 10, S. 1755–1758, 2009.

[346] V. Kazemi und J. Sullivan, „One millisecond face alignment with an ensemble of regression trees“ in *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, 2014, S. 1867–1874, doi: 10.1109/CVPR.2014.241.

[347] N. E. Gillian und J. A. Paradiso, „The gesture recognition toolkit“, *Journal of Machine Learning Research*, Jg. 15, Nr. 1, S. 3483–3487, 2014.

[348] P. Jaccard, „Étude comparative de la distribution florale dans une portion des Alpes et du Jura“, 1901, doi: 10.5169/seals-266450.

[349] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto und E. A. B. Da Silva, „A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit“, *Electronics*, Jg. 10, Nr. 3, S. 279, 2021, doi: 10.3390/electronics10030279.

[350] A. Bochkovskiy, C.-Y. Wang und H.-Y. M. Liao, „YOLOv4: Optimal Speed and Accuracy of Object Detection“, 23. Apr. 2020. [Online]. Verfügbar unter: http://arxiv.org/pdf/2004.10934v1.

[351] *FastLED*. 3.1. https://github.com/FastLED/FastLED, 2015. [Online]. Verfügbar unter: https://github.com/FastLED/FastLED

[352] F. Güntner, *Bergunfälle 2020: So wenige Tote wie noch nie*, 2021. [Online]. Verfügbar unter: https://www.alpenverein.de/der-dav/presse/pressemeldungen/dav-bergunfallstatistik-2020_aid_37146.html

[353] M. Klier, E. R. Wanka-Pail, C. Gehring und S. Prueckner, „Luftrettung rund um die Uhr – Welchen Einfluss hat das Wetter?“, *Notfall Rettungsmed*, Jg. 18, Nr. 2, S. 130–138, 2015, doi: 10.1007/s10049-015-1987-3.

[354] K. Datar, M. N. Gandhi, P. Aggarwal und M. Sohani, „A Review on Deep Learning Based Lip-Reading“, *IJSRCSEIT*, S. 182–188, 2020, CSEIT206140, doi: 10.32628/CSEIT206140.

[355] P. F. Christ, F. Lachner, A. Hösl, B. Menze, K. Diepold und A. Butz, „Human-Drone-Interaction: A Case Study to Investigate the Relation Between Autonomy and User Experience“ in *Lecture Notes in Computer Science, Computer Vision – ECCV 2016 Workshops*, G. Hua und H. Jégou, Hg., Cham: Springer International Publishing, 2016, S. 238–253, doi: 10.1007/978-3-319-48881-3_17.

[356] I. S. MacKenzie, *Human-computer interaction: An empirical research perspective*. Amsterdam, Boston, Heidelberg, London, New York: Elsevier Morgan, 2012.

[357] S. G. Hart, „Nasa-Task Load Index (NASA-TLX); 20 Years Later“, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Jg. 50, Nr. 9, S. 904–908, 2006, doi: 10.1177/154193120605000909.

[358] Institut für Industrial Design, Hochschule Magdeburg-Stendal, *NASA-TLX (Kurzfassung deutsch).* [Online]. Verfügbar unter: http://interaction-design-group.de/toolbox/wp-content/uploads/2016/05/NASA-TLX.pdf (Zugriff am: 9. September 2021).

[359] J. Deriu *et al.*, „Survey on Evaluation Methods for Dialogue Systems“, *Artif Intell Rev*, Jg. 54, Nr. 1, S. 755–810, 2021, doi: 10.1007/s10462-020-09866-x.

[360] *Ergonomie der Mensch-System-Interaktion - Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte,* DIN EN ISO 9241-11:2018-11, Berlin. [Online]. Verfügbar unter: https://dx.doi.org/10.31030/2757945

[361] J. Cohen, „A power primer“, *Psychological Bulletin*, Jg. 112, Nr. 1, S. 155–159, 1992, doi: 10.1037/0033-2909.112.1.155.

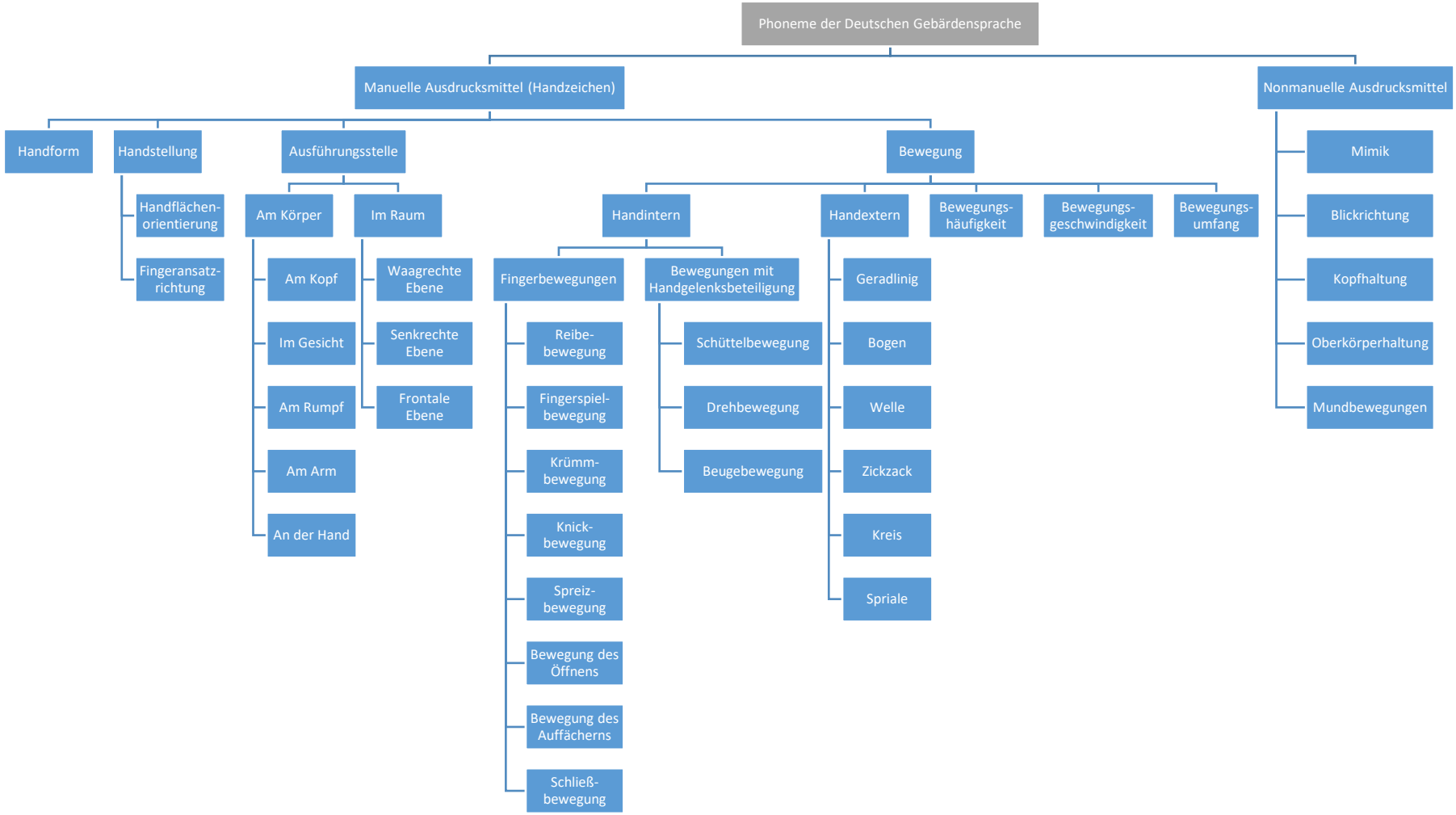
[362] K. J. Vicente und J. Rasmussen, „Ecological interface design: theoretical foundations“, *IEEE Trans. Syst., Man, Cybern.*, Jg. 22, Nr. 4, S. 589–606, 1992, doi: 10.1109/21.156574.

[363] J. Reason, *Human error.* Cambridge, UK: Cambridge University Press, 1990.

[364] A. Kluge, „Grundsätzliche Überlegungen für die Planung von Evaluationsmaßnahmen im eigenen Projekt-was, wie und wann evaluieren?“ in *Kognitive Systeme Workshop*, Duisburg, 2012.

Anhang

A1 Phoneme der Deutschen Gebärdensprache nach [99]



A2 Klassifikationsleistung des Merkmalbeschreibers für statische Armstellungen

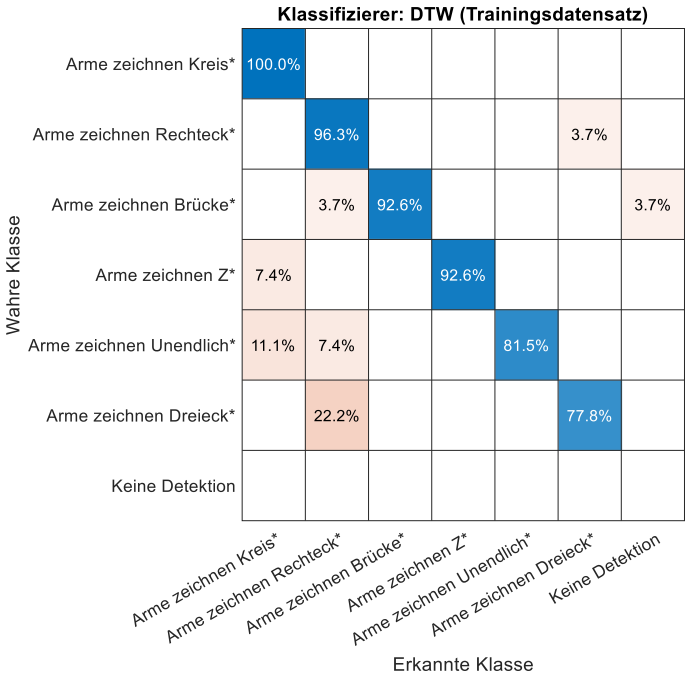
Klassifizierer: SVM (Trainingsdatensatz)

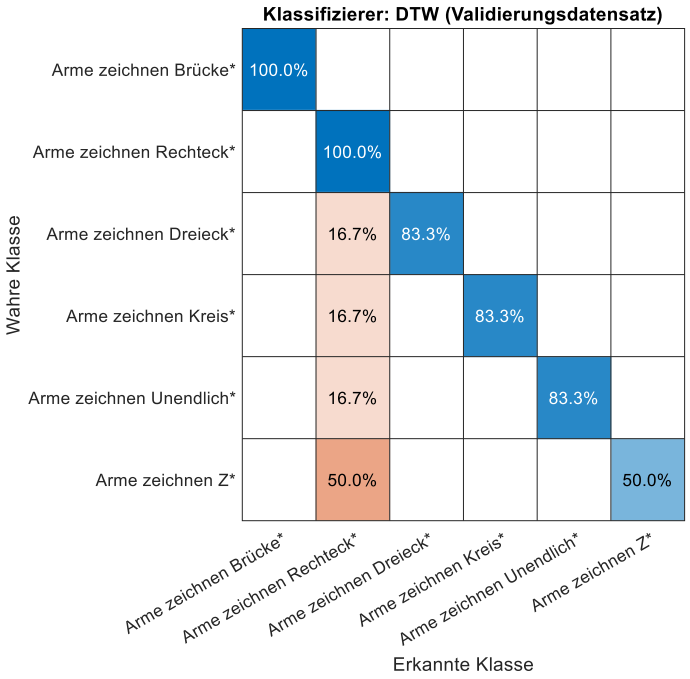
Wahre Klasse \ Erkannte Klasse	Arm zeigt auf UAV	Arm zeigt auf Brust	Arm zeigt L-förmig nach oben	Arm zeigt auf Ohr	Arm gekreuzt unter Hüfte	Arm zeigt auf Gesicht	Arm zeigt L-förmig nach außen	Arm gekreuzt über Kopf	Arm zeigt auf Hals	Arm zeigt nach oben	Arm zeigt nach außen	Arm hängend	Arm gekreuzt vor Brust	Arm zeigt auf Knie	Arm zeigt nach unten	Keine Detektion
Arm zeigt auf UAV	99.2%								0.4%							0.3%
Arm zeigt auf Brust		99.2%									0.4%					0.4%
Arm zeigt L-förmig nach oben			99.2%				0.4%									0.4%
Arm zeigt auf Ohr		0.4%		99.2%												0.4%
Arm gekreuzt unter Hüfte					99.2%			0.4%								0.4%
Arm zeigt auf Gesicht				0.5%		99.2%										0.4%
Arm zeigt L-förmig nach außen							99.2%								0.5%	0.4%
Arm gekreuzt über Kopf								99.1%								0.9%
Arm zeigt auf Hals									99.1%					0.5%		0.4%
Arm zeigt nach oben	0.5%									99.1%						0.4%
Arm zeigt nach außen			0.5%								99.1%					0.4%
Arm hängend												99.1%	0.5%			0.4%
Arm gekreuzt vor Brust					0.5%								99.0%			0.4%
Arm zeigt auf Knie						0.6%								99.0%		0.4%
Arm zeigt nach unten												1.8%			97.0%	1.2%
Keine Detektion																

Klassifizierer: SVM (Validierungsdatensatz)

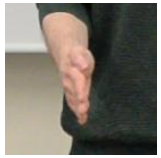
Wahre Klasse \ Erkannte Klasse	Arm zeigt auf UAV	Arm gekreuzt über Kopf	Arm zeigt auf Brust	Arm gekreuzt unter Hüfte	Arm zeigt auf Ohr	Arm zeigt nach oben	Arm zeigt L-förmig nach außen	Arm zeigt auf Hals	Arm zeigt nach unten	Arm hängend	Arm zeigt L-förmig nach oben	Arm zeigt auf Knie	Arm zeigt auf Gesicht	Arm zeigt nach außen	Arm gekreuzt vor Brust	Keine Detektion
Arm zeigt auf UAV	96.8%						1.8%									1.4%
Arm gekreuzt über Kopf		96.7%														3.3%
Arm zeigt auf Brust			96.7%											1.9%		1.5%
Arm gekreuzt unter Hüfte		1.9%		96.6%												1.5%
Arm zeigt auf Ohr			1.9%		96.5%											1.6%
Arm zeigt nach oben	2.0%					96.5%										1.6%
Arm zeigt L-förmig nach außen							96.4%		2.0%							1.6%
Arm zeigt auf Hals								96.4%				2.0%				1.6%
Arm zeigt nach unten									96.3%	2.1%						1.6%
Arm hängend										96.2%					2.1%	1.7%
Arm zeigt L-förmig nach oben							2.1%				96.2%					1.7%
Arm zeigt auf Knie												96.2%	2.1%			1.7%
Arm zeigt auf Gesicht					2.1%								96.2%			1.7%
Arm zeigt nach außen											2.3%			95.9%		1.8%
Arm gekreuzt vor Brust				1.9%											95.8%	2.3%
Keine Detektion																

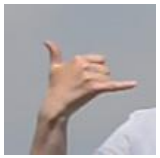
A3 Klassifikationsleistung des Merkmalbeschreibers für gezeichnete Symbole



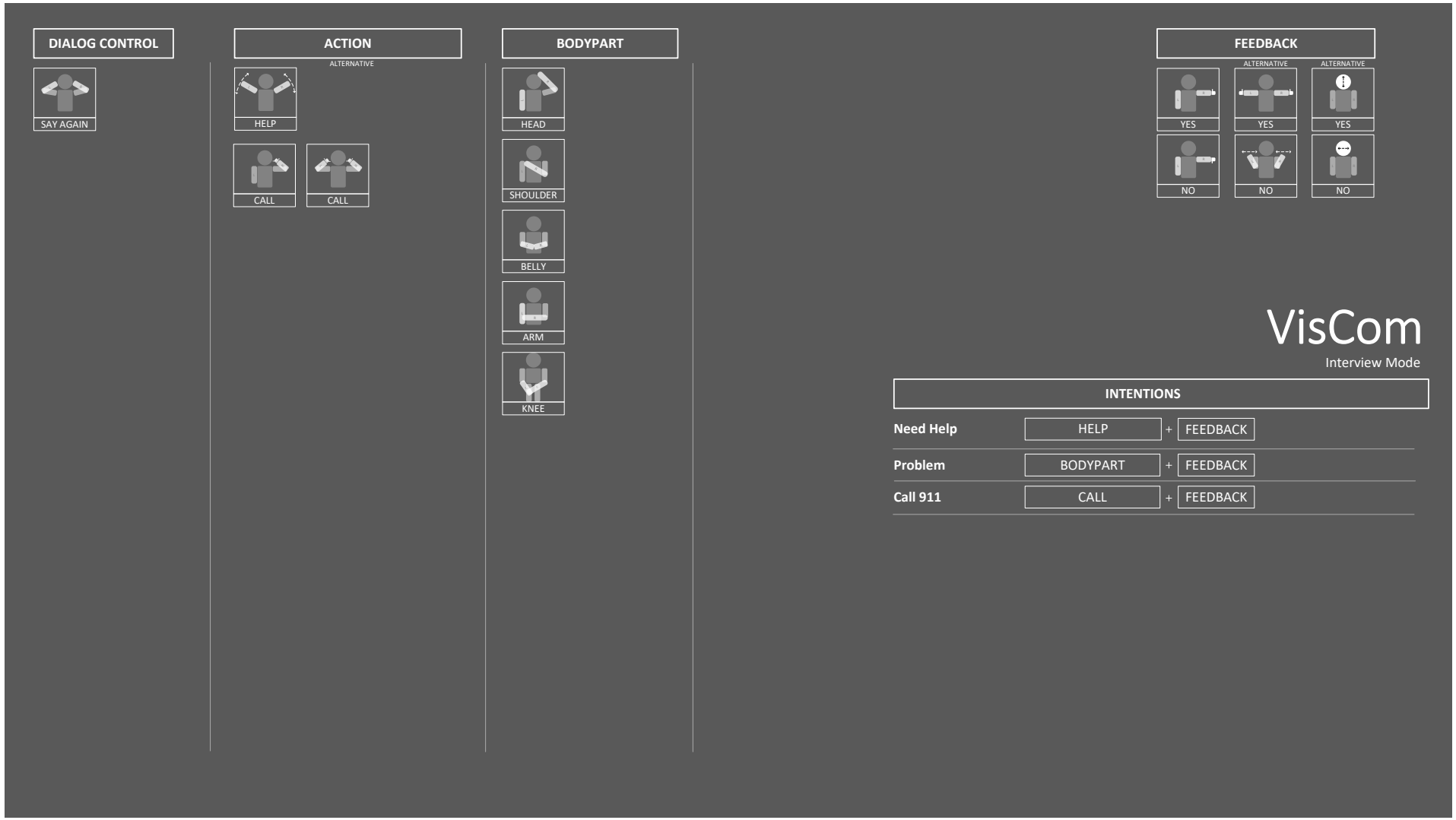


A4 Auszug aus dem Handformendatensatz

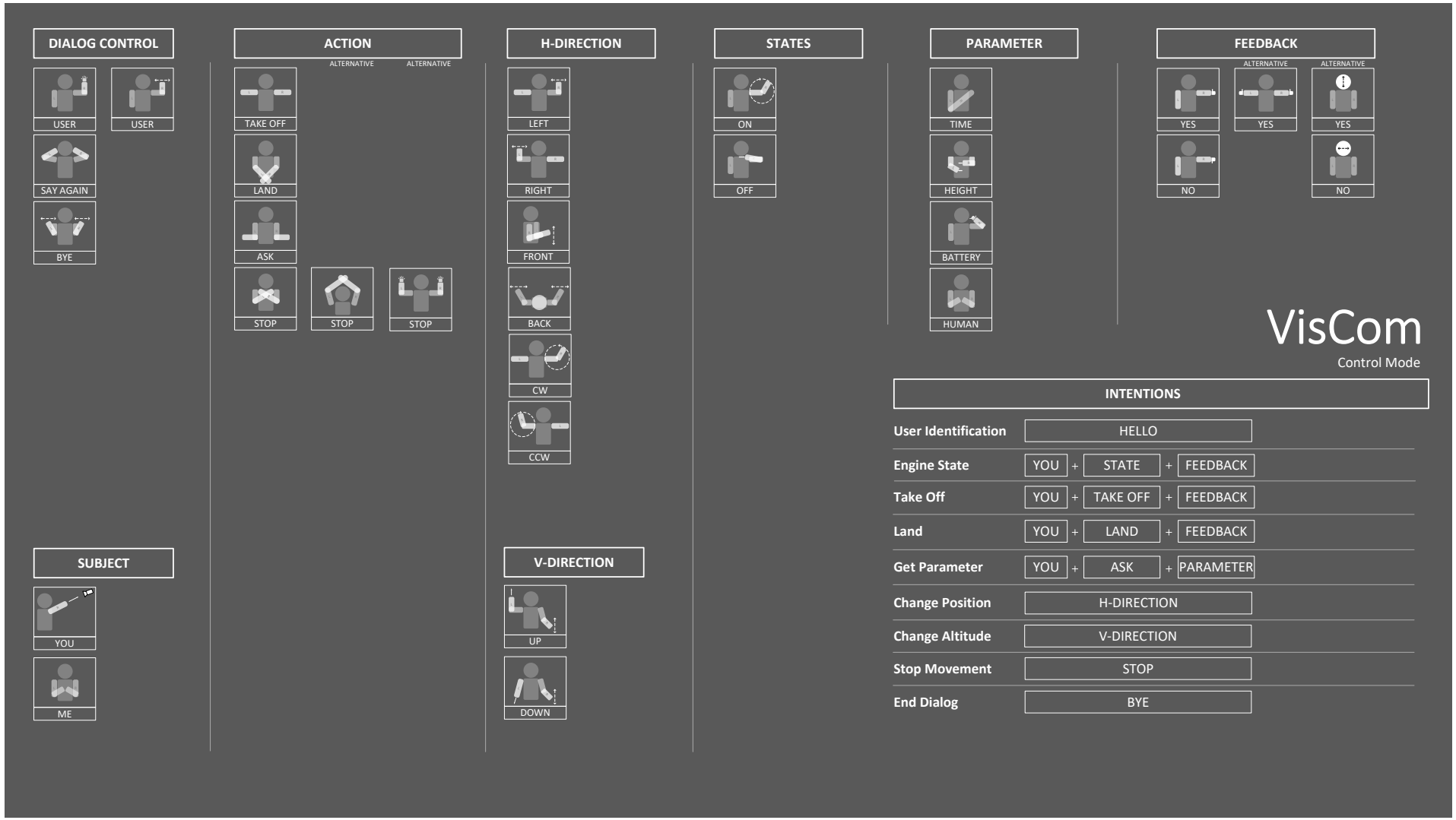
Klasse	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Bezeichnung	Geschlossen	Faust	Flach frontal vertikal	Flach seitlich horizontal	Flach seitlich vertikal	Geöffnet	Zeigefinger	Zeigefinger, Mittel-finger, Ringfinger, kleiner Finger
Beispiel								

Klasse	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Bezeichnung	C	O	V	Daumen runter	Daumen, kleiner Finger	Daumen, Zeigefinger	Daumen, Zeigefinger, Mittelfinger	Daumen hoch
Beispiel								

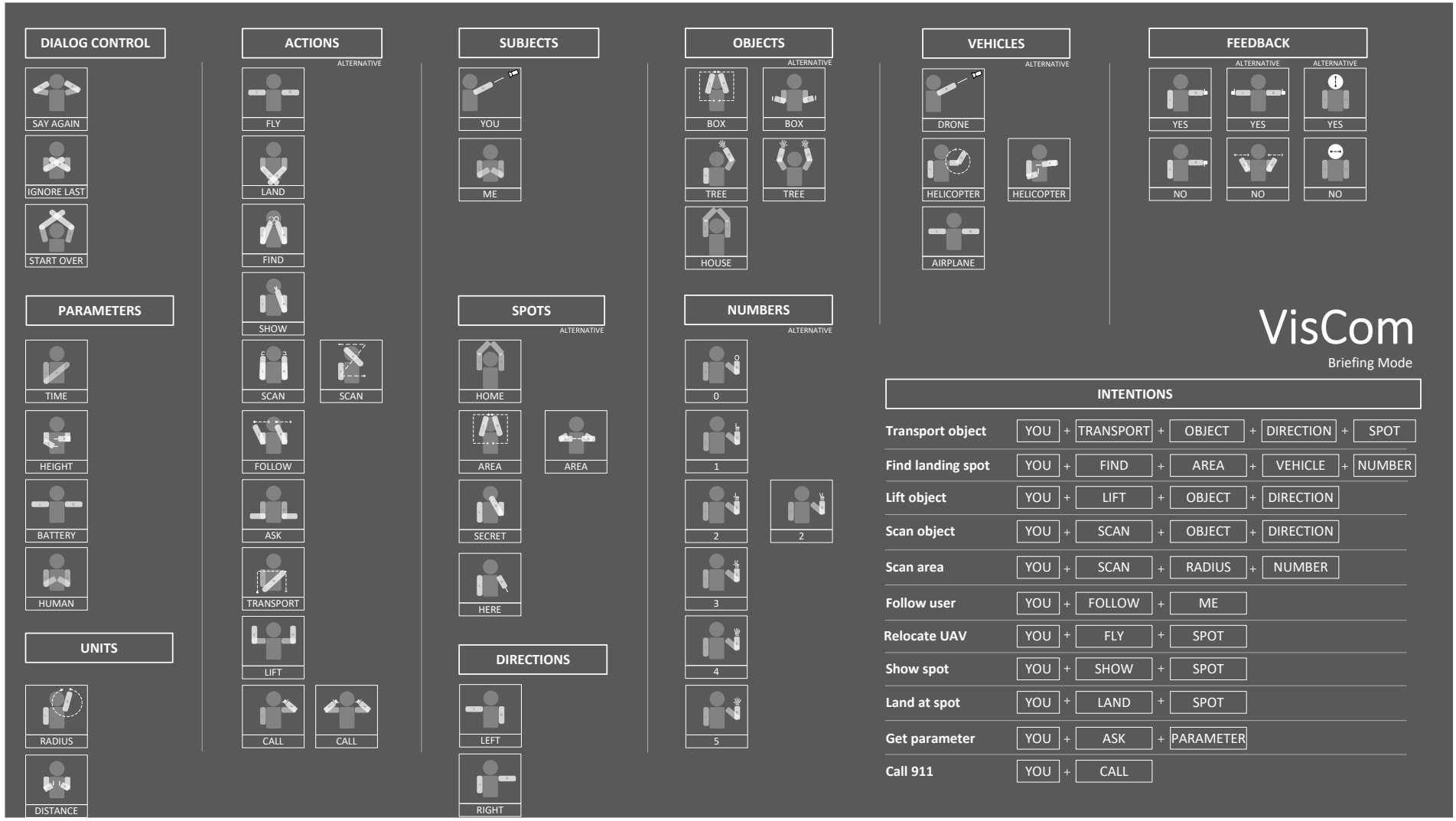
A5 Vokabular und Syntax des Experimentalsystems (Interview Mode für Experiment A)



A6 Vokabular und Syntax des Experimentalsystems (Control Mode für Experiment B)

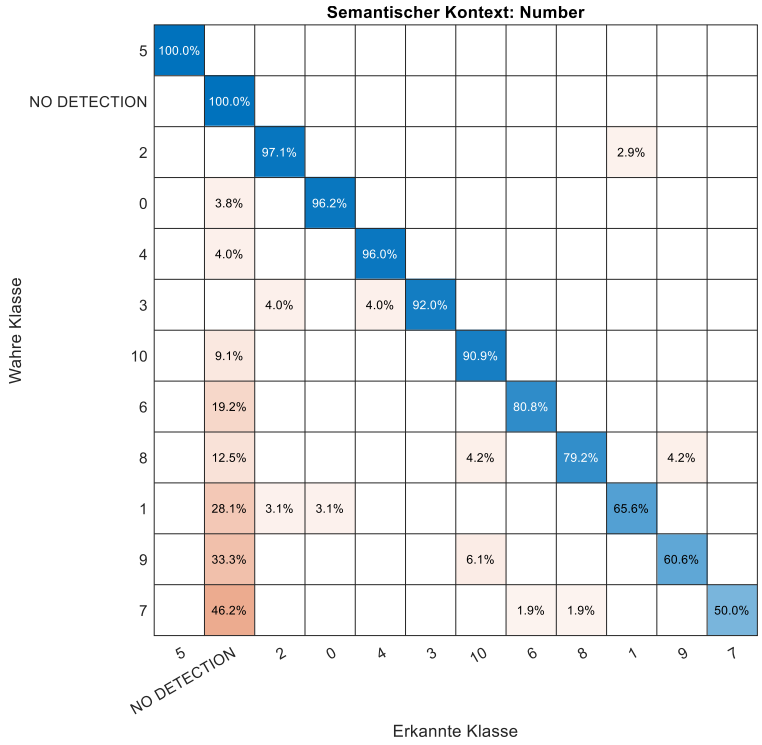


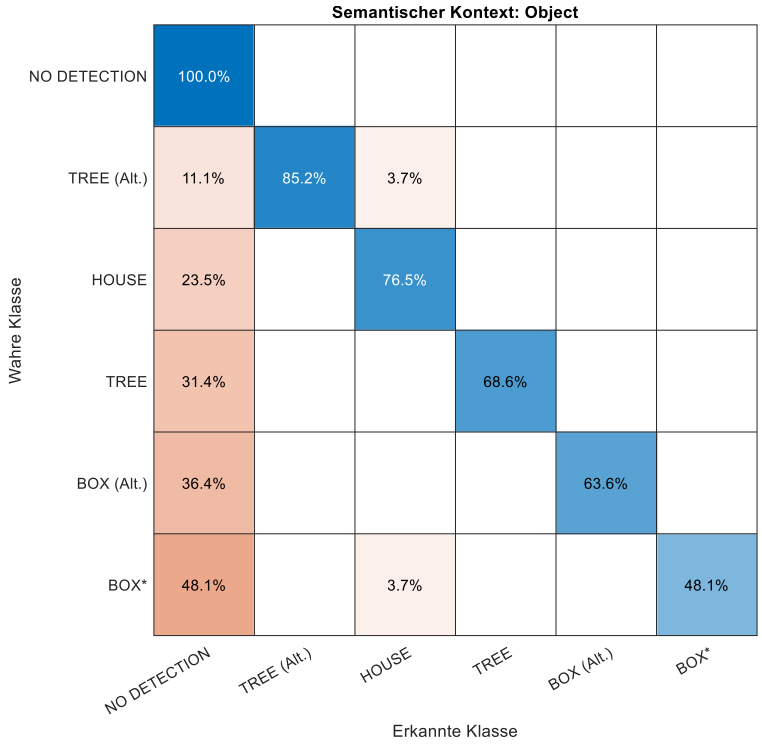
A7 Vokabular und Syntax des Experimentalsystems (Briefing Mode für Experiment C)

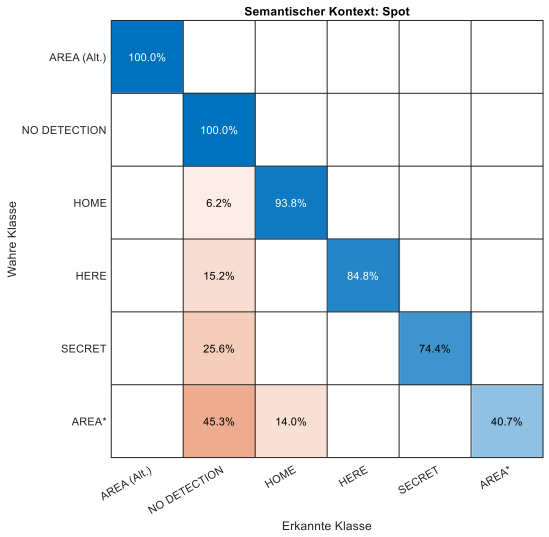


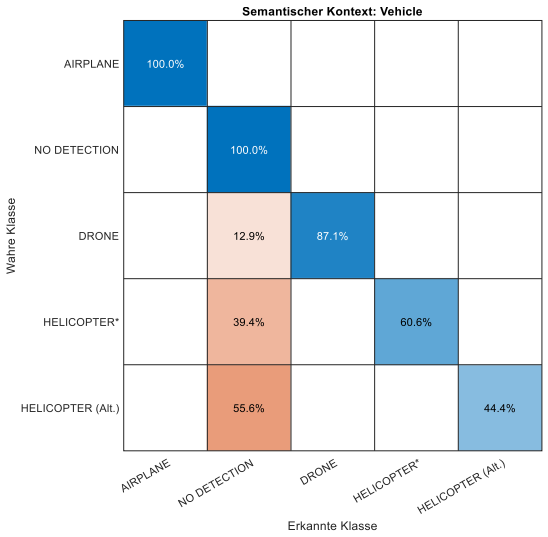
A8 Klassifikationsleistung des Experimentalsystems

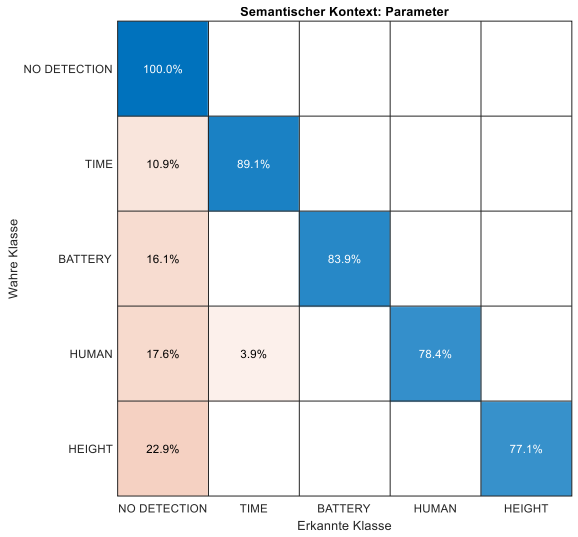
Die nachfolgenden Abbildungen zeigen die Klassifikationsleistung des VisCom Experimentalsystems für jeden semantischen Kontext von dekodierten Körpersignalen unter statischen Laborbedingungen in Form einer Konfusionsmatrix. Die blau hinterlegten Prozentangaben entsprechen der *True-Positive-Rate* (TPR), während orange hinterlegte Felder die *False-Positive-Rate* (FPR) innerhalb des semantischen Kontextes darstellen.

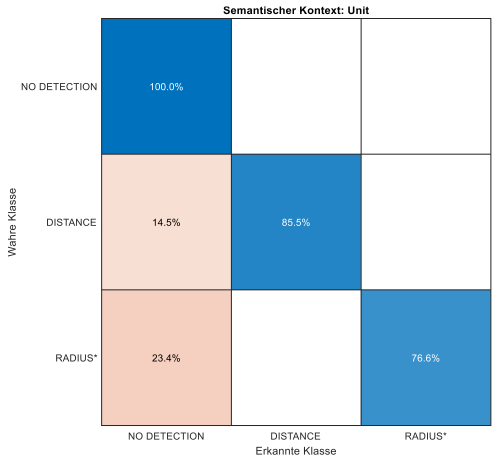


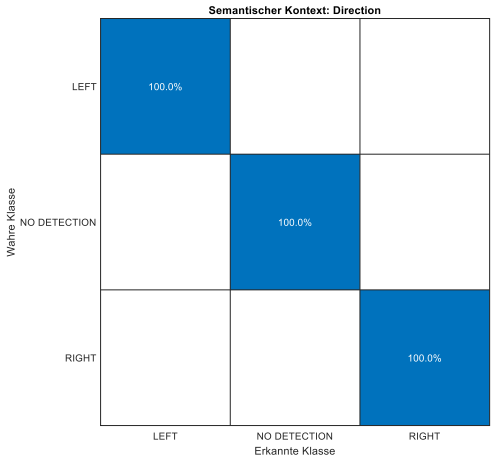


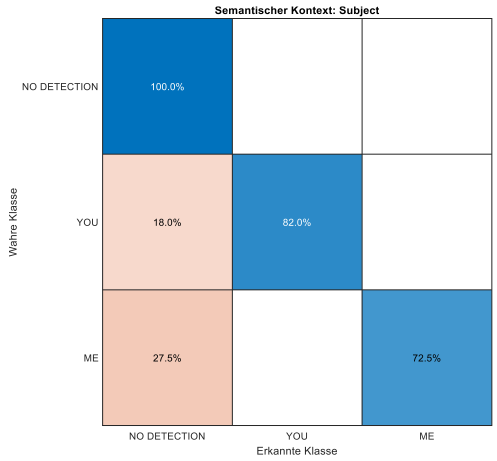

Semantischer Kontext: Spot
Wahre Klasse
AREA (Alt.)
NO DETECTION
HOME
HERE
SECRET
AREA*
100.0%
100.0%
6.2%
93.8%
15.2%
84.8%
25.6%
74.4%
45.3%
14.0%
40.7%
Erkannte Klasse


Semantischer Kontext: Vehicle
Wahre Klasse
AIRPLANE
NO DETECTION
DRONE
HELICOPTER*
HELICOPTER (Alt.)
100.0%
100.0%
12.9%
87.1%
39.4%
60.6%
55.6%
44.4%
Erkannte Klasse


Semantischer Kontext: Parameter
Wahre Klasse
NO DETECTION
TIME
BATTERY
HUMAN
HEIGHT
100.0%
10.9%
89.1%
16.1%
83.9%
17.6%
3.9%
78.4%
22.9%
77.1%
Erkannte Klasse


Semantischer Kontext: Unit
Wahre Klasse
NO DETECTION
DISTANCE
RADIUS*
100.0%
14.5%
85.5%
23.4%
76.6%
Erkannte Klasse


Semantischer Kontext: Direction
Wahre Klasse
LEFT
NO DETECTION
RIGHT
100.0%
100.0%
100.0%
Erkannte Klasse


Semantischer Kontext: Subject
Wahre Klasse
NO DETECTION
YOU
ME
100.0%
18.0%
82.0%
27.5%
72.5%
Erkannte Klasse

A9 NASA Raw TLX Fragebogen nach [358]

NASA-TLX (Kurzfassung deutsch) Seite 1

Beanspruchungshöhe

Geben Sie jetzt für jede der unten stehenden Dimensionen an, wie hoch die Beanspruchung war. Markieren Sie dazu bitte auf den folgenden Skalen, in welchem Maße Sie sich in den sechs genannten Dimensionen von der Aufgabe beansprucht oder gefordert gesehen haben:

Beispiel:

gering X hoch

Geistige Anforderungen
Wie viel geistige Anstrengung war bei der Informationsaufnahme und -verarbeitung erforderlich (z.B. Denken, Entscheiden, Rechnen, Erinnern, Hinsehen, Suchen...)? War die Aufgabe leicht oder anspruchsvoll, einfach oder komplex, erforderte sie hohe Genauigkeit oder war sie fehlertolerant?

gering hoch

Körperliche Anforderungen
Wie viel körperliche Aktivität war erforderlich (z.B. Ziehen, Drücken, Drehen, Steuern, Aktivieren,...)? War die Aufgabe leicht oder schwer, einfach oder anstrengend, erholsam oder mühselig?

gering hoch

Zeitliche Anforderungen
Wie viel Zeitdruck empfanden Sie hinsichtlich der Häufigkeit oder dem Takt, mit dem Aufgaben oder Aufgabenelemente auftraten? War die Abfolge langsam und geruhsam oder schnell und hektisch?

gering hoch

NASA-TLX (Kurzfassung deutsch) Seite 2

Leistung
Wie erfolgreich haben Sie Ihrer Meinung nach die vom Versuchsleiter (oder Ihnen selbst) gesetzten Ziele erreicht? Wie zufrieden waren Sie mit Ihrer Leistung bei der Verfolgung dieser Ziele?

gut schlecht

Anstrengung
Wie hart mussten sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?

gering hoch

Frustration
Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?

gering hoch

Kontrollieren sie bitte, ob Sie zu allen Fragen Angaben gemacht haben. Bei Unklarheiten wenden Sie sich bitte an die Versuchsleiterin / den Versuchsleiter.

A10 Ermittelte Daten aus Experiment A

	Proband		P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
Stichprobenbeschreibung	Beruflicher Hintergrund (0=zivil, 1=militärisch)		0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1
	Alter in Jahren		27	27	29	21	35	32	26	55	25	26	27	25	28	35	28	27	25	28	37	42	23	32
	Geschlecht (0=männlich, 1=weiblich)		0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Interaktion	Realismus der Darstellung (0=niedrig, 1=hoch)		0	1	1	0	1	1	0	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1
	Rückmeldung ähnelt Interaktionswunsch (0=nein, 1=ja)		1	0	1	1	1	1	0	1	1	0	0	0	1	0	0	0	0	0	1	1	1	1
Visuelle Notsignale (A1)	Darstellungsform	Interaktionswunsch[99]	0	5	0	6	0	0	1	0	0	0	0	5	2	0	0	7	0	0	0	0	0	0
		Rückmeldung[100]	2	1	3	2	3	2	0	1	3	0	7	1	3	0	1	0	8	8	5	2	5	5
		Problembeschreibung[101]	0	9	2	2	5	1	3	6	6	0	1	0	0	8	3	7	5	2	0	5	8	-
	Arm-/Handbeteiligung	Interaktionswunsch [102]	1	1	2	0	1	0	2	1	2	1	1	1	2	1	1	1	1	1	1	1	1	1
		Rückmeldung[103]	1	3	1	3	1	0	1	1	1	0	1	3	1	0	3	0	3	3	1	1	1	1
		Problembeschreibung[104]	1	1	0	0	1	0	0	1	1	0	1	1	0	2	1	0	1	0	0	2	0	-
	Mundbeteiligung[105]	Interaktionswunsch	0	1	1	1	1	0	0	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1
	Handform[106]	Problembeschreibung	0	0	1	0	1	1	1	1	1	0	1	0	0	0	1	0	1	1	0	1	0	-
Erkennung von Interaktionswünschen (A2)	Interaktionswunsch erkannt (0=nein, 1=ja)		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Erkennungsdauer in Sekunden		11	2	21	12	1	2	4	1	27	2	2	2	38	10	1	6	1	1	1	1	1	2
Interaktionserfolg (A3)	Problematisches Körperteil korrekt identifiziert (0=nein, 1=ja)		1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	0	1	1	1	-
	Notruf erfolgreich abgesetzt (0=nein, 1=ja)		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0
	Interaktionsdauer in Sekunden		91	74	31	68	166	17	60	67	56	49	65	25	77	168	18	94	113	42	105	28	15	63

-: nicht erfasst

[99] 0=Seitliches Winken, 1=Heranholende Bewegung, 2=Seitliches Winken u. heranholende Bewegung, 3=Kreisen, 4=Auf den Körper Zeigen, 5=Seitliches Winken und Springen, 6=Kreisen u. heranholende Bewegung, 7=Seitliches Winken u. auf den Körper Zeigen

[100] 0=Daumen hoch, 1=Kopfnicken, 2=Heranholende Bewegung, 3=Winken, 4=Winken und Kopfnicken, 5=Wie Aufmerksamkeit, 6=Kopfnicken u. Daumen hoch, 7=Arme hoch (Y-Signal), 8=Verbale Antwort

[101] 0=Zeigen, 1=Berühren, 2=Klopfen, 3=Streichen, 4=Umgreifen, 5=Klopfen u. Umgreifen, 6=Klopfen u. Streichen, 7=Zeigen u. Streichen, 8=Zeigen u. Klopfen, 9=Zeigen u. Umgreifen

[102] 0=Einhändig, 1=Zweihändig, 2=Einhändig, dann zweihändig

[103] 0=Einhändig, 1=Zweihändig, 2=Einhändig, dann zweihändig, 3=Ohne Hände

[104] 0=Einhändig, 1=Zweihändig, 2=Einhändig, dann zweihändig

[105] 0=Ohne, 1=Mit

[106] 0=Zeigefinger, 1=Flache Hand

A11 Ermittelte Daten aus Experiment B

		Proband	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15
Stichproben-beschreibung		Beruflicher Hintergrund (0=zivil, 1=militärisch)	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
		Alter in Jahren	26	27	26	28	28	29	25	35	58	35	32	29	32	44	36
		Geschlecht (0=männlich, 1=weiblich)	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
		Erfahrung mit Fernsteuerung von Fluggeräten (0=nein, 1=ja)	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1
		Anzahl an Jahren mit praktischer Erfahrung	13	12	1	1	7	1	0	1	47	1	5	0	9	30	9
Flug 1 (A nach B)	Var.	Führungsmodus (1=kontinuierlich, 2=diskret)	1	1	1	1	1	1	2	2	2	2	2	2	1	1	2
	Subj. Metriken	Geistige Anforderung	11	4	2	4	3	9	6	18	10	8	14	7	11	11	17
		Körperliche Anforderung	6	14	2	8	9	16	14	16	14	8	10	17	11	1	13
		Zeitliche Anforderung	4	2	2	8	5	4	14	18	10	12	4	13	3	5	15
		Leistung	5	4	2	6	4	4	2	12	10	10	6	18	3	2	9
		Anstrengung	11	6	2	4	5	10	14	18	14	10	18	14	5	5	15
		Frustration	8	4	2	0	7	4	0	14	2	8	7	9	1	1	7
		Beanspruchung	45	34	12	30	33	47	50	96	60	56	59	78	34	25	76
	Obj. Metriken	Entfernung zum UAV nach Landung in Metern	-	7,5	11,2	7,8	13	7,5	10,1	9	9,9	8,63	6,3	8,8	7,2	5,8	5,86
		Landegenauigkeit in Metern	1,87	1,76	0,8	1,5	1,1	2	0,55	1,65	1,56	1,94	1,7	0,74	0,67	1,5	1,3
		Flugdauer in Minuten	6,5	4,59	3,11	4,13	4,82	4,93	3,92	5,97	6,96	4,16	4,94	5,06	3,35	4,78	6,21
		Flugführungspräferenz während Start/Transit/Landung (1=Push, 2=Pull, 3=Parallel)	2/2/3	2/2/2	3/3/1	2/3/1	2/2/2	2/2/2	2/3/1	2/3/3	2/3/1	2/3/1	2/2/2	2/2/2	2/2/3	2/2/2	2/3/1
Flug 2 (B nach A)	Var.	Führungsmodus (1=kontinuierlich, 2=diskret)	2	2	2	2	2	2	1	1	1	1	1	1	2	2	1
	Subj. Metriken	Geistige Anforderung	4	2	2	2	3	8	3	16	8	12	18	16	5	11	15
		Körperliche Anforderung	1	3	2	8	4	2	8	12	4	6	6	17	4	0	7
		Zeitliche Anforderung	4	3	2	10	2	4	8	14	4	12	13	14	2	4	13
		Leistung	5	2	6	6	4	2	2	16	8	8	2	10	4	3	13
		Anstrengung	5	1	2	2	4	3	12	16	4	10	18	16	2	3	15
		Frustration	0	2	8	2	11	2	0	6	0	14	6	14	1	1	3
		Beanspruchung	19	13	22	30	28	21	33	80	28	62	63	87	18	22	66
	Obj. Metriken	Entfernung zum UAV nach Landung in Metern	7,5	6,2	16,1	8,2	10,1	8,4	10,1	11,8	11,36	7,4	8,9	8,9	7,1	4,7	7,75
		Landegenauigkeit in Metern	0,8	0,45	0,35	0,9	0,83	0,5	0,65	2,42	1,25	1,8	0,8	0,57	1	1,6	1,46
		Flugdauer in Minuten	5,47	4,77	5,75	8,29	6,68	4,45	5,14	6,12	5,22	6,05	4,31	5,79	4,6	5,12	5,6
		Flugführungspräferenz während Start/Transit/Landung (1=Push, 2=Pull, 3=Parallel)	1/3/1	1/1/1	2/3/1	1/3/2	1/3/1	1/3/2	1/3/1	2/3/3	2/3/1	2/3/1	1/2/1	1/2/2	1/3/2	1/1/1	2/3/2

-: nicht erfasst

A12 Ermittelte Daten aus Experiment C

	Proband[107]	P01	P02	P03	P04	P05	P07	P08	P09	P10	P11	P12	P13	P14	P15	P17	P18	P19	P20	P21	P22
Stichprobenbeschreibung	Beruflicher Hintergrund (0=zivil, 1=militärisch)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1
	Alter in Jahren	27	27	29	21	35	26	55	25	26	27	25	28	35	28	25	28	37	42	23	32
	Geschlecht (0=männlich, 1=weiblich)	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Gruppenzuweisung (0=ohne, 1=mit Optionsanzeige)	0	1	0	1	0	0	1	0	1	0	1	0	1	0	0	1	1	1	1	0
Durchgang 1	Anzahl korrekter Übermittlungen (Dienstklasse A) in Prozent	87,5	100	100	75	87,5	100	87,5	100	87,5	100	87,5	100	100	100	87,5	87,5	100	100	100	100
(mit Kontextsensitivität)	Summe der Eingaben bei erfolgreichen Übermittlungen	60	54	68	40	51	66	53	61	40	75	67	54	55	53	54	52	49	70	50	49
	Gesamtdauer erfolgreicher Übermittlungen in Sekunden	311	233	251	233	188	289	329	396	217	373	419	268	317	156	278	277	212	482	203	184
Durchgang 2	Anzahl korrekter Übermittlungen (Dienstklasse A) in Prozent	62,5	50	100	37,5	50	100	75	87,5	100	100	87,5	87,5	75	87,5	75	87,5	50	87,5	87,5	87,5
(ohne Kontextsensitivität)	Summe der Eingaben bei erfolgreichen Übermittlungen	39	54	73	18	30	84	47	78	70	186	88	116	58	43	69	114	21	98	83	62
	Gesamtdauer erfolgreicher Übermittlungen in Sekunden	145	230	255	50	107	294	241	401	392	737	452	526	256	142	339	576	61	564	457	195
Frustrationsgrenzen	Mittlere Anzahl Eingaben bis Dialogabbruch	11,25	19,34	24	30,75	22,5	38,25	35,75	31	11,25	26,25	22,75	40,5	34,5	25,75	25,75	29,5	60,75	20,25	47	30,5
(Dienstklasse B)	Mittlere Interaktionsdauer bis Dialogabbruch in Sekunden	56,5	78,67	93	122,5	96,75	210,5	199,75	201,75	300,25	64,75	138,25	139,25	211,75	156,5	123	301,75	189,5	181	225,75	116,5

[107] Aufgrund von Ausreißern in den Daten (insbesondere in der Anzahl der Eingaben und der Dialogdauer) wurden die Probanden P06 und P16 aus der numerischen Auswertung ausgeschlossen.

A13 Veröffentlichungen des Autors

Veröffentlichung	Peer Review (Full Paper)	National/ International	SCOPUS gelistet	Art der Veröffentlichung
A. Schelle, F. Völk, R. T. Schwarz, A. Knopp und P. Stütz, „Evaluation of a Multi-Mode-Transceiver for Enhanced UAV Visibility and Connectivity in Mixed ATM/UTM Contexts“, *Drones*, Jg. 6, Nr. 4, 2022, doi: 10.3390/drones6040080	●	**I**	●	Journalbeitrag
A. Schelle und P. Stütz, „Visual Communication with UAS: Estimating Parameters for Gestural Transmission of Task Descriptions“ in *Lecture Notes in Networks and Systems, Advances in Human Factors in Robots, Unmanned Systems and Cybersecurity*, M. Zallio, C. Raymundo Ibañez und J. H. Hernandez, Hg., Cham: Springer International Publishing, 2021, S. 16–24, doi: 10.1007/978-3-030-79997-7_3	-	**I**	●	Konferenzbeitrag
A. Schelle und P. Stütz, „Evaluierung eines symbolisch motivierten Gestenwortschatzes zur visuellen Kommandierung von unbemannten Flugsystemen“, *8. Interdisziplinärer Workshop Kognitive Systeme*, Duisburg, 28.03.2019	-	**N**	-	Vortrag
A. Schelle und P. Stütz, „Visual Communication with UAV: Use Cases and Achievements“ in *Communications in Computer and Information Science, Computer Analysis of Images and Patterns*, M. Vento et al., Hg., Cham: Springer International Publishing, 2019, S. 120–128, doi: 10.1007/978-3-030-29930-9_12.	●	**I**	●	Konferenzbeitrag
A. Schelle und P. Stütz, „Gestural Transmission of Tasking Information to an Airborne UAV“ in *Lecture Notes in Computer Science, Human Interface and the Management of Information. Interaction, Visualization, and Analytics*, S. Yamamoto und H. Mori, Hg., Cham: Springer International Publishing, 2018, S. 318–335, doi: 10.1007/978-3-319-92043-6_27	-	**I**	●	Konferenzbeitrag
A. Schelle und P. Stütz, „Visual Communication with UAS: Recognizing Gestures from an Airborne Platform“ in *Lecture Notes in Computer Science, Virtual, Augmented and Mixed Reality*, S. Lackey und J. Chen, Hg., Cham: Springer In-ternational Publishing, 2017, S. 173–184, doi: 10.1007/978-3-319-57987-0_14	-	**I**	●	Konferenzbeitrag
A. Schelle, „Gesture Based Communication with UAVs“. *16th PSCE Conference*, München, 05.05.2017	-	**I**	-	Vortrag
A. Schelle und P. Stütz, „Modelling Visual Communication with UAS“ in *Lecture Notes in Computer Science, Modelling and Simulation for Autonomous Systems*, J. Hodicky, Hg., Cham: Springer International Publishing, 2016, S. 81–98, doi: 10.1007/978-3-319-47605-6_7	●	**I**	●	Konferenzbeitrag