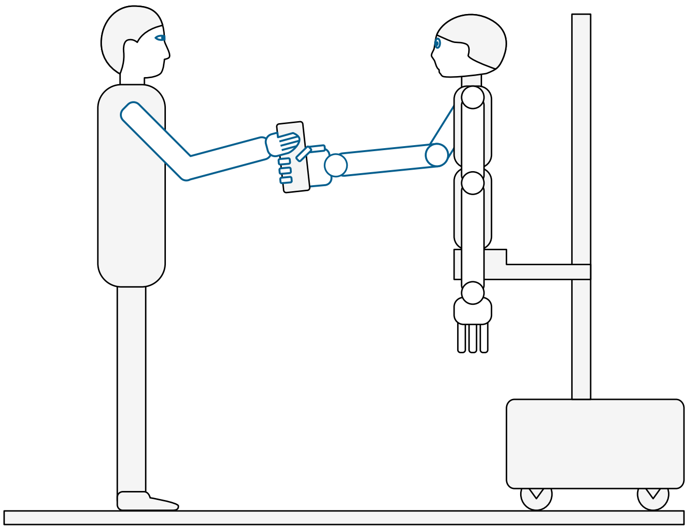
Sebastian Meyer zu Borgsen

NONVERBAL COMMUNICATION DURING HUMAN-ROBOT OBJECT HANDOVER

Improving Predictability of Humanoid Robots by Gaze and Gestures in Close Interaction



Sebastian Meyer zu Borgsen

NONVERBAL COMMUNICATION DURING HUMAN-ROBOT OBJECT HANDOVER

Improving Predictability of Humanoid Robots
by Gaze and Gestures in Close Interaction

A doctoral thesis presented for the degree of
Doctor of Engineering (Dr.-Ing.) at

Cluster of Excellence Cognitive Interaction Technology (CITEC)
Faculty of Technology
Bielefeld University
Inspiration 1
33619 Bielefeld
Germany

REVIEWERS
PD Dr.-Ing. Sven Wachsmuth
Prof. Dr. Dr.-Ing. Matthias König

EXAMINATION BOARD
Prof. Dr.-Ing. Stefan Kopp
Dr. Guillaume Walck

DEFENDED AND APPROVED
May 18, 2020

Printed on permanent paper as per ISO 9706.

ABSTRACT

This doctoral thesis investigates the influence of nonverbal communication on human-robot object handover. Handing objects to one another is an everyday activity where two individuals cooperatively interact. Such close interactions incorporate a lot of nonverbal communication in order to create alignment in space and time. Understanding and transferring communication cues to robots becomes more and more important as e.g. service robots are expected to closely interact with humans in the near future. Their tasks often include delivering and taking objects. Thus, handover scenarios play an important role in human-robot interaction. A lot of work in this field of research focuses on speed, accuracy, and predictability of the robot's movement during object handover. Still, robots need to be enabled to closely interact with naive users and not only experts.

In this work I present how nonverbal communication can be implemented in robots to facilitate smooth handovers. I conducted a study on people with different levels of experience exchanging objects with a humanoid robot. It became clear that especially users with only little experience in regard to interaction with robots rely heavily on the communication cues they are used to on the basis of former interactions with humans. I added different gestures with the second arm, not directly involved in the transfer, to analyze the influence on synchronization, predictability, and human acceptance. Handing an object has a special movement trajectory itself which has not only the purpose of bringing the object or hand to the position of exchange but also of socially signalizing the intention to exchange an object. Another common type of nonverbal communication is gaze. It allows guessing the focus of attention of an interaction partner and thus helps to predict the next action.

In order to evaluate handover interaction performance between human and robot, I applied the developed concepts to the humanoid robot Meka M1. By adding the humanoid robot head named Floka Head to the system, I created the Floka humanoid, to implement gaze strategies that aim to increase predictability and user comfort. This thesis contributes to the field of human-robot object handover by presenting study outcomes and concepts along with an implementation of improved software modules resulting in a fully functional object handing humanoid robot from perception and prediction capabilities to behaviors enhanced and improved by features of nonverbal communication.

ACKNOWLEDGMENTS

Thanks to everyone involved in making this thesis possible!

At first, I want to express a special thanks to PD Dr.-Ing. Sven Wachsmuth as my supervisor, for his guidance and support. He not only helped me through all the stages of the biggest project in my life, but also made it possible for me to work in such a great environment as well as letting me experience ten RoboCup@home competitions. I am grateful that this Thesis was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation. Furthermore, I would like to thank Prof. Dr. Dr.-Ing. Matthias König for taking time to review this thesis and to Prof. Dr.-Ing. Stefan Kopp as well as Dr. Guillaume Walck for joining the examination board.

Thanks to everyone who helped me to get and keep the robot running. Phillip and Luca did a great job, enhancing and fixing functionalities of the platform. Jan and Florian provided the required tooling with the RDTK, which contributed to keeping track of all the involved software components and libraries of the resulting system during development and experiments. I owe special thanks to Simon for sharing the office with me, repairing the robot so many times with me and becoming a good friend. I am also indebted to all my colleagues from Central Lab Facilities (CLF), Applied Informatics (AI), Neuroinformatics (NI), Cognitive Systems Engineering (CSE), and Cognitive Service Robotics Apartment (CSRA) for helping and discussing ideas with me. I am grateful to our writing group (Birte, Eva, Jasmin, Mara, Marian, Sebastian, Viktor) that provided guidance, feedback and motivation and to Johannes for providing the latex template.

As pursuing a PhD is an overall hard endeavor, it requires not only substantial support. Thanks to my family and friends, who provided emotional relief during this time and to Matthias and Leon, who made me start the journey of becoming a scientist. Tijn who told me, during the festivities after becoming RoboCup@Home world champion, "pursuing a PhD is like rural firewalking and simply has to be done". I would like to add thanks to all the RoboCup students and especially student assistants: Dominik, Johannes, Leroy, Lukas, Nils, Jan, and Janneke. It was a great time with all of you. I not only learned a lot about robots but also about people, places and myself. Thanks to Kirsten and Ayla for recruiting and all the participants for taking part in my studies.

Thanks to my lovely wife Katherine.

CONTENTS

LIST OF RESEARCH QUESTIONS

LIST OF HYPOTHESES

LIST OF REQUIREMENTS

LIST OF FIGURES

LIST OF TABLES

NOTATION

MARGIN NOTES

- Key point
- Definition

MEAN AND STANDARD DEVIATION

Values that where deduced from multiple repetitions are given as the mean and the standard deviation in the form 1.234 ± 0.056 m. The first part gives the mean and the number after $\pm$ denotes the standard deviation. The given unit is valid for both parts.

ATTRIBUTION OF AUTHORSHIP

I will speak of myself using *I* in case of work originally done by myself alone. In case the results of a collaboration with others are presented, I will use *we*. The respective collaborators are indicated by the co-authors of the publication the results are based on.

Part I

RESEARCH TOPIC

This is the introductory part of this thesis, starting with an overview of the topic. It sets up research questions, hypotheses, and proposed requirements. After that, I introduce related work in the field of object handover, non-verbal communication, and robot control research. It motivates my work and explains the mandatory concepts.

1 INTRODUCTION

As early as the 10th century BC, people already started building mechanical automata. An automaton is a machine that mechanically (re-)produces behavior. Since then, such machines were steadily improved and gained functionality. Today we call such a behavior producing machine robot, a term first coined by Čapek in the year 1920 with his play *R. U. R. - Rossum's Universal Robots* [Čap20]. In this play robots are artificial people, but not even machines rather biological copies of humans. Today the term robot evolved to be used for describing machines that can perform tasks automatically. What started as a dystopia a century ago becomes more and more reality. Not in the way that we have biological copies of ourselves but machines that come closer to human characteristics each day. There are already robots with hundreds of motors. Equipped with sensors and computation capabilities, they can be programmed to show sophisticated behavior. As these machines become more complex, they also start to interact with humans. Such human-robot interaction (HRI) [GS07] can be basic co-working tasks in a factory or really complex interactions involving social cues.

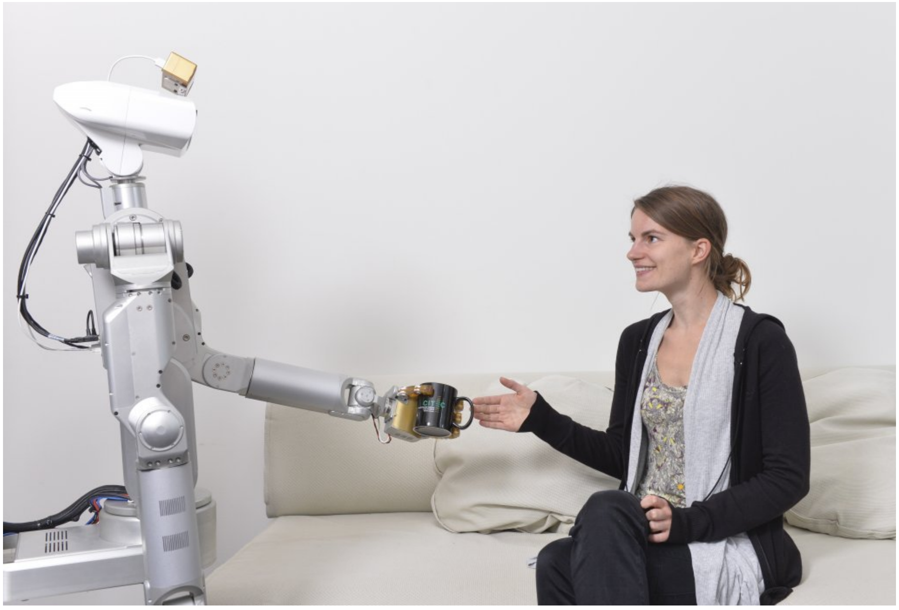


Figure 1.1: The humanoid robot Meka M1 Mobile Manipulator (Meka M1) hands a cup to a sitting woman. The hand and body posture is interpretable as an object handover. Meka M1 is reaching out with the cup and the woman is holding her hand open and keeping eye contact, signaling readiness. Photo: [CITEC]

"Robot, please bring me a coffee", or variations with beer, coke, etc., is one of the most common sentence scientists hear when visitors see their robots. While such a sentence sounds simple, it involves a lot of tasks for the robot like understanding the spoken words, navigating to the coffee machine, pouring a coffee, (let alone brewing it), grasping it, navigating back to the person, and concluding with handing over the cup. Even those tasks can be split into sub-tasks that involve sensing, processing and acting on it. Figure 1.1 shows such a situation where the Meka M1 gives a cup to a sitting woman. What is shown in that static picture is a highly dynamic process. Handing over objects is a complex task that entails collaboration and precise synchronization in space and time. Object handover takes place everywhere in our daily lives [Str+12; Cha+13; SS15]. Tasks that future robots are expected to handle include household tasks like delivering a drink, helping with the dishes, handing the TV remote, and collaborating at work where tools or parts need to be exchanged between co-workers. Therefore, the ability to exchange objects with humans is mandatory for socially accepted interactions with service robots. Such close interactions also demand a social behavior of the robot. One aspect for collaboration is communication which helps to create joint action understanding. It has been shown that the integration of nonverbal cues like gaze and head orientation improves robot-to-human object handover [Gri+13; Str+13; Zhe+14; Moo+14].

Handing over objects is a type of HRI that combines multiple research topics that need to fluently interact and overlap for a successful interaction. Good interaction always starts with perceiving the partner. The obvious part for an object handover is detecting the hand, whereby the body and facial expression play an important role as well. Such information is used to derive the intent of the interacting person. As object handover is a highly dynamic process not only the posture but also movements have to be taken into account. For a smooth transfer it is important to predict the future state to estimate an object transfer point. With such a prediction about position and time the robot can generate a trajectory for its end-effector (EEF). Here it has to be taken into account that not only the EEF needs to reach the target position at the right time. This is only half of the story as it only describes the functional part of exchanging an object. For a good and predictable interaction the nonverbal communication (NVC) takes an important role. Already in 1872, Darwin started to systematically analyze NVC with the book "The expression of emotions in Man and Animals" [Dar72]. With this work he created the foundation for research of nonverbal communication according to Pease and Pease [PP04]. They explain in their book "The Definitive Book of Body Language: How to read others thoughts by their gestures" how humans can read thoughts and plans of others by just observing them. This shows that we consciously or unconsciously transfer a lot of information with our movements. If we propose that machines with human-like physical properties are able to communicate in the same nonverbal way, it becomes really important to address this topic in robotics. It needs to be discussed which chances and possibilities as well as limitations are in the applicability of human NVC onto robots. Especially at the current state of robot capabilities, where they start to match the human in some regions but are still steps behind in others, it is important to scale the "body"-language accordingly.

Mori wrote about "The Uncanny Valley" already in 1970 as the effect that as robots become more and more manlike the likability increases to a point where they are really disliked. He states that this effect exaggerates with movement. This means when we design movements for robots we have to stay away from this valley to improve nice and likable behavior [Mor70]. Although almost 40 years later Bartneck et al. found out that the valley is more like a cliff [Bar+07], it highlights even stronger that we should not directly copy humans but design behavior that fits the robots that makes use of it. This underlines that we need to design specific robot behavior instead of copying it directly from humans.

Evaluating interaction between humans and robots is a research area on its own. One way of quantifying a robot's performance is to put it in a competition with other robots. The *RoboCup@home* [ZW06; ZW07] league committed itself to this task by testing and evaluating robots in real-world scenarios. The competition starts with the *robot inspection*, where the robots are tasked to autonomously register themselves for the competition. Since 2010 they can deliberately hand over a registration form to a member of the technical committee [IRS10, p. 23]. Two years later, in 2012 this object handover (handover) became mandatory to receive points for the autonomous registration [HS12, pp. 31–33]. Besides the inspection there are other tests that involve handover, like *Restaurant* and *Cocktail Party* where the robot acts as a waiter, or the *General Purpose Service Robot* where almost every imaginable service task can be given to the robot. The experience in the *RoboCup@home* [Zie+13; Zie+14; Zie+15; MKW16; Wac+17; WLM18; Mey+17] showed that the current human-robot handovers are still far from natural. While synchronization between two humans happens subconsciously in this task, handovers between a human and a robot still require an explicit protocol. Robots use special commands or sensors that either require explanation by the team or robot or an experienced technical committee. These methods involve holding the object and waiting for someone to pull it out of the gripper with enough force or touching a sensor or button to trigger opening the robot's EEF. This underlines the necessity of research in this domain.

In this thesis I discuss how NVC helps to improve the predictability of service robots during human-robot object handover.

1.1 OUTLINE

This thesis is structured into three parts and an appendix.

The first chapter of the first part introduced and motivated the research topic. In Chapter 2: THESIS PROSPECT I address the research questions, derived hypotheses, and system requirements. The Chapter 3: TERMINOLOGY AND CONCEPTS introduces the necessary terms and further motivate the requirements of this work. Based on the literature I deduce a coherent concept of handover including its types and structure. I establish the relationship and differences between human-human and human-robot object exchange. In addition, this part goes into the topic of nonverbal communication for handover with gaze and gestures.

Part II: The Object Handing Robot starts with an introduction of the platform I implemented and evaluated my concepts with. Hereafter, I give a description of a baseline implementation and experiment which led to the enhancements and extended concepts required for fluent handover. This is verified in an evaluation study with users having different levels of experience with robots, presented in Chapter 5: HUMAN-ROBOT HANDOVER EXPERIMENT. Derived from that study, I present three new additions for improved human-robot handovers. Chapter 6: USING ROBOT GAZE FOR PREDICTABILITY gives details of an enhanced gaze scheme. In Chapter 7: FUNCTIONAL GESTURE MOTIONS I introduce a concept and implementation of a module for smooth, predictable, and flexible reaching motions. Therefor, in Chapter 8: PERCEPTION AND PREDICTION OF OTPS I present modules for tracking of hands and predicting the location of handover. These improvements are combined in a coherent behavior presented in Chapter 9: COMBINED HANDOVER SYSTEM and put to a test described in Chapter 10: FINAL EVALUATION.

In the third and last part (Part III: Perspectives) the results are summed up and discussed. Furthermore, I give an outlook on future work in the field of human-robot object handover in terms of open challenges that emerged during my work or were out of the scope of this thesis. I conclude with the last chapter that retrospectively elaborates my research questions. The Appendix contains supplementary material that was used and created in the context of this work.

2 THESIS PROSPECT

In this thesis I address the topic of nonverbal communication (NVC) during human-robot object handover. I will especially investigate the integration of gaze and gestures executed during an object handover by a robot. As robots and humans are still different in appearance and their movement capabilities, one can not directly copy movement patterns already known from human-human interaction. This might be due to hard- and software limitations or for security reasons. Thus, it needs to be found out what kind of adaptations are needed and what are the most crucial party of NVC for a robot. This leads to the following research questions (RQs), the derived hypotheses (Hs), and subsequent system requirements (SRs) that I explain and deduce later-on during this thesis.

2.1 RESEARCH QUESTIONS

The research questions in this thesis address the general pattern of object handover, the involved nonverbal communication, the influence of experience with robots, and the perceptive capabilities of a robot to achieve a smooth object handover interaction. While the existence of a functional part of an object handover is obviously transferring an object physically from someone to someone else, I claim that the whole movement involved is a gesture, meaning that the arm motion conveys information. For exchanging an object it is required that two hands or EEFs meet in space and time. For the robot this means that its EEF needs to be moved to a predicted target object transfer point (OTP). While this is the core part of an object handover, it requires some preceding steps for successful and fluid execution. The gesture part, on the other hand, transfers information between the interactants. During handover this can include the *what?*, *when?*, *where?* and *how?*.

Based on the existing literature and own experiments I examine the question on which kind of pattern exists and how it can be transferred from humans to robots. If such a pattern can be found it likely needs to be scaled and adapted to fit a robot. The question on how to scale the behavior that has already been observed in human-human handover is thus tightly coupled to the question on a general pattern of human handover strategies. As most robots are currently not as fast as humans and it most likely will be like that for a long time, I want to find out how to adjust the behavior such that it maintains the general structure but allows execution with less speed and acceleration for a safe and predictable interaction experience. This includes the search for a general structure that is applicable for a variety of situations. Research question 1 summarizes this research topic in one question.

Research Question 1: Handover Interaction Pattern
What is the underlying structure of human object handover and how can it be implemented on a robot companion?

Another topic addressed in this thesis is nonverbal communication during human-robot object handover as a slightly less obvious but not less important part of such an interaction. Thus, with RQ 2, I will explore which types of gestures and gaze can be integrated in object handover. With the results of the aforementioned question I will look for useful social signals that can be incorporated in the robots during each phase of the handover. The parts that are coupled, also need to be checked for their interplay and how they should be synchronized to the functional parts of the handover.

Research Question 2: Impact of Nonverbal Communication
Which types of nonverbal communication can be utilized to improve object handover in terms of predictability and comfort?

As robots are expected to leave controlled environments and laboratories it needs to be ensured that they are able to interact with the people outside of such controlled environments. They should be easily operated, not only by researchers, students, and other technical people but also by everyone. For example the ongoing nursing staff shortage might be tackled with service robots in the near future. In nursing homes the robots need to interact with elderly people that might be naive in regard to interacting with non-human care takers. Research question 3 rests upon the question on how to design object handover behaviors of robots that are understood by everyone. On the other hand, the interaction of experts should not be slowed down by forcing them to use fixed pattern. Thus, it needs to be found out what kind of shortcuts can be allowed to be taken by experts to use robots as useful tools.

Research Question 3: Influence of Expert Knowledge
How can we create robotic object handover behavior that is understood by everyone, being people that are inexperienced with robots or experts that could be even faster by taking shortcuts?

I propose that we would like to have responsive interaction partners that handle some effort in a shared task instead of leaving it all to the human. Thus, I aim to propose solutions that react to the human and create a synchronization between both interactants. Therefore, the robot needs to be able to perceive the human. I will check for existing solutions and propose new ones that fulfill the requirements to extend the robot's capabilities such that a reactive human-centered behavior is created.

Research Question 4: Perception Requirements
Which kinds of perception or understanding of the world does a robot need for smooth human-robot object handover?

These questions are tightly coupled and need to be addressed in unity for comfortable object transfer by shifting cognitive as well as physical load from the human to the robot.

2.2 RESEARCH HYPOTHESES

From the aforementioned research questions I deduced hypotheses that I am going to address in this thesis. While some of them have been established and are reoccurring, I attend to them under different perspectives. In Chapter 3: TERMINOLOGY AND CONCEPTS I give a more distinct classification and demarcation of the given hypotheses. Research question 1: Handover Interaction Pattern addresses the existence of a structure of object handover. The following hypothesis adds to it by presuming that a distinct pattern with individual phases exists. Such pattern would allow to implement a repeatable behavior on a robot, modeled as states in a finite-state machine (FSM).

Hypothesis 1: Handover Has a Distinct Pattern
The process of handing over an object follows an implicit model with distinct phases that is repeatable and can be implemented on a robot.

With RQ 2: Impact of Nonverbal Communication, I am exploring the types of NVC during a handover. The H 2: Second Arm Helps to Synchronize addresses body language during human-robot object handover that I will analyze. Assuming that the object is transferred unimanual, a humanoid robot (humanoid) has additional body parts that can be freely used to create nonverbal social signals. As an addition to the body posture that mainly includes the posture of the torso and the robot's legs or base, the second arm can be utilized to nonverbally communicate the intention and state of the robot. With H 2 I propose that moving or positioning these body parts correctly might help to transfer the state of the robot to the interacting person without verbally explaining it, leading to an improved synchronization.

Hypothesis 2: Second Arm Helps to Synchronize
Gestures with body parts of the robot that are not directly involved (e.g. the second arm) in the object transfer help to synchronize position and timing between human and robot.

As addressed with RQ 3: Influence of Expert Knowledge experience changes the way we, humans interact. I propose that interaction partners with no, or only little experience in the interaction with robots need the machine to move and behave similarly to what they are used to from interaction with humans. While experts in contrast learned the particularities of the robot's behavior and adapted to it. Hypothesis 3 proposes that such differences can be observed in object handover. I will address this hypothesis in a study to assess the disparities based on differences in the level of experience with robots.

Hypothesis 3: Experience Changes Interaction
Naive and (robot-)experienced users handle object handovers with robots differently.

2.3 SYSTEM REQUIREMENTS

Based on the research questions, hypotheses, and related work published in this field I derive ample SRs. The following requirements determine capabilities and features that are required for smooth object handovers or are useful for an improved system design.

2.3.1 *Handover Structure*

Based on the properties of Hypothesis 1: Handover Has a Distinct Pattern, the overall behavior need to be defined. At first, it is important to replicate most of the human behavior in the robotic system to create an easily readable overall behavior (SR 1(a)). System requirement 1(b) stresses the importance of scalability under abidance of communicational properties. This requirement is coupled to choosing a robot system and its capabilities but should provide a general way of scaling the behavior in terms of its speed and acceleration. The behavior pattern also needs to be reactive for a human-centered, comfortable interaction (SR 1(c)).

System Requirement 1(a): Human-Like Pattern
The robot needs to replicate the basic behavior pattern humans show during object handover, to be easily understood by humans in the first place.

System Requirement 1(b): Pattern Scalability
The pattern needs to be scaled down in terms of speed and acceleration to be performed safely with today's robots without losing the characteristics of the basic pattern.

System Requirement 1(c): Reactive Pattern
Reactivity and adaptivity to the human need to be embedded into the pattern for increased user comfort.

2.3.2 *Predictable and Natural Motions*

The Hypothesis 2: Second Arm Helps to Synchronize hypothesizes that gestures with an additional arm can be added to an object handover for a more predictable interaction. Linking to it with the following SRs with which I propose additional properties of motion by means of NVC. These movements start with approaching and the whole body posture (SR 2(a): Body Orientation). I suggest the body posture and orientation of the robot as an important element that allows adopting to the human as well as social signaling the intent of the robot to exchange an object. System requirement 2(b) focuses on the whole trajectory of the robot's arm itself. While the arm, hand, or EEF that transfers the object has an obvious functional task, it also has the task of communicating the intent of exchanging an object. Which means, that while being functional it is desirable that the arm follows a gesture-like characteristic. Such communication can be assumed to be non-binary, meaning it does not only communicate the *what*, but also *where*, *when*, and even *how*. Thus, I will address the properties of the movement itself.

Looking at something or someone can be perceived by an interaction partner. Such referential gazes can be seen as a tool to create joint attention. If such a tool is used in human-robot object handover, the human might predict the robot's plans and motions more easily. This way the cognitive load of the human is reduced, which will lead to a more comfortable interaction, being one of the goals of HRI research. Therefore, I propose gaze cues as one of the essential SRs (SR 2(c): Gaze for Predictability).

System Requirement 2(a): Body Orientation
The robot should communicate readiness and willingness to interact by adjusting the torso position towards the interactant.

System Requirement 2(b): Gesture Motion
A handover motion should not only reach a given position but be a gesture that clearly communicates the information on what*, will happen* when*, and* where.

System Requirement 2(c): Gaze for Predictability
Directed gaze should be incorporated to increase predictability of the robot's behavior.

2.3.3 *Interaction with Everyone*

In the near future robots that are able to interact with everyone are needed. Especially a regular activity like handover should be accomplishable by everyone without further training. Being it young or elderly people at home or in care-homes, restaurants, or at work, the robot should interact such that there is no limitation of interaction partners. Differences in behavior based on the level of experience with robots (see H 3: Experience Changes Interaction) should be expected. Bringing in RQ 3: Influence of Expert Knowledge, might lead to the assumption that the robot's behavior pattern needs to have most aspects expected by naive interactants (SR 3(a): Understandable by Everyone). On the other hand experts should be enabled to take shortcuts for an efficient and fast interaction as long as robots do not meet human performance.

System Requirement 3(a): Understandable by Everyone
The handover should be executable by everyone without explanation and especially novices should be able do understand the robot directly.

System Requirement 3(b): Shortcuts for Experts
Experts should be able to skip parts of the embedded phases to speed up the process that they are used from interaction with machines.

2.3.4 *Required Perceptional Capabilities*

The RQ 4: Perception Requirements uncloses the field of perception capabilities of robots in HRI. This large topic of research will be addressed here in regard to object handover. A fixed movement of the robot itself does not require much perception — if one does not consider the control level and does not interpret position encoders and such as perception. A reactive robot on the other hand requires a certain level of awareness of the situation. For an interaction like object handover, the contemplable aspects are the perception of the interacting person, the object, and also the surroundings for context and collision avoidance. In order to initiate a handover, a certain level of nearness is required which can only be achieved if the robot sees the interactant. This might initially be a rough coordinate of the person that allows the robot to approach. For a smooth and reactive handover, the requirements in regard to information about the person increase. Therefor, the location of the hand needs to be perceived to position the EEF correctly. In case the robot is tasked to act proactive during the handover it needs to estimate a future or even final position of the human hand. This behavior is the most desirable which requires the least amount of waiting for the human.

In regard to NVC I propose the need for a more sophisticated perception of the interacting person to facilitate bidirectional exchange of social signals. A perception of the human face can enable the robot to act on the produced gaze cues for an improved synchronization. Additionally, the exchange needs to be detected in terms of when to grasp or release the object and determine whether the exchange was successful for the possibility of a regrasp. Another important requirement for the perception is that one should not require to alter the environment or interactant. There must not be the need for attached (artificial) markers to objects or persons interacting with the robot.

Based on the discussed capabilities I propose the following list of system requirements in regard to perception capabilities of the robot which will be discussed further throughout my thesis:

System Requirement 4(a): Hand Tracking
During the handover the robot should be able to detect and track the position of the partner's hand to achieve reactive behavior.

System Requirement 4(b): Object Transfer Point Prediction
For a proactive behavior a prediction of the future hand position is needed for further reduction of the human's effort.

System Requirement 4(c): Person Tracking
The robot needs the current position of the interacting person for approach and alignment.

System Requirement 4(d): Face Tracking
The position of the human face provides information that can be used to create joint attention and thus help to create synchronization between human and robot.

System Requirement 4(e): Contact Detection
A contact detection is required to detect when to open/close the hand and whether the exchange was successful.

System Requirement 4(f): Visual Transfer
A visual transfer detection is required to detect when to exchange the object without applying force.

System Requirement 4(g): Markerless Perception
All perception has to be done without markers in a natural environment.

2.3.5 *Design and Integration*

Human-robot interaction requires a robot to be participating. Hence, the design needs to consider the properties of the existing system. Modules and behaviors have to be implemented for as well as to be integrated in such a system with existing hard- and software. The desire is to integrate it with only little modifications of the system to make sure existing behavior is not degraded during the process. The types of NVC need to be combined with the attendant phases. As stated earlier, most robots still lack the physiological capabilities of humans. Even if some constructions attain the speed, the acceleration, and smoothness, there might be the need or wish for limitations due to safety concerns. Additionally, there can always be simpler and cheaper constructions for which the requirement of slower movements with good readability stay important. I target to address as much of the human-like NVC properties of object handover as possible. Thus, a robot that possesses the requirements, as human-like arms or a head for making examinations of such cues possible, is required. Choosing a specific robot poses requirements on the concept and implementation of the system by e.g. the availability of sensors and processing capabilities. These will become apparent when introducing the chosen platform.

As we expect autonomous robots that interact naturally, all created capabilities need to be implemented with onboard sensors and computation power. Especially mobile robots should not need to depend on any external hard- or software to be able to execute the everyday activity handover in all places.

Though I implement the system on a specific robot, generalization should be kept in mind during design to not limit the usage to one specific system. A modularization concept needs to be designed to adapt parts of the system for individual evaluation and exchangeability. On the other hand the possibility to transfer the system or at least parts of it to different robots increases the reuseability. This way, more robots could enhance their handover capabilities.

These properties of design and integration result in the following system requirement:

System Requirement 5(a): Robot Integration
The system needs to be integrated on a robot which itself creates requests upon the system.

System Requirement 5(b): Onboard Sensing and Processing
All perception has to be implemented with onboard sensors and computation for mobile and autonomous interaction.

System Requirement 5(c): Modularization
The system needs to be modularized for interchangeability of software parts, adaptability to different robots, and the possibility of individual optimization.

2.3.6 *Summary*

I propose system requirements in five categories building upon the stated research questions of my thesis. Together with the hypotheses, they form the foundation of this work. I further motivate these when discussing related work and my own object handover experiments. Subsequent I discuss how these affect my concepts and final implementation.

3 TERMINOLOGY AND CONCEPTS

This chapter starts with defining important terms and concepts in human-robot object handover as well as nonverbal communication (NVC). As there might be multiple definitions and interpretations that evolved over time, I define how these terms are used in the context of this work. Especially terms like robot, handover, and NVC have a wide range of meanings even in the field of computer science and robotics. Based on the literature I introduce my derived concept and further explain along with motivating the system requirement posed earlier.

3.1 ROBOT

As already stated in the INTRODUCTION, robot is a term first coined by Karel Čapek [Čap20]. The term itself originates from the Czech word *robota*, which can be translated as 'forced labour' [OLD:robot]. In Čapek's play *R. U. R. - Rossum's Universal Robots* the robots are biological copies of humans, which are used for hard work. This description does not fit today's usage of the term robot. The dictionary of the Oxford University Press offers five different definitions, which evolved over time, for the term robot with a variety of meanings [OED:robot]. These range from persons without emotions over traffic lights and flying bombs to software programs. This shows that a definition of the term in the context of this research is important for establishing a common understanding. Their first definition: "Chiefly Science Fiction. An intelligent artificial being typically made of metal and resembling in some way a human or other animal." [OED:robot] is quite similar to the one in The Oxford Living Dictionaries with "(especially in science fiction) a machine resembling a human being and able to replicate certain human movements and functions automatically." [OLD:robot]. The main difference between those definitions is the term intelligent added in the dictionary of Oxford University Press. The first definition in the dictionary of Merriam-Webster is "a machine that resembles a living creature in being capable of moving independently (as by walking or rolling on wheels) and performing complex actions (such as grasping and moving objects)" [MW:robot], which also is a definition that is compatible with the usage in this thesis. Based on the previous definitions, I use the term *robot* for any machine that can act on the environment by grasping and moving objects.

robot

The dictionary of Oxford University Press also suggests that dif-

the generic term robot can be specified with compound words ferent compound words can be added to the term robot to specify which kind of robot is meant [OED:robot]. Here, I narrow down the requirements for a robot in the case of an object handover. At first, the machine needs to be able to grasp and move objects. Therefore, an *end-effector (EEF)*, a tool to pick, is required by the robot. While there are multiple types of such tools, an impactive EEF is suggested, as this type is closest to being general purpose. Such a tool has two or more fingers to physically grasp an object by direct impact. Also, this type of EEF is similar to a human's hand and thus easiest to predict for humans during HRI. In general, a robot with similarities to the human appearance is called *humanoid robot* or short: *humanoid*. As we, humans are used to interact with humans as well as our ability to empathize with someone else makes interacting with such robots easier [Isho6]. This way, predicting movements as well as actions becomes straightforward compared to an industrial robot with a complex link/joint-structure. Another important category are robots which are able to emit social signals. Such *social robots* are able to create a link with their interactants beyond functional requirements. Object handover often involves scenarios in which someone is serving a human. A robot that provides such services for people, is defined as a *service robot*.

end-effector (EEF)

humanoid robot (humanoid)

social robot

service robot

For object handover tasks, moving from one place to another is needed to establish the required proximity between interactants. Such unaided moving from one place to another is called *locomotion*. A *mobile robot* that has a way of locomotion is able to position itself being accomplished by using wheels or legs.

locomotion

mobile robot

The definition of Dautenhahn for a *robot companion* also fits the kind of robot discussed in this thesis:

robot companion

> A robot companion is a robot that (i) makes itself 'useful', i.e. is able to carry out a variety of tasks in order to assist humans, e.g. in a domestic home environment, and (ii) behaves socially, i.e. possesses social skills in order to be able to interact with people in a socially acceptable manner. [Dau07]

The requirements of (i) match the definition of a service robot. Although the mobility of the robot is not explicitly mentioned, I would expect a robot that is able to carry out a variety of tasks to be able of moving around. In the second part (ii) of the aforementioned definition, the requirement for a social robot is given and I propose at least some features of a humanoid are also required.

As the focus of this thesis is NVC during human-robot object handover, multiple requirements on the type of robot are posed. A suitable robot for the topic of this thesis has to ...

- ... move, to position itself → mobile robot
- ... emit nonverbal social signals → social robot
- ... human-like appearance → humanoid robot
- ... sense and act on the environment → service robot

Based on the discussed properties, I term a robot that combines all requirements for this thesis a *mobile social humanoid service robot (mshs-robot)*. Such a robot is able to approach someone, hand an object, and emit the necessary social signals while doing so for a smooth and predictable object handover. Although the robot does not need to be a full humanoid with all body parts similar to a human, at least some traits need to be present. A head with eyes is mandatory to research the influence of gaze and the closer the arm is to human properties the more natural an object handover can take place.

mobile social humanoid service robot (mshs-robot)

3.2 HUMAN-ROBOT INTERACTION

human-robot interaction (HRI)

HRI is a multidisciplinary field of research

Human-robot interaction (HRI) is a term that does not only describe a situation where a human and a robot interact but is used to describe a multidisciplinary field of research. The International Conference on Human-Robot Interaction (HRI) is an annual conference in this area [hri:conf]. According to Goodrich and Schultz HRI is shaped by a number of disciplines: "[...] the disciplines that contribute to the field, such as human factors, robotics, cognitive psychology, and design." [GS07]. In the inaugural conference there even was an operating system for HRI presented which underlines the complexity for robots and their designers [Fon+06]. An analysis after the fifth year of the conference showed that it is a fast growing field of research [Bar11]. Such a broad range of research topics shows the challenge involved in researching a scenario in this field. The work presented in this thesis addresses parts of most of the mentioned topics. Human factors play an important role during object handover as this field is concerned with the interaction between humans and optimizes their well-being and the team performance. As the developed concepts are implemented and fitted for a robot, the robotics relation is apparent. Robotics even is an interdisciplinary field itself with multiple topics involved. In this work the most contributing areas are integration, behavior design, and control. Topics like electrical and mechanical engineering as well as hardware design were only addressed secondary to adapt the robot the way it was needed. Concepts from cognitive psychology are integrated, like attention and relevant methods are applied for evaluation.

Other classifications, I situate my work in, distinguish HRI by e.g. the distance or number of interactants. Goodrich and Schultz split HRI into two categories, remote and proximate interaction [GS07]. This work and object handover in general can be classified as the latter. The taxonomy for HRI of Yanco and Drury allows classifying this work into a category A interaction. In this class one human interacts with one robot in contrast to their other categories B-H, where larger groups of humans or robots interact with each other [YD02; YD04].

Another aspect deals with rating the performance of a robot in HRI. When a robot and a human perform a task collaboratively it is hard to measure the quality of the collaboration. Here, delay is an important objective measure to detect idle times and brake of fluency. Downstream measures like cognitive load or trust are not less important when designing collaborative systems [Hof13].

3.3 NONVERBAL COMMUNICATION

Watzlawick, Beavin-Bavelas, and Jackson established five Axioms of communication. Probably the best known is "Axiom 1: One cannot not communicate" [WBJ67]. This translates to: every behavior communicates, meaning that if people are aware of each other, they permanently communicate. Even the absence of actions might be interpreted as communication [Coa09]. *Nonverbal communication (NVC)* is the type of communication that is not expressed verbally by speech but by behavior. A multitude of such behaviors exist. The ones most interesting for object handover are those expressed by body language like proxemics, posture, arm motions, and gaze. Even subtle hand motions like a little grabbing motion in the air can communicate the information of readiness to receive an object [Lee+11]. This kind of communication has social properties and impact because it expresses feelings, emotions, relationships or other statements that influence how we interact. Such expressions are called *social signals* and can be expressed either directly or indirectly [PD10]. While such signals can be deliberate as well as subconscious, the term *signaling* is mostly used in the context of intentionally usage or expression of NVC [Ekm97; Pan+11b; DDP18]. It was found that information leakage applies for humanoids as long as the robot has enough human-likeness and performance. Also, it does not matter whether the interactants notice the nonverbal social signals or whether they are perceived subconsciously [Mut+09b]. Finding the correct behaviors and parameters for subconscious NVC is a tough challenge as you can not simply ask people about their preferences. Reading such signals from the interacting human might enable robots to adjust their own signal emitting [Mit+08].

nonverbal communication (NVC)

social signal

signaling

Besides the explicit design of gestures and gaze (see Sections 3.3.1 and 3.3.2) the robot needs to show openness and willingness to interact. System requirement 2(a): Body Orientation motivates that the robot turns towards the interactant with an open posture that shows the stated willingness to interact.

3.3.1 *Gestures as Information Channel*

Gestures constitute an essential part of NVC and thus can act as an information channel. A *gesture* is an intentional movement of the body or limbs as a means of expression [MW:gesture]. It was found that some gestures even cause the same brain activity as spoken language [Xu+09]. This gives a hint that NVC is able to transfer information similar to speech. The special case of sign language is often considered as speech instead of gesture research. Thus, the boundary between speech and gesture is hard to draw. *gesture*

A common categorization of gestures is between manual and non-manual gestures. While the former addresses motions of the hands and arms, the latter is about movements of other limbs, like nodding with the head. As object handover is done with the hands, it can be classified as a manual gesture. Another categorization for gesture are types like *lexical* (iconic) or *motor* (beat) which are the groups of gesture which support speech. A *lexical* gesture has a lexeme-like function. A giving or receiving motion can be directly used as a replacement for the lexemes themselves or to support them during speech. In contrast, motor gestures have no semantic meaning, they are mostly used to underline the rhythm of speech [Ken04]. The aforementioned types of gesture appear mostly together with spoken language, which is not addressed in this thesis.

Important categories of nonverbal gesture are the symbolic gestures (emblematic) and deictic gestures (indexical). *Symbolic gestures* can be a replacement for words. These gesture are conventional and might differ in cultures and context. With the example of giving and taking, a gesture can be used as a substitution for the words and is mostly understood with an object in the hand or holding the hand like there is one. A *deictic gesture* is used to reference something by pointing at it. It can be used supplementary to speech but also on its own [Ken04]. M.C. Caselli discussed the giving of objects by children as a deictic gesture. Considering the object handover task, a deictic gesture directly creates a reference to the given object. The object might also transfer additional information based on its type or context, like handing a closed box that the child wants to be opened [VE90, pp. 56–59]. Such interactions contain high-level information that require a lot of reasoning and context to understand the gesture correctly. *symbolic gesture* *deictic gesture*

I argue that the motion itself already contains information on the emblematic level, by communicating the intent of exchanging an object. Also, there is more subtle information like the location and time of the object handover transferred by the motion giving it also deictic properties. Thus, I conclude that object handover contains properties of a gesture, that can be understood by most people, while it still has a special role, especially when not only acting but actually performing the interaction.

While humans use gestures naturally, it is still a challenge to incorporate such a concept into robots. Riek et al. conducted a study to find out how people perceive three different gestures (beckon, give, shake hands) executed by a humanoid. They also compared abrupt (fast) with smooth (slow) motions and different orientations being frontal and sideways. For their smooth condition they limit the velocity to $\approx 250\,\mathrm{mm\,s^{-1}}$, while the faster condition had about five times average speed while executing the gesture. They found that gestures should not be too smooth to maintain their characteristic appearance. Also, vis-à-vis configuration makes it easier to recognize the gestures. This effect was strongest for the give scenario [Rie+10]. Zheng and Meng discuss the problem of transferring gestures from humans to humanoids because of their physical differences. Stressing the importance of testing the gestures with humans for evaluation, when designing gestures for humanoids [ZM12]. Pfeiffer and Angulo suggests to train humanoids to execute gestures by employing dynamic movement primitives. While *dynamic movement primitives (DMPs)* comprise a good tool to create rhythmic and discrete motions that can be adapted to the environment or goal, they can be hard to record and parameterize for a specific robot, especially on a joint-level [Sch06]. Thus Pfeiffer and Angulo recorded the trajectories in EEF space and calculated the joint-motions with inverse kinematics (IK). The generation of a self-collision free trajectory takes multiple seconds [PA15]. Although the authors report a successful reproduction of gestures, an evaluation with humans would be needed to check the recognizability of the resulting motions. Wang et al. present a different approach for incorporating gestures in human-robot object handover. They make use of a wearable sensor bracelet that tracks the motions of the human forearm. This data is used to classify the human motions into gestures that generate commands for a robot with the task to hand something over. Such gestures include signals like closer, faster, and stop. Their results show that humans tend to show a lot of variation when performing gestures [Wan+19].

dynamic movement primitives (DMPs)

Strabala et al. stress the importance of human-like gestures and cues for seamless object handovers. For a smooth interaction, robots are required that are able to show human-like social signals, as these are inherently easy to understand for humans. This way a consent of *what*, *when*, and *where* to object handovers can be established between the interactants. Additional non-human gestures can be used when human and robot share the meaning of those [Str+13]. Integrating such special gestures has the problem of needing to explain or negotiate them, which makes it harder for naive users to interact with the robot. Thus, sticking to human-like gestures makes first time interactions easier as there is no need to explain.

3.3.2 *Gaze for Mutual Understanding*

"Making eye contact is the most powerful mode of establishing a communicative link between humans." [Far+02] This quote expresses the importance of eye motion cues in interactions. Such eye fixations on an object or a person for a while are called *gaze*. Cook further specifies gaze to only occur when looking at the upper half of the face of someone [Coo77]. I suggest not to exclude fixations on other parts like the lips, the hands or even objects following the usage of Argyle and Ingham [AI72]. *Mutual gaze* is described as two persons making eye-contact by looking each-other into the eyes/upper part of the face [Coo77; AI72; KC69]. These gazes can follow distinct patterns. In social interactions a consistency of such reoccurring behaviors could be found. While there are some patterns exhibited by most people in a similar way, there are still individual differences [KC69]. Gaze is a term that is discussed in linguistics and communication. Humans do not only look at something to perceive it visually but also to show someone what they are talking about by directing the eyes on it as a pointer like rays that emanate from the eyes [Coo77]. Such *referential gaze* means the use of gaze to explicitly direct the attention of someone to something. This can happen during spoken interaction as well as in silent interactions between humans. If both (or more) interactants direct their gaze on the same subject a *shared gaze* is established.

gaze

mutual gaze

referential gaze

shared gaze

Shared gaze is the first step and lowest level of *joint attention*. The literature defines two additional levels of joint attention: *triadic* and *dyadic*. On top of looking at the same object, the dyadic level adds a basic form of understanding. For triadic joint attention it is required that there is understanding of the interaction partner, knowing that the other is focusing on the same thing or person [OG04]. Gaze also plays an important role in turn-taking. *Turn-taking* poses a term originating from a linguistic concept that describes when someone takes its turn in spoken dialog. This concept can also be applied to other forms of interaction where some actions can be taken either turn-by-turn or even with parallelism involved. Object handover has these turns as I discuss in Section 3.4.2.

joint attention

turn-taking

Looser and Wheatley showed that humans not only perceive the gaze but generally infer the lifelines of artificial faces with human properties based mostly on the eyes. The eyes contributed more than other facial features [LW10]. These results show the importance of gaze and the transferability of human behavior to animated faces of agents. It was shown that concepts of facial cues like gaze also apply for robots. People are able to tell when a robot looked at them and feel addressed, though they might have problems telling when it addresses someone else if the robot lacks human eyes [Ima+02]. Later, Imai, Ono, and Ishiguro extended the results to an artificial joint attention mechanism. With this system, gaze showed to be an effective way to draw a human's attention to the robot or direct it to an object. Also eye-contact showed to be important to create a form of joint attention between human and robot [IOI03]. Mutlu showed that manipulation of the gaze cues of a robot show significant impact on social and cognitive outcomes. Humans show increased task attentiveness, task performance, liking, and feelings of groupness. Though all humans seem to read gaze cues, humans that own pets had it easier to read the cues of robots [Mut09; Mut+09a]. Mumm and Mutlu implemented a mutual and averted gaze behavior on a humanoid and evaluated how it influences the physical as well as psychological distance humans keep to the robot. They found that it is important for the robot to establish some likability before seeking closeness with the humans. They also found significant impact on pet ownership and gender on the effect of a robot's gaze. Pet-owners and women react more positively to gaze cues [MM11]. It was also shown that gaze cues applied by a humanoid significantly improve the performance in cooperative tasks. The gaze of a robot can be read by a human to predict the target of the robot and thus decrease reaction time. If the eyes of the robot are covered by sunglasses, the performance only slightly reduces as the human is left to using the head orientation of the robot and thus loses some precision [Bou+12]. We also analyzed how gaze can be utilized to recognize addressing and use head orientation of a humanoid to give feedback. We could verify that mutual gaze is a meaningful signal for turn yielding [Ric+16].

concepts on gaze cues apply for robots

For object handover similar concepts apply as for turn-taking in dialog. Kirchner, Alempijevic, and Dissanayake conducted an experiment in which a single robot was interacting with multiple persons and had to deliver an object. They showed that the robot was able to apply NVC in the form of gaze to communicate to the group who should take the object by individualizing the group members into selected for object handover and spectator [KAD11]. Sisbot and Alami state that fixating the object with the robot's camera during object handover improves the legibility of the motion. They suggest that the ideal case would be a chain of motion protocols of the robot's gaze that establishes joint attention with the human. The robot should follow a pattern like e.g. looking at the object, the human and again on the object [SA12]. On the other hand, incorporating the detection of the human gaze into the decision-making of the robot during handover can decrease the rate of false-positive object releases by making sure joint attention is established beforehand. Tracking the eye gaze is still a challenge and decreases the reaction time of the robot and increases the false-negative object releases [Gri+13]. Strabala et al. examined the importance of mutual gaze before object handover and found that availability of asynchronous gaze is more important to communicate one's state and intentions. Especially for communicating the intent to exchange an object, gazing in the direction of the interactant is important [Str+12; Str+13]. Admoni et al. state that manual tasks like object handover might draw the attention of the human interactant away from the robot's face and thus its gaze by requiring attention on the hands or the object. They found that delays in the manual interaction draw the attention back to the face. People actively search for actions in the robot's gaze behavior and read its communication [Adm+14].This finding might lead to the assumption that until robots are as fast as humans, their facial cues are even more important.

asynchronous eye gaze is more important in object handover than mutual gaze

Moon et al. showed that a robot's gaze plays a key role in improving human-robot handovers. First, they conducted a human-human object handover study to find typical gaze pattern human exhibited during such interaction. With that information they implemented two gaze pattern for a PR2 (see Fig. 3.0(d)). In one, the robot slowly turns its head towards the future object handover position. In the other condition, the robot additionally gazes at the interactants face to establish mutual gaze which they call the turn-taking behavior. A comparison with a condition where the robot did not move its head revealed a subjective preference toward the turn-taking behavior. Anyhow, the task performance, measured by timing the interaction, decreased stronger for the shared gaze. Also, their results suggest that the gaze cues are more important for naive users, as novelty effects might cause the interactants to look at the head [Moo+14]. However, these results are limited to a robot without eyes and moving the head might be slow and less precise than needed. With the same systems and conditions Zheng et al. performed a video analysis of the HRI. Here they could confirm that if the robot looks ahead at the transfer position participants reach that location earlier. In their annotations they also found that 92 % of the participants looked at least once at the head of the robot during the object handover [Zhe+14]. In a later study the authors compared three different gaze strategies to further investigate gazing at the human face during object handover. Their first condition had the robot looking only at the target object handover position. The second condition had the robot solely looking into the human face. A third variation let the robot shift its gaze from the human face to the transfer position. In both conditions that exhibit a face gaze the participants gave higher ratings for likability and anthropomorphism, although with their implementation the shifting gaze caused a small delay in the interaction compared to the non shifting conditions [Zhe+15]. Nevertheless, Gharbi et al. found that the shifting is important. Especially a shift of the gaze at the end of an object handover from the object to the human's face signals that the robot is done.

Participants where observed to search for such an acknowledgment in the robot's face. Thus, pattern where the robot first looked at the face, then at the object and back at the face where perceived as more natural [Gha+15]. While the results are limited to a rather slow robot and the experiment was only done in a video study, it shows that the correct implementation of shared gaze and mutual gaze leads to a more natural interaction.

Whereas for only moving the head, a rough human position might be enough to look in the face, for a humanoid with precise eyes, the face position of the interactant needs to be detected by the robot. This motivates SR 4(d): Face Tracking for mutual gaze capable robots. It might even be possible that gestures are easier understood if the robot looks livelier or more human-like by creating gaze. Maybe even mutual reinforcement effects exist that cannot be considered individually.

3.4 OBJECT HANDOVER

The scenario that is researched in this thesis is a situation where a human and a robot exchange an object. As the word *hand* already has a long list of different meanings [OED:hand], the derived noun *handover* and verb *hand over* need to be defined in the context of this thesis. As an example for the variability of this term, emergency nurses do handover before and after their shifts and there is also a lot of NVC involved [Fra+12; EM16]. Thus, the exact same wording is used in a completely different scenario and hence needs to be further specified. According to the dictionary of Oxford University Press there are two main usages of the term handover [OED:haov]. "Telecommunications and Computing. An instance of handing over a connection or call from one base station, network, etc., to another" [OED:haov] Although this is a thesis in computer science the handover is meant physically and not in terms of connections. Thus, the second definition is closer to the meaning in this work: "The act or an instance of handing over a person or thing (lit. and fig.); spec. the transferring of power or responsibility from one country, administration, body, etc., to another, or the period during which this is done." [OED:haov] As well as the definitions of The Oxford Living Dictionaries and the dictionary of Merriam-Webster fit: "An act or instance of handing something over." [OLD:haov] "to yield control of" [MW:haov]. As these definitions keep open *what* is transferred, *object handover* is the correct term for the handovers described in this thesis. Where an object is something, that can be carried in one hand. From here on I use the shorter term handover subsidiary for object handover.

object handover (handover)

Another term found in literature regarding this research area is *hand-off* [MSS11; Lee+11; Str+12]. According to Oxford University

Press, The Oxford Living Dictionaries and Merriam-Webster this term is only used in American Football.

When at least one of the involved persons is moving by means of locomotion, meaning by either walking or running it is sometimes referred to in motion handover. Relay races are another example were a baton is passed while continuing to run. Some also call a handover, *in motion handover* when the arm does not stop while releasing or grasping the object [Hen+14; Hen+16]. The handover addressed in this work is the everyday activity which does not involve persons walking or running while handing over. Nevertheless, I address transferring the object in a non-static way in terms of the EEF motion.

in motion handover

3.4.1 *Classification of Handovers*

Handovers can be classified by who gives the object and who receives it. This is the most common type of classification found in literature. Current articles focus most of the time on either human to robot object handovers [Kaj+95; YRA13] or robot to human handovers [PAN17; Pra+14; HS15; Moo+14; BSW13; AT97] or sometimes both cases [MM05; Nag+98; Hen+16].

the giver has the object in the beginning, the receiver in the end

I propose a second dimension for classification: the intention. Intention might be even more important than who has an object in the beginning of a handover. Someshwar and Edan took a psychological analysis of the handover task to find differences in the perception based on the roles of giver and receiver. They found out that the giver and receiver role still change with the type of experiment, like putting objects in a lower or higher shelf [SE17]. This underlines the effect of intended state of the object in the beginning. Especially for the NVC it is of importance to communicate one's intent to the interaction partner. This type of discrimination separates between an active and passive party in the interaction. While the handover scenario is highly cooperative most of the time there is someone that initiates the process and another one that cooperates or follows.

there are four distinct HTypes

If one assumes that only one object is transferred, the distinction between giver and receiver is a binary decision and clearly decidable. With multiple objects involved there might also be multiple givers and receivers. Such situations can then be divided into a series of single-object handovers, which allows categorizing them individually using the aforementioned classifications. The division by intention into active and passive roles during a handover might not have the same clear boundary and is thus more complicated. There might be even cases in-between were both interaction partners are active or at least the passive one transitions from a passive state to an active state during the interaction. Nevertheless, it might remain important for the robot's behavior during the initiation.

I propose a cross classification based on who has the object at the beginning of handover and who has the iniative to transfer the object. I call the four resulting categories *handover types (HTypes)*. Table 3.1 shows the proposed cross classification of handovers. HType 1: I:H, G:H is the class of handovers where the human wants to give an object to the robot. To give an example, this could be a situation where someone hands over an empty cup to a robot for refill. In contrast, HType 2: I:R, G:H differs in terms of intent. This matches a situation where the robot has the task to fetch an object from someone, such as a waiter that has to collect all empty cups from guests. HType 3: I:H, G:R and HType 4: I:R, G:R only differ in the way that here the robot is the giver and the human the receiver, like someone tries to fetch a drink from a robot or respectively the robot moving around with the intent of handing out drinks.

handover type (HType)

		Initiative	
		Human	Robot
Giver	Human	**Handover Type 1: I:H, G:H** *Someone wants to give an object to the robot.*	**Handover Type 2: I:R, G:H** *The robot "wants" to take an object from someone.*
Giver	Robot	**Handover Type 3: I:H, G:R** *Someone wants to get an object from the robot.*	**Handover Type 4: I:R, G:R** *The robot "wants" to hand an object to someone.*

Table 3.1: The classification of handover types regarding the initiative and object possession in the beginning. The one that starts the handover process has the initiative. The one holding an object in the beginning is called the giver.

Carfi et al. recorded a dataset on human-human handovers and chose a similar configuration with a 2x2 experiment design. They also had the giver and receiver role as one discriminator. For the second dimension they chose who approaches whom [Car+19]. One could see the approach as the initiative, although I see a difference between being told to approach and someone having the initiative on its own. The problem here is that it is hard to design a study where only one participant has the intention. Such a mind state might be created but the absence of initiative in a study that is about handover might be impossible to create.

Another distinction between handovers present in the literature is whether the handover is direct or indirect. Indirect handovers do not involve a real interaction as the object is placed on e.g. a table by someone and then grasped by another one [Str+12]. Indirect handovers are much simpler than direct ones due to the absence of communication and synchronization, hence I focus on the direct handovers. Furthermore, there is the distinction between single-handed and bi-manual handover. Here, I focus on single-handed handovers while the second arm is free to be used for NVC.

In regard to configuration or scenarios there is a wide variety of options. Handovers while sitting or standing creates another two classes that influence the process. If the case where both people are sitting is considered, a third class exists. The sitting person becomes immobile and thus loses the ability to approach. The seating furniture also creates limitations regarding the way the other partner can approach. Other furniture like a table complicates the situation even further and needs to be considered for planning.

3.4.2 *Phases of Handover*

For the discussion of a process like handover it is important to create a common ground on the involved sub-processes partitioning the handover process into distinct phases enables us to add timings and measurements to the phases. In the current literature there are differences in terms of level of detail as well as the sub-processes that occur before or after the handover. If the focus is on who has control over the object, a simple segmentation could be: Person A has control over the object ⇒ Person A and B together have control ⇒ Person B has control. Huber et al. present a similar concept that segments the process into three parts: In contrast to the aforementioned definition the authors use the peak velocities of the giver and receiver for segmentation. These were used to measure the duration of each step. Their first segment was called the reaction-time which starts as one of the subjects reach their peak velocity. The next one is called manipulation-time and starts with the peak velocity of the second subject. The third phase (post-handover) is defined by the descending movements [Hub+08a; Hub+08b]. While this approach might be useful to assess the efficiency of the handover, the usage of global peak velocities can only be safely calculated in post. This fact makes it hard to employ this approach at runtime for realizing the handover task.

Cakmak et al. argue that a handover already starts with picking up the object and ends with retracting the arm while the actual transfer is a crucial moment. It might not seem obvious that picking the object is part of the handover, they show that there are preferences by humans on how to receive an object. To fulfill the user's expectations it might be useful to take them into account already at the time of picking the object [Cak+11a].

A technical analysis is done by Hendrich et al. who carved out seven distinct handover phases for a robot. Starting with 1) an arm motion to the user, 2) grasping of the object by the user, 3) detection that the object was grasped, 4) releasing of the object 5) user taking the object, 6) an arm motion away from the user, and in the final phase 7) the robot is idle [Hen+14]. In 2016 they presented a revised version of their phase-scheme for in motion handover: "(a): robot idle, (b): arm motion towards the user, (c): user touches and grasps the object, (d): gripper opens (and forces decrease), (e): arm stops, (f): robot stopped" [Hen+16]. In general the labels and phases changed only slightly but for in motion handover the boundaries become less strict. In their control scheme they added a step that observes forces during the trajectory execution that can trigger a soft-stop [Hen+16].

Moon et al. model the robot's behavior in five phases: grasp object, move gripper, wait for grasp of interactant, release object and return gripper [Moo+14]. As they focus on the gaze of the robot during the handover, they have all other steps hardcoded beforehand. Still, this approach proves to be valid and is another example for distinct phases with grasping as the first phase. Although they argue that their work is based on "Investigating Human-Human Approach and Hand-Over" [Bas+09], Basili et al. state that phases of handover blend smoothly into each other. They name that one can identify an approach, a lifting of the object and the handover itself but without being separate programs. Some phases even have an overlap that creates parallelism [Bas+09; Gla+10].

Suay and Sisbot state that there is a "generic flow of phases" while changes to the parameters of the process have an impact on the resulting interaction. The authors claim that a handover always starts with a trigger for initiation. Some form of communication is involved in starting the interaction, either verbal or nonverbal. This phase is followed by a hand movement of the giver towards the receiver. After that, the object is exchanged from giver to receiver. The finale phase is the post-handover where the receiver has (and uses) the object while the giver moves on [SS15].

A similar structure is proposed by Medina et al., who propose the phases approach, passing, and retraction. As they use a fixed mounted industrial robot, the approach only means the extension of the arm towards the user and not a positioning of a mobile robot. They present a system that tries to fluidly transition between those phases [Med+16].

Lee et al. proposed a three-phase model with: carrying, signaling, handing off. These phases contain different actions based on an adaptivity model, that adjusts parameters like position or even politeness. Their signaling phase has two sub-phases where the giver and respectively the receiver signal readiness [Lee+11]. Strabala et al. modified the phase-model of Lee et al. by removing the carrying phase in the beginning and adding a termination phase where the arms retract [Str+12]. Later they went back to a model that is closer to the proposed version of Lee et al. They now propose three physical phases: approach, reach, transfer. These three phases contain similar actions, where approach matches carrying, signaling matches reach, and handing off is equal to the transfer phase. Besides updating the physical phases they propose the concept of social-cognitive phases with *what*, *when*, and *where*. These concepts are lend from Clark's *Using language* [Cla96] on theories of common ground and joint activity [Str+13].

Vogt et al. describe four phases for a handover. The *Give* phase is where the movement of the hands towards each other happens. It is followed by the object transfer which they call *Hand Over*. This is followed by the *Retract* where the hands move back to their starting position and transitions into the *Idle* phase. They also chose to segment the whole handover trajectory into five segments which seems to be an artifact of their algorithm choice without semantic correlation [Vog+18]. Naming both the full process and the transfer part handover might be confusing.

Parastegari states that he uses the four phases presented in "Towards Seamless Human-Robot Handovers" [Str+13] [Par18; PAN17] Although Strabala et al. describe only three phases, the proposed four phases: "grasping the object, approaching the receiver, reaching out and transferring the object" [Par18] match the model of Strabala et al. except the added grasping as a first phase. In his work, Parastegari focuses on the last two phases, namely reaching out and transferring [Par18].

As the literature shows, there is a variety of different definitions of *handover phases (HPhases)*. However, they all have a common ground on the described overall structure. The biggest differences appear to be in the naming of the phases. Thus, one can already preliminary verify H 1: Handover Has a Distinct Pattern but I readdress it later on in my implementations and studies. I propose a combination of the previously discussed phases that covers the whole handover process starting with the acquisition of the object and ending with the retreat. This results in a total of five distinct phases with soft boundaries and fluid transitions. My definition of the human object handover phases, which are derived from the previously discussed literature, can be seen in Table 3.2. In the following chapters I further specify and discuss each of the phases individually in the context of human-human as well as human-robot handover. I go into details of the pre-handover phase HPhase 0: Acquire in Section 3.4.3, followed by the HPhase 1: Approach in Section 3.4.4. After that, I explain the HPhase 2: Reach along with their two sub-phases in Section 3.4.6 after introducing the object transfer point and discussing where the EEF has to be moved. In Section 3.4.7 I deal with transferring the control of an object between two individuals HPhase 3: Transfer. The process concludes with the HPhase 4: Retreat. These phases also motivate SR 1(a): Human-Like Pattern which aims at transferring this pattern to robots.

handover phase (HPhase)

handover has discriminable physical phases with soft boundaries

Handover Phase 0: Acquire
This phase takes place before the actual handover. It is special in the sense that it is solely executed by the giver. This acquisition can take place by picking or by a handover. This phase might also influence the receiver by watching the giver acquiring something. As research showed that the intent to handover influences the way the object is held, this phase is already part of the whole handover process.

Handover Phase 1: Approach
In the approaching phase the giver and receiver position themselves by means of locomotion in a way that they are in reaching distance. Who moves to whom is mostly defined by the initiative. The way of approaching signals the intention to the interaction partner.

Handover Phase 2: Reach
The reaching phase deals with the hand/arm motion while the body stays mostly stationary. In this phase both partners move their hands close. This can either happen consecutively by starting with a proactive movement and then waiting for the other one or at the same time. It can be divided into two sub-phases.

Handover Phase 2(a): Base Reach
The first part of the reaching motion has the goal of roughly reaching the object transfer point. This sub-phase has a bigger impact on NVC then the subsequent sub-phase.

Handover Phase 2(b): Adapt
Here, the fine adjustment of hand positions is done until both touch the object.

Handover Phase 3: Transfer
In the transferring phase the object is passed from the giver to the receiver. Both have extended their arms such that both are touching the object. Visual and force feedback can be used to perform a safe transfer of control over the object.

Handover Phase 4: Retreat
After the object is transferred to the receiver both interactants retreat their hands. This phase concludes a handover and both participants can carry on with other tasks.

Table 3.2: My definition of the handover phases. The resulting definition consists of five main phases and an additional two sub phases for the HPhase 2: Reach.

3.4.3 *Grasping for Handover as a Giver*

As a giver, to hand over an object, one has to be in possession of it in the first place as stated with HPhase 0: Acquire. This phase is often also called *picking*. I propose *acquisition* as there might be multiple ways to get hold of an object. One way can be even a preceding handover. There are two reasons why I count this phase to the handover process. The first reason is, that watching a giver acquiring an object might influence the receiver and the second reason is that the intended handover can influence the way an object is acquired. For tools like a hammer the final state is the handle in the hand and the hammer head correctly aligned. Also, the safety needs to be addressed by keeping dangerous parts like sharp edges away from the receiver. The minimal requirement for a giver is not to cover the whole object so that the receiver has enough space to grasp it.

Kim et al. propose three different approaches for preparing an object for human-robot handover. For one-handed handover they either directly consider the correct grasp site while grasping or if that is not possible, they propose a second approach which adds a step where the robot grasps the object with one EEF and corrects the grasp while the robot transfers the object from one EEF to another before finally handing it to the human with the correct orientation. The third approach uses a two-handed handover to grasp an object from both sides to leave enough space in the middle part of the object. This approach only applies to larger objects and makes the planning more complex [Kim+04].

Studies showed that users have preferred ways of receiving objects. Here, Cakmak et al. identified specific configurations for specific types of objects with e.g. handles [Cak+11a]. While they focus on finding the preferred state of the object during the transfer, they show that the giver needs to hold the object in a particular way, which was verified with a user study with 25 participants. Aleotti, Micelli, and Caselli even go a step further by considering the final state of the object in the hand of the receiver for an affordance-sensitive handover. To this purpose they perform a full three-dimensional (3D) reconstruction of the object in front of the robot and apply a priori knowledge on which part to present to the human receiver. Then the grasps are planned and executed maximizing user comfort [AMC12; AMC14]. Their system is implemented in a static industrial type setting with functional requirements instead of social aspects, in contrast to the work in this thesis. Lopez-Damian et al. presented an approach that decomposes objects to leave enough room to grasp the object during an interactive handover. The decomposed object approach calculates grasps for each half of the object [Lop+06].

Meulenbroek et al. showed in a study that observing someone during object manipulation influences the way this person interacts with the object. People were even able to estimate the object mass by observing it being picked [Meu+07]. This shows the impact of grasping to the handover.

Cini et al. confirmed that the giver [with intent] considers the receiver while picking an object. They designed a taxonomy based on "The GRASP Taxonomy of Human Grasp Types" [Fei+16]. With the three main types *power*, *intermediate* and *precision* and 28 (sub-)grasp types they classified how giver and receiver hold the object for handover. The *power grasp* clamps an object between fingers and palm, according to Landsmeer, Napier first coined the terms *power grip* and *precision grip* [Lan62; Nap56]. As opposed to this, a *precision grasp* clamps an object between the fingers and thumb. It could be observed that the giver used a precision grasp more often when someone has the goal to handover instead of using the object, where power grasp was preferred. The change in grasp type shows that the giver plans ahead to make object transfer and application easier for the receiver. A precision grasp of the giver provides more room for the receiver to take the object. There is also a tendency that receivers use the power grasp more frequently [Cin+19].

power grasp

precision grasp

a giver grasps differently when planning to handover

3.4.4 *Positioning for Handover*

The HPhase 1: Approach is concerned with reducing the distance between giver and receiver until an interaction distance is reached. Also, how the approach takes place nonverbally transfers information e.g. what the interactants intend to do and likely also the position and time. Handover requires to have a common *interaction space* which is established by the overlap of the peripersonal spaces of both interactants. The *peripersonal space* is the reachable area of a person. This area is important for object exchange.

interaction space

peripersonal space

Close encounters between human and robot can occur in a variety of scenarios. Besides handover, they can happen with e.g. a robot receptionist where similar concepts of social approaching behavior take place [Hol14]. While Holthaus' focus is the social behavior of a robot when being approached, similar strategies apply for a barkeeper robot that should be able to attract people in a social and pleasant way.

Focused interactions require interactants to be close to each other. In such situations they position each other in relation to each other. Kendon coined the term *f-Formation* to describe such systems [Ken76; CK80]. Ciolek and Kendon defined six basic types of configurations in which two interactants can be positioned to each other. They used the letters N, H, V, L, C, and I to describe the spatial layout. The lines of the written letters actually represent the positions of persons. The first three describe encounters where the interactants face each other though being shifted in the N formation and slightly rotated in the V-formation. An H-Formation describes the nearly straight arrange-

f-Formation

Space	Description	Distance	
1. Intimate	*touching, lovers, or close friends*	Close	<0.02 m
		Far	0.15 to 0.46 m
2. Personal	*interaction of friends or family*	Close	0.46 to 0.76 m
		Far	0.76 to 1.22 m
3. Social	*communication of acquaintances*	Close	1.22 to 2.13 m
		Far	2.13 to 3.66 m
4. Public	*public speaking*	Close	3.66 to 7.62 m
		Far	>7.62 m

Table 3.3: The four space categories as described by Hall with near and far phase [Hal+68; Hal69] converted from inch to m.

ment and is also called *vis-à-vis* [CK80]. While these f-Formation were originally designed to describe conversational scenarios, they equally fit to describe handover situations.

vis-à-vis

Hall established the term *proxemics* and categorized the interpersonal distance in four distinct categories [Hal+68; Hal69]. Proxemics is also known as a type of NVC (see Section 3.3). Table 3.3 lists his four proposed main categories, Spaces 1 to 4, with their corresponding far and close distances. This concept is a widely accepted way of describing and categorizing social interaction. As handover is a special type of social interaction with specific requirements, the individual handover phases can be assigned to the defined distance classes of Hall. HPhase 0: Acquire can happen at a distance, even at the Space 4: Public. HPhase 1: Approach always ends in the Space 2: Personal, where the subsequent HPhases 2 to 4 take place.

proxemics

Walters et al. conducted a study to check whether social approach distances apply to robot companions [Wal+05a]. Unfortunately the safety-system of their robot prevented interaction closer then 0.5 m. The resulting inability of the robot to reach Space 1: Intimate made approach distances of robot to human and human to robot hard to compare. Additionally, Walters et al. addressed the influence of personality traits on approach distances again. The main result was that people express, compatible approach distances to robots, even if those are not humanoids, as expressed towards other humans. Although, a part of the test group approached the robot closer than one would expect from human-human interaction. In regard to personality traits, they found that the more proactive a person is, the greater the human-to-robot approach distance is [Wal+05b]. Later they also chose handover as a scenario to cover the whole range of distances including close interaction, but with the same safety limitation of 0.5 m. Most of the participants let the robot interact with them in Intimate or Personal Space. For a frontal interaction a frontal approach was preferred. In constraint environments, like being seated, where the chair and legs block the robot from establishing a vis-à-vis configuration, an ap-

people express compatible approach distances towards robots as to humans.

proach from the side was preferred with a slight preference to from the right [Wal+06]. With the results of a follow-up study Walters et al. developed a *robot etiquette* that further specified approach directions in constraint settings. They recommend a distinction between sitting and standing when designing an approaching behavior for a mobile robot. For seated and backed against a wall persons an approach from the front left or front right is preferred over a direct approach. Someone freestanding prefers being approached from the front as this is the most efficient behavior [Wal+07]. Their experiments are limited to one interactant having the initiative. That means either the human or the robot approaches, mixed initiative scenarios might differ. In contrast, Koay et al. found most participants prefer a frontal approach, even when sitting. The preferred distance from the robot was on average 0.67 m. The preferences fall into two clusters where people with higher Intellect/Openness traits approach closer and Agreeableness personality traits keep ≈0.1 m bigger distance. The direction does also correlate to the distance. Another finding showed that the robot should start moving the arm just after reaching the personal space and not before as being approached with an outstretched arm is least preferred [Koa+07]. Based on these results Sisbot et al. presented a human-aware motion planner that positions the robot in a socially acceptable manner. Although this planner focuses on spoken interaction, the results can be extended to fit handover scenarios [Sis+07].

Basili et al. examined the human-human approach in a handover scenario. They asked participants to stand in a room, that had a high accuracy tracking system installed. Another person should then approach this person to hand an object. They started with a distance of 4.2 m. They found that some participants started moving the arm/hand at a distance of 2.16 ± 0.42 m. Approaching, in form of walking, stops at 1.16 ± 0.19 m interpersonal distance. No significant correlation to the height and/or arm-length could be found. Their results showed that in general humans approach in a straight line [Bas+09; Pha+07]. This matches the findings of Hicheur et al. who found out that human locomotion follows an optimality criterion (maximal-smoothness). They also found that the overall trajectory is more important than the individual footsteps executed by humans [Hic+07]. This allows detecting and predicting an approach to the robot. For the proxemics categories a handover approach can start in the most distant category (Space 4: Public) and comes as close as the second category where the actual interaction takes place (Space 2: Personal/interaction space).

In a situation where a robot is tasked to e.g. distribute brochures it becomes even more important to approach the receivers hastily. Strategies for social robots to initiate interaction become even more complex when the person to approach is moving and needs to be intercepted. Previous work already showed that in such situations a simple strategy like approaching and starting to talk has a very low rate of acceptance. Nonverbally communicating the intent to interact significantly increased the rate of effect [Sat+12]. In cluttered scenes or complex scenarios where a vis-à-vis configuration and a straight approach are not possible, alternatives need to be found. When the user is taken into account and simulated before the handover takes place, the robot is able to find a comfortable solution. The approach of integrating the mobility of the user into the planning of the robot might create effort in the first place but reduce the overall system and interaction effort by sharing the load between the interactants. A grid based sampling algorithm that takes human, robot, and the scene with obstacles might generate solutions that are not possible if the robot only plans for itself [Mai+12]. The notion of *mobility* can tune the amount of effort required by the human [Gha+13; Mai+12]. Waldhart, Gharbi, and Alami present an approach that even considers multiple agents in a graph-based approach. Here several handovers can be integrated to find a solution to bring an object from one location to another. Each human and robot is sampled in a number of places to rate the comfort and effort for them. In the end the planner presents a list of locations where the objects should be transferred for the solution with the least effort for the whole system [WGA15].

The work of Kruse et al. deals with socially accepted navigation and positioning in more complex scenarios for wheeled robots. They collect and summarize multiple works in this field that solve individual problems of navigation. According to the authors the combination of those algorithms and techniques remain an open challenge, especially for handover where the whole body movement needs to be considered and coordinated [Kru+13].

This section showed the importance of an interaction partner being able to position or orient itself in regard to each other. For such an ability on a robot, SR 4(c): Person Tracking is vital.

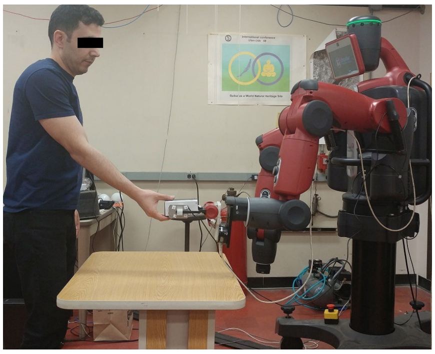
3.4.5 *The Object Transfer Point*

After the robot found a suitable position to execute the handover, it should reach out with the arm towards the interactant. Before I discuss the reaching motion itself (see HPhase 2: Reach), I want to specify where to reach. The actual exchange happens at a distinct point in space during the HPhase 3: Transfer. Shibata, Tanaka, and Shimizu conducted experiments and analyzed the motion towards the *handing point*. This point was found to be in front of the receiver [STS95]. Another term used to describe this point is the *position of the handover* [Str+13] and *handover position* accordingly [Hub+09a; SS15]. This term might be easily confused with the position that a person or mobile robot reached by means of locomotion (cf. Section 3.4.4: Positioning for Handover). Same applies for the *object delivery position*, which was defined as a static point at the height of the interacting person and a fixed distance [AMC12]. A reason for the different position terms might be the application in static scenarios, for instance, with sitting persons or immobile robots.

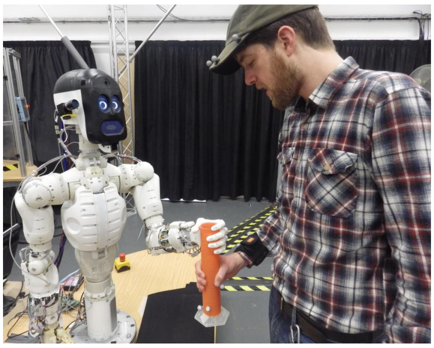
Pandey et al. investigate the spatial reasoning for proactive HRI. Here, cluttered environments enforce constraints on the handover. Their concept of *where* is used to generate feasible locations for the object transfer. This concept shall enable the robot to behave proactive in such cooperational tasks [Pan+11a]. Other work has also shown that humans have an a-priori expectation of where the exchange of the object will happen. This is of importance for a successful and comfortable handover. Basili et al. as well as Someshwar and Edan call this the *point-of-handover (p-o-h)* [Bas+09; SE17], which is located 0.65 ± 0.08 m in front of the giver [Bas+09]. Basili et al. also found out that the handover position is located closely between the midpoint of the giver and receiver with an offset to the right side of the giver (for right-handed persons) [Bas+09; Han+17]. While Someshwar and Edan's experiments show that there is an initial point which is decided upon subconsciously and might be updated during the interaction, humans might update their internal model that generates the point during multiple interactions with the same person. In general, both giver and receiver expect the handover to take place in roughly the same location as before [SE17]. This might be due to the scenario in front of a shelf which forced the receiver to focus on a second task during the receiving. The point of handover might not necessarily be decided only by the giver (called transporting subject) as stated by Basili et al. [Bas+09]. According to Huber et al. the point appears to be subconsciously negotiated by both interactants and stays mostly consistent during an interaction [Hub+09a]. Vahrenkamp et al. present an approach that analyses the interaction workspace of a human and a robot to find a good pose for handover. A rasterization of the workspace allows rating multiple poses in terms of reachability and quality. An implementation for the ARMAR-III robot (see Fig. 3.0(e)) showed to generate viable poses even for a person on e.g. a ladder [Vah+16].

humans have an a-priori expection of where to exchange an object with another human

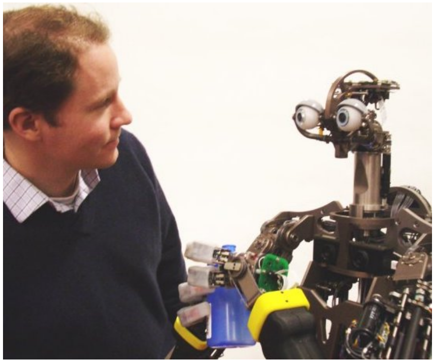
Koay et al. call the final position for the HPhase 3: Transfer the *handing over hand position*. In their experiments the preferred height of this position was determined as 0.789 m for sitting participants [Koa+07]. Cakmak et al. introduce a final object configuration (C_{obj}) that defines, together with the grasp (P^r_{grasp}) and base position of the robot (P^r_{base}), the handover configuration ($C^r_{handover}$) for the delivery position. Orientations of objects during handover used by humans were tracked and analyzed for efficient robot to human handovers. It was found that some objects have certain preferred orientations when being handed to someone else [Cha+15]. These affordances' deviations



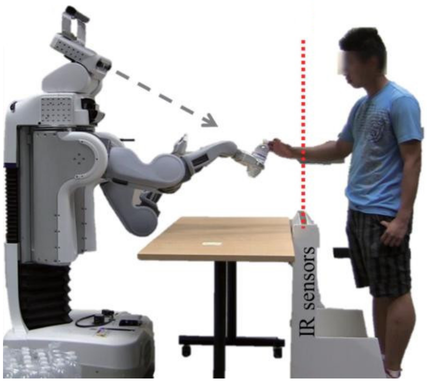
(a) Baxter [Par18]



(b) BERT2 [Web+16]



(c) Domo [EK07]

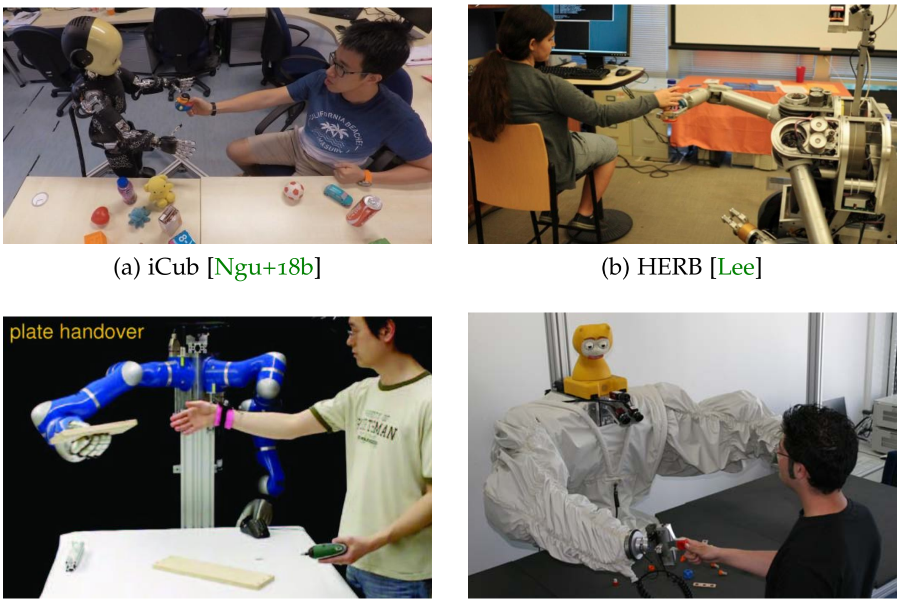


(d) PR2 [Zhe+14]

(e) ARMAR-III [Vah+16]

(f) HRP4R [Cha+14]

Figure 3.0: A collection of robots performing handover. Sorting is based on the pictures' aspect ratio. The citations reference the work addressing handover with the stated robots and not the publication of the robot alone. Some pictures have been slightly cropped to fit the grid.



(a) iCub [Ngu+18b]

(b) HERB [Lee]

(c) KUKA LWRs & DLR-HIT [Mae+17b]

(d) JAST [Hub+08b]

Figure 3.1: Continued Fig. 3.0. A collection of related projects of humanoid/bi-dexterous robots used in the context of object handover. Of the robots pictured here, only Fig. 3.1(b) comprises a mobile robot. Figure 3.1(c) has no facial features.

were shown to be reducible with a RANdom SAmple Consensus (RANSAC) based approach for increased human comfort [Cha+19].

object transfer point (OTP)

Sisbot coined and defined the term *object transfer point (OTP)* in his dissertation with the following definition: “In a scenario where a person A hands and object to another person B, we call ‘object transfer point’, the spatial point in 3D workspace where A and B will reach and hold the object together momentarily for its transfer.” [Sis08]. Recently the modified term *Object Transfer Position* was also used [PAN17]. Huber et al. use a definition, that specifies a name as well as a position: “The average handingover position lies close to the middle of the experimental table. The mean is slightly shifted toward the taking subject.” [Hub+09b] They argue that the OTP is shifted towards the receiver which I would argue against as I suggest it does not depend on who gives and receives but on who has the initiative (see Section 3.4.1). In their experiment the giver had the initiative and thus was a little faster than the receiver which explains the offset (see Section 3.4.1). This confirms the results of Basili et al. that the OTP is located closely to the midpoint [Bas+09; Gla+10]. In Nemlekar, Dutia, and Li’s research, which is based on the work of Li and Hauser [LH15], on predicting the transfer point, the author also uses the term *object transfer point* [NDL18; NDL19]. They also differentiate a static, a dynamic and an integrated OTP. I deduce my definition from Sisbot’s [Sis08] and extensions of Nemlekar, Dutia, and Li [NDL18].

Following from the above I define the different types of OTPs. The *static object transfer point (OTP$_{static}$)* is an a-priori prediction of such a point without incorporating the current motion. In contrast to this, the *dynamic object transfer point (OTP$_{dynamic}$)* is based on the current observations of a hand-detection module while the *integrated object transfer point (OTP$_{integrated}$)* combines the concepts of OTP$_{\text{static}}$ and OTP$_{\text{dynamic}}$ [NDL19], which is in line with the findings of Someshwar and Edan [SE17].

OTP$_{static}$

OTP$_{dynamic}$

OTP$_{integrated}$

3.4.6 *Reaching Out*

The reaching motion (see HPhase 2: Reach) is one of the most apparent topics in handover research. It is the step just before the actual transfer happens. It can be done by mobile as well as fixed robots and in multiple poses, being standing, sitting, lying, etc. Multiple aspects need to be taken care of when designing the reaching motion for a robot. These quality aspects include safety, speed, accuracy but also comfort, and predictability. In general this phase takes about 1.24 ± 0.28 s [Bas+09]. A nonverbal human-human handover experiment by Huber et al. showed that over multiple interactions of the participants, the timing might change in favor of faster interaction but the position and trajectory do not change much [Hub+09a].

The previous research regarding the reaching motion has two directions. One is collecting data of human-human interaction to understand how humans transfer objects among themselves. The other one is implementing algorithms on robots as controllers or by transferring/adopting trajectories from humans and for instance by means of machine learning. While some focus on the trajectory the EEF takes in the Cartesian space, I propose that this is only the absolute minimal requirement for a robot to reach the OTP. It is of equal importance to create trajectories that take the whole robot into account and thus have a human-like characteristic. The last distinction is whether the movement is adapted during the interaction. If there is no adaption during the process I call the OTP a *fixed object transfer point (OTP$_{fixed}$)* which is similar to an OTP$_{\text{static}}$ but does not require any prediction. That means, the robot is not able to sense the movement of the interactant and adjust its behavior during execution. Lorenz et al. already found that such an adaptability is vital in HRI for successful movement synchronization [Lor+11; LMH13]. As individual properties show a strong impact on the hand trajectory during handover, adaptable robots that do not treat everybody the same are needed [BMR19]. These findings motivate SR 1(c): Reactive Pattern. At the same time a general approach that generalizes well over all possible interactants is required to be able to interact successfully without requiring manual adjustment.

OTP$_{fixed}$

Kato et al. compared handovers of seeing and blind-folded participants to evaluate the contribution of visual feedback during the HPhase 2: Reach. The effect they found differs for the three directions of medio-lateral (X), antero-posterior (Y), and inferior-superior (Z). This shows that the correct reaching behavior and visual feedback help to establish successful handover. The authors found the gap between giver and receiver in the blind-folded condition to be smallest in the Y with ≈6 cm and the overall visual correction to be ≈10 cm [Kat+19].

visual feedback helps to establish handover

3.4.6.1 *Considering Only the End-Effector Motion*

The related work presented in this section considers only the motion and trajectory of the EEF to a desired Cartesian goal (see OTP). That means, not taking the whole robot into account while moving, but only the functional part of the handover. The following descriptions distinguish between work where an fixed object transfer point (OTP_{fixed}) is used and publications where the OTP is updated dynamically.

FIXED OTP Moving just the EEF to a predefined Cartesian goal without considering the arm motion only requires IK. It was shown though, that acceptance and predictability of robots can be improved by generating smooth and human-like motions: Legible trajectories during collaboration help to decrease the coordination time [Dra+15].

In 1995 Kajikawa et al. and Shibata, Tanaka, and Shimizu independently analyzed handover with robots [Kaj+95; STS95]. Both focused on EEF motion pattern in the horizontal plane. This reduces the problem to a two-dimensional task. There, a mostly linear motion could be observed in human interaction. Moreover, perturbations introduced by obstacles like a table between giver and receiver influence the trajectory. The velocity profile is bell-shaped with its peak just before half the distance [Kaj+95; STS95]. Later, Shibata et al. analyzed different velocity pattern for a one-dimensional robot and found that human-like bell shaped pattern gave the best ratings by study participants [Shi+97]. Jindai et al. developed an handover system with a two degree of freedom (DOF) robot with user-adjustable parameters like handover position for maximum comfort and peak velocity for reduced fear during interaction. The EEF was controlled with minimum jerk for a human-like bell-shaped velocity profile [Jin+03]. Later, they combined gestures and voice commands to adjust these parameters online [Jin+07]. Edsinger and Kemp present a robot that is able to receive objects from a human operator. The authors suggest detecting a face and then let the robot reach below that location. Thus, their OTP is close to the person's abdomen. When the arm reaches this OTP_{fixed}, and thus finished moving, a reduced stiffness of the series elastic actuator (SEA) allows the human to easily move the arm of the

reaching motion is nearly linear

robot which facilitates detecting the joint deviation and trigger an object release. They qualitatively showed that people are able to adapt to this strategy [EK07].

In 2000, Kajikawa and Ishikawa detected in a human-human handover experiment that the reaching motion can be split into two parts, which they call *Mode1 motion* and *Mode2 motion*. While the first part is straight and rapid it transits to the second part with a slight rotational movement and finally reducing the velocity to converge to the interactant [KI00]. While the target point was not tracked, they used a prerecorded EEF trajectory as input to their system, which was a first step in the direction of a dynamic OTP. Later they extended the results for a three DOF arm and compared the EEF motion to the trajectory of a human hand [KSO02].

reaching motion consists of two parts

Following these findings in human handover, I define the subphases HPhase 2(a): Base Reach for *Mode1 motion* and HPhase 2(b): Adapt for *Mode2 motion* during the HPhase 2: Reach.

Cakmak et al. conducted a study where users selected configurations for handover in a robot simulation. Final positions and orientations of objects as well as the robot's distance could be modified. Therefore, they precomputed all IK solutions in a discretized space in front of a (simulated) human. They outline that there are different optimal configurations for different objects but do not focus on the trajectory generation [Cak+11a]. In another study, Cakmak et al. also added temporal and spatial contrast to test whether poses that are easier to distinguish from the carry pose of the robot help it to move to that pose without the human interfering [Cak+11b]. While this showed to be a valid approach for robots that are not able to handle in motion handover, it lacks naturalism and introduces delay to the process. Sisbot and Alami employ a simulation of the human by calculating its kinematic with an approach called Inverse Kinematics using ANalytical Methods (IKAN) [TGB00] which they found to be fast in generating ergonomic postures. They also state that they optimize the whole trajectory towards the OTP in a way that the intentions of the robot are clearly expressed [SA12]. However, their planning procedure takes $\approx$6 s, which needs improvement for fluent interaction. Another approach making use of a simulation has been developed by Quispe, Ben Amor, and Stilman. It simulates both the giver and receiver to find a suitable solution for both interactants. Beginning with the midpoint between the interactants their algorithm samples the space and finally selects the configuration with the best manipulability [QBS15]. Dehais et al. use a human-aware motion planner by Sisbot et al. [Sis+07] that precalculates an OTP based on the human position. They compare three different motion types for the reaching motion. Best results for safety, legibility and comfort were achieved with smooth and not too fast motions [Deh+11]. Unfortunately they only analyzed three of twelve possible parameter combinations. This limitation prevents to attribute the individual contribution of each parameter. Hendrich et al. make use of the *MoveIt Motion Planning Framework* [MoveIt] to solve the IK while preventing self collisions to validate their force-based transfer algorithm. Although most of their participants where able to exchange objects with the robot system, they state that in the future they would suggest improving the system by tracking the hand pose of the interactant to react to the human's motion [Hen+14].

DYNAMIC OTP For a reactive robot the OTP needs to be updated to bring the human's and robot's hands together. For this purpose, systems need to track or even predict the human's hand position to react accordingly.

For a robot that is able to deal with more than the expected straight movements humans do typically, Agah and Tanie propose to employ a contention control architecture, in which multiple agents provide competing outputs that influence the robot's motions, can be applied to generate a reaching motion. Multiple competing agents can be used to create legible as well as most comfortable reaching motions [AT97]. With this system, Agah and Tanie implemented a reactive system, though all results were only tested in simulation (even the human) with arms that had only three or fewer degrees of freedom. Medina et al. presented an industrial robot system that makes use of a high precision external tracking system with markers attached to the object and the wrist of the human. This allows the system predicting an OTP and adapt the target during execution. Their algorithm based on dynamical systems applies torque to a torque controlled robot that converges to the target [Med+16]. An approach of synthesizing object receiving motions of humanoids based on a human motion database might create legible movements but might be hard to adapt during execution. Such a database-based approach was implemented by Yamane, Revfi, and Asfour, while the IK were approximated by exploring multiple configurations with their humanoid combined with an external marker-based tracking system [YRA13]. DMPs proved to generate predictable as well as reactive trajectories on an industrial object handing robot [Pra+13; Pra+14]. A notable advantage of this approach is the adaptivity of the trajectory during execution. In another study Koene et al. showed that timing might be even more important than position. Timing had more influence on the perceived safety than the actual trajectories [Koe+14]. Adaptive coordination strategies such as waiting for the human are often a trade-off between team performance and user experience [HCM15].

Another approach that relies on external tracking data generates one-shot learned triadic interaction meshes to generate a trajectory in EEF space. Therefore, Vogt et al. use multiple external Microsoft

Kinect cameras to get the body part positions of the interactant. This information is used to generate an EEF trajectory which is mapped to their robot with an unspecified IK solver [Vog+18]. Another approach with AR markers is used by Sidiropoulos, Psomopoulou, and Doulgeri, which feed the current hand position into a dynamical system (DS) that generates smooth EEF motions, trained from human-human handover wrist positions. First order IK are used to obtain the joint velocities based on the DS output [SPD18]. Park, Park, and Manocha present an approach for human-robot collaboration that predicts the future actions of a human based on depth data and a learning approach. They call the implementation of this approach Intention-aware motion Planner (I-Planner). It generates trajectories for a 7-DOF robot with a replan rate of 2 Hz. Their main goal is online collision avoidance during HRI [PPM19]. Kupcsik, Hsu, and Lee proposed an approach where an expert designed DMPs in EEF space that generate a trajectory towards the user's hand. In their approach, some parameters like the stiffness of the robot's joints during transfer were adjusted. Additionally, the distance when the robot switches from visual tracking to force sensing can be adapted by policy learning. The author showed that for dynamic handover, for instance, while running a reduced stiffness of the robot might help the human to grasp the object [KHL16]. Recently, Kshirsagar, Kress-Gazit, and Hoffman presented a controller that is synthesized from a Signal Temporal Logic (STL) formalism. With this technique they could easily implement different reaching strategies. Yet they only tested this approach in simulation and it remains open how it performs in the real world [KKH19].

3.4.6.2 *Full Joint Trajectory*

In this section, work is considered that takes the whole robot into account when executing the movement. This results in trajectories that either mimic the style of human-human interaction or aim at creating predictability in another way. Here the same distinction between fixed and dynamic OTPs can be made. While the former selects a goal and executes a trajectory, the later permanently updates the goal as a reaction to the interaction partner. It is standing to reason that the latter is the desired type of robot behavior, as it combines both the predictability and the adaption to the human. This motivates SR 2(b): Gesture Motion as the authors, of work discussed in the following, consider the NVC aspect in the arm's reaching motion to be important.

FIXED OTP Huber et al. compared a minimum jerk profile in joint space with minimum jerk in spatial coordinates. Although they state that their resulting trajectories are not human-like, considering the joint trajectory is important. They found shorter reaction times in the interaction for the minimum jerk profile in joint space. In their experiments they did track the human hand but the arm trajectories where precalculated and fixed during all interactions [Hub+08b; Hub+08a]. Rasch, Wachsmuth, and König analyzed the motion of humans during handover to transfer it to humanoids. The authors recorded not only the wrist, but also shoulder and elbow movements [RWK17]. With that data, they determined a joint motion model (JMM) for robots that aims for a human-like appearance of the handover reaching motion. A comparison with a linear joint space trajectory showed that humans could only tell a difference if they were made aware of them, by telling them to look for differences in movements of the arm and joints. The authors also compared the perceived differences by implementing it in two different robots. On a humanoid Pepper robot the noticed difference was even smaller than for an industrial type robot. Nevertheless, when difference is recognized, the JMM was rated as safer and more human-like [RWK18].

DYNAMIC OTP Having dynamic transfer points that are adapted to the human while maintaining a human-like and natural appearance can be considered the desirable goal for human-robot handover. Only few have tried to solve that task. Nguyen et al. present a system on a humanoid (iCub Fig. 3.1(a)) that uses onboard 3D markerless perception. With this system, they were able to react to the human as well as the environment. Although the presented system operates in the Cartesian space, they claim to incorporate an IK library that generates human-like joint motions [Ngu+18b]. This IK library is specially designed for the iCub robot (see Fig. 3.1(a)). The central idea is based on "Reaching with multi-referential dynamical systems" [HB08], which employs the Vector Integration to Endpoint (VITE) model [BG88a; BG88b]. This biological inspired model is likely to generate human-like trajectories in Cartesian- as well as the joint-space. They compared it with a dynamical linear system and minimum jerk based IK. As the authors did not get to run and test the implementation of the VITE algorithm as the computational complexity of the solver prevented an online application, the performance remains open [Pat+10]. For the handover task the controller and IK were combined into a single formulation that generates joint velocities. It also aims for collision-free interaction and even incorporates tactile information to handle post-collision situations [Ngu+18a]. Though Nguyen et al. present an interesting approach and aim at human-like arm motions, it remains unclear how natural and human-like the joint motions of the resulting system are. Also, they do not predict future hand positions of the interactant. In their experiment the adaption distances of 0.2 m are rather short and it takes about 2 s for that distance [Ngu+18b]. Likely, the authors only addressed HPhase 2(b) and skipped HPhase 2(a) by having the robot's arms reach out right from the beginning. Maeda et al. make use of Probabilistic Movement Primitives to teach movements to the robot that can be adapted during execution. This approach yields joint trajectories for the robot that correlate to the Cartesian trajectory of the human interactant [Mae+17a]. As they use two industrial Kuka LWR robots (see Fig. 3.1(c)) the human-likeness is limited. Recently, Pan et al. presented a system implemented for a Kuka LBR where they added a bear head to the second joint. This resulted in a character-like robot with 2 DOF for the head and 5 DOF for the arm. Here they also followed a two stage approach with a precomputed and updatable trajectory which receives the object pose with an external marker-based tracking system. With this robot they tested different delays and speeds in regard to how the interaction is perceived by humans. They found that moving the robot faster than the human as well as having no delays at all might cause discomfort [Pan+19].

3.4.7 *Transferring Physical Control*

After the hands/EEFs found together in space and time the transfer needs to happen. In the HPhase 3: Transfer, the object is transferred between subjects' hands. Hence, the robot needs some kind of sensing when to release/grasp the object. When objects are transferred there is always forces involved, being it gravity "pulling" on the object or the interactant applying force by pulling/pushing the object. These forces can be measured in the robot's joints or with tactile sensors on the EEF's surface. So the questions are, when, who and how much force is applied in this phase of handover. Additionally, it needs to be decided when the giver can release the object. It was found in experiments that it is likely that humans use a combination of visual, tactile and force sensing in that process. Regulation and adjustment is performed based on the sensed object mass and surface properties [KI92]. The grasp force also varies on a trial-by-trial basis even within subjects and anticipatory forces are applied on an initial handover [MM05].

Existing approaches often make use of force/torque-sensors (FTSs) in the wrist to sense when the human applies force to the robot's EEF either by pulling on the object or by pushing it into the EEF of the robot [BSW13; Cha+13; HS15]. This motivates SR 4(e): Contact Detection as a basic requirement. More advanced approaches add optical or tactile sensors in the gripper to optimize grip-force control and contact detection [Hen+14; Hen+16; Par+16]. Nagata et al. already presented an approach for a robot to position the fingers of a robotic EEF such that a stable grasp is established. This system was only able to grasp objects that were put directly in the gripper [Nag+98].

Chan et al. designed a baton equipped with an FTS in the middle and a force sensing resistor on the bottom and top. It was found that the giver and receiver employ similar control strategies during the transfer. They also found that the duration from first contact to full load transfer takes about 0.500 s. So this would be the ultimate goal to achieve for human-robot handover [Cha+12]. The same baton was used to evaluate a force-based transfer controller implemented for the PR2 robot (see Fig. 3.0(d)). The results show that the release threshold should not be larger than 40 % of the object weight as humans do not pull stronger than that [Cha+13]. A later implementation on the HRP4R humanoid (see Fig. 3.0(f)) without a FTS made use of the elbow position joint error to estimate the load force. This approach required a higher threshold of $42.3 \pm 15.0\,\%$ to be successful. Nevertheless, the system was only evaluated with a single subject [Cha+14]. Controzzi et al. build a similar baton to show that humans use visual-feedback to trigger the release of an object. This was done by blindfolding participants in some runs and measuring the changed grip force with the baton. They also showed that humans adapt their behavior based on the interactants movement speed when giving an object [Con+18]. Han and Yanco implemented three different release policies for the Baxter humanoid (see Fig. 3.0(a)) with a touch sensor added to the gripper. Two of them were threshold-based. While one of those was active when the EEF reached its goal, the other was already active during the movement (in motion handover). The third policy was implemented by an approach that observes the forces gravitational component and triggers on a drop to detect a load transfer. They call this approach *proactive release policy* where the data is smoothed over 180 ms in a moving average for 40 windows. If 35 % of the windows are decreasing, the object is released [HY18]. In an HRI experiment they found the third policy to be perceived best by participants [HY19]. The authors used a rate of 1000 Hz while in the other policies only 4 Hz were used. This introduced a delay of 250 ms for the classic threshold-based and 220 ms for the moving average window-based approach. Also, it remains open how they set the thresholds.

force threshold depends on objects mass and should be as low as possible.

Hendrich et al. confirmed in a touch and force based setup that users might prefer force thresholds that depend on the mass of the object. They also state that in the future visual perception might be required [Hen+14]. In a later experiment they added that the threshold should be as low as possible without the robot dropping the object. For in motion handover this requires precise sensors and a model of the robot that removes self-induced measurements of the moving robot. Here also higher thresholds are excepted on heavier objects [Hen+16]. In a learning approach where the slope of a threshold function was learned over multiple interactions, the resulting threshold was about 80 % of the object mass [KHL16]. This quite conservative value might have been introduced due to the in motion handover that applied forces on the wrist sensor just by moving the robot. The resulting robot controller was rather compliant which allowed the robot following the human motion over some distance. Also, they collected only five trials which might not be enough to give a good generalization for everyone and especially naive interactants. Medina et al. refrain from a threshold and use a load force transfer function that gradually transfers the load from giver to receiver. With their industrial type robot they achieve precise measurements and decomposition of forces on the wrist mounted FTS into internal and external forces. Their evaluation, although only performed with the author, showed that the passing time, where both interactants support the object, could be reduced by 0.1 s [Med+16]. Sidiropoulos, Psomopoulou, and Doulgeri followed a similar approach for a giving robot. Here, again an FTS in the wrist was used to detect when the load is zero to open the EEF of the robot to transfer the load from robot to human. They state that this approach still has some risks and thus suggest adding visual perception in the future for sanity checks [SPD18]. A different approach by Singh et al. made use of a vibrator attached to one side of a robot's EEF and measure the vibration on the other side with an inertial measurement unit (IMU) for successful detection of grasp/handover [Sin+18]. This approach has the disadvantage of needing to produce vibrations on the robot which might lead to unnatural interaction and might strain the robot.

All the force-based handover approaches expect the user to actually apply force above a static or dynamic threshold. In a preliminary study Chan et al. discovered that this is not the case for all interactions and decided to instruct the participants to pull on the object until it is released [Cha+13]. Although such instructions might be a good approach to compare a set of algorithms, the need to get handling instructions for an interaction with a robot contradicts the anticipation of natural human-robot handover. Which motivates SR 4(f): Visual Transfer to enable natural handover for everyone (see SR 3(a)).

humans do not always apply force in HPhase 3: Transfer

3.4.8 *Laterality and Handedness*

As this thesis discusses one-handed handovers, it is important to discuss which hand to use primarily. While SR 3(a): Understandable by Everyone states that the resulting system should be usable by everyone, most humans have a primary or preferred hand to which the system defaults to. Still, they have no problems exchanging objects with someone having other preferences without trouble. Klußmann et al. stresses the importance that especially at work one should not be forced to work with the non-primary hand. They suggest that machines should be able to be operated from both hands e.g. by making them symmetric [Klu+14]. A humanoid robot with two arms offers the possibility to provide such a symmetric interaction. Koay et al. investigated preferences for handover in terms of approach direction and final position. Most of the participants preferred a transfer directly at the front [Koa+07]. This shows that either with the right or left manipulator a symmetric handover is possible. "An International Study of Human Handedness: The Data" showed that only 5.9 % are strongly left-handed [PE94]. Thus most humans are used to interact regularly with right-handed people.

3.5 SUMMARY

The review of human-robot object handover approaches and the influence of nonverbal communication revealed that this topic is important and widely addressed in the human-robot interaction research community. While most approaches address only parts of the handover interaction, by taking shortcuts in some areas, I focus on approaching the topic as a whole from HPhases 1 to 4 and back. This means that I develop, implement, and validate concepts on a mobile social humanoid service robot that comprises most of the human capabilities and properties like mobility, two arms, a head, a torso, and without the need for any external sensing, computation, or artificial markers. While such a robot might not reach the human capabilities in terms of speed and precision, it allows examining the whole spectrum of human handover and the involved nonverbal communication. Moreover, I do not distinguish by who is giving and receiving instead, but I use the more suitable concept of determining who has the initiative.

Part II

THE OBJECT HANDING ROBOT

This part describes the system, modules, and experiments developed during my doctoral studies. It starts with introducing the mobile social humanoid service robot Meka M1 Mobile Manipulator as platform basis, followed by an initial experiment that evinced further system requirement. I present three additional contributions in form of a gaze behavior, an improved reaching motion module, and an object transfer point prediction. After demonstrating individual results, I discuss my combined behavior and its evaluation.

4 SYSTEM FOUNDATION AND ENVIRONMENT

This chapter describes the hard- and software the system was build on, as well as the environment it was evaluated in. The system requirement (SR) 5(a): Robot Integration demands to define the environment and the constrictions by the selected robot. It also sets the basis for the SR 5(c): Modularization by showing where individual hard- and software components can be combined or exchanged for different levels of object handover (handover). Also, these modules aim to be reusable on different robots.

4.1 MEKA M1 MOBILE MANIPULATOR

I chose to use the *Meka M1 Mobile Manipulator (Meka M1)* (see Fig. 4.1) as base for my implementations and experiments as it fits the description of robot companion and mobile social humanoid service robot (mshs-robot). The Meka M1 is a robotic platform created by *Meka Robotics* [Meka:robot]. It comes in different configurations and appearances. The version used in this thesis mainly consists of a mobile base, a z-lift, a torso, two arms with hands, a sensor head and a computation backpack. Up from the "hip", which is the mounting to the z-lift, the robot is a humanoid robot (humanoid) in the sense that the proportions and degree of freedom (DOF) are similar to a human adult. Most of the joints feature *series elastic actuators (SEAs)* to improve human safety by decoupling the motor inertia from the link [PKM02]. This provides also an improved shock tolerance which protects the drives and gears of the robot from damage by external forces [Meka:tec].

Meka M1 Mobile Manipulator (Meka M1)

series elastic actuator (SEA)

The mobile base is a *B1 Omni Base* [Meka:base] with a computation backpack and a prismatic z-lift from Festo to move the torso up and down. The *Holomni* powered caster wheels allow omni-directional motion. Two integrated battery banks provide 24 V for mobile operation. An onboard real-time personal computer (PC) (using *The RealTime Application Interface for Linux* [RTAI]) is responsible for the control of the robot. It runs *M3 Core* [M3 Core] realtime control and sends the control messages over Ethernet for Control Automation Technology (EtherCAT) [JB04] to all the actuators and collects the data from all torque and position sensors. For processing of perceptions and higher level control of the system two additional off-the-shelf PCs in the mITX form-factor were added. One of them contains an *Nvidia GTX960M* graphics processing unit (GPU). Two 2D light detection and ranging (LiDAR) scanners provide distance measurements of the surroundings. A *Hokuyo UTM-30LX* mounted at the front and a *Sick TIM781* mounted at the back of the robot together provide a 360° perception around the base. This can be utilized to track interactants and perform simultaneous localization and mapping (SLAM).

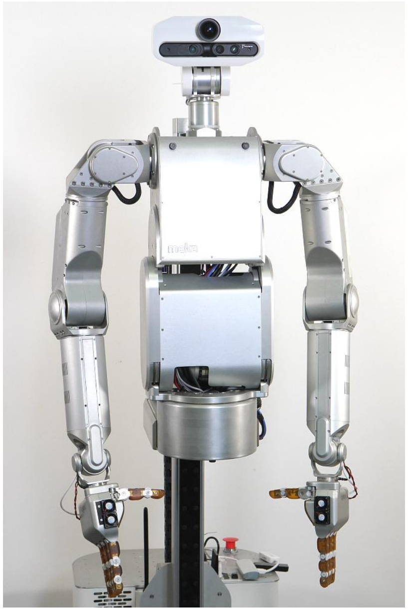


Figure 4.1: The humanoid robot Meka M1 Mobile Manipulator (Meka M1), utilized in this thesis to study human-robot handover, pictured in a portrait.

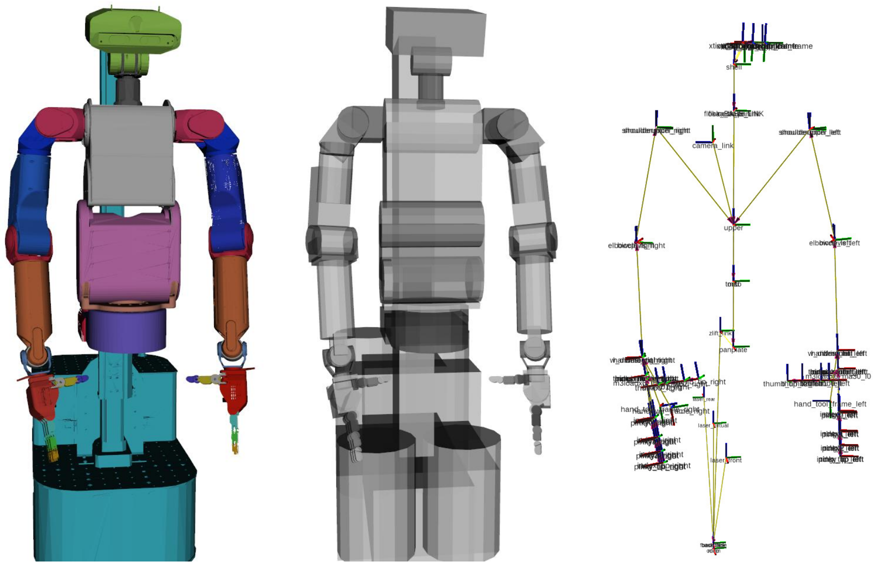
The torso (*T2 Humanoid Torso* [Meka:torso]) has two controllable DOFs and an additional coupled hip joint. This allows the torso to rotate sideways as well as to bend back and forth. Zero backlash harmonic drives allow precise control without slack. The load cell based torque sensing and control allows compliant actuation. A sensor head (*S3 Sensor Head* [Meka:head]) incorporating two of the harmonic drive driven joints allows panning and tilting of a Primesense Carmine 1.09 short-range RGBD-Sensor [Pri12]. Two *A2 Compliant Arm* [Meka:arm] are attached to the upper part of the torso. They consist of seven DOF SEA with near human appearance and workspace. They also provide torque control at every joint for safe and compliant interaction. The singularity free roll-pitch-yaw wrist has a *6-Axis Force and Torque Sensor (Mini40 Series)* [ATI:Mini40] force/torque-sensor (FTS) attached. To each of the FTSs a hand/end-effector (EEF) of the type *H2 Compliant Hand* [Meka:hand] is attached such that the FTS measures forces between hand and arm. Contrary to a human hand this robotic hand has three fingers and a thumb due to size reasons. The fingers of that hand are underactuated, thus they can be only actively closed with a single actuator per finger to actuate three joints. The thumb additionally features abduction with a second actuator. The opening of fingers is implemented by an elastic band running on the outside of them. Making it a total of five SEAs for twelve joints compared to the 27 DOF of a human hand, it significantly reduces the possible complexity of grasps. While precision grasps without incorporating the palm when grasping might be possible, the robot can not exploit the gained movability of the object as the fingers do not allow lateral movements. Thus, power grasps are the preferred grasp type for this EEF. The fingers are made of urethan which is stretched on closing and used to pull open the hands. We added a Dold Safemaster RE 5910/001 [E D12] with four two-step push buttons. This wireless controller is not only used for safety reasons as an emergency stop but the additional buttons provide input to control the robot during human-robot interaction (HRI) experiments.

4.2 SYSTEM ARCHITECTURE

The actuators of the Meka M1 are connected to an integrated EtherCAT bus [Meka:tec]. The control PC is connected to the same bus, which runs with 4 kHz. It synchronizes the data over shared memory with a semaphore between the bus driver and the *m3rt_server* which is updated at 1 kHz. It controls the 155 components [m3bie] running inside the RTAI user space. In this software the torque control, joint control and gravity compensated, compliant joint position control is implemented. This functionality is exposed with *ROS Control* [ros_-control] to the Robot Operating System (ROS) eco-system via the *Position Joint Interface*. The *Robot Operating System (ROS)* is a collection of tools and software to interface and program robots [Qui+09]. It is used extensively in this thesis to foster reuse of concepts and modules. Inside ros_control, access is provided to the gravity compensated joint position via a joint_trajectory_controller (JTC). The position of all the joints is provided via a *joint_state_controller*. Additionally, the stiffness of the joints can be adjusted between 0 and 1. Zero means that the robot is completely compliant and does not try to reach the requested position at all. This mode can be used to e.g. teach new positions to the robot. A value of one sets the robot to actively reach the desired position without losing the advantages of SEA. Values in between allow joint position control with increased compliance.

Robot Operating System (ROS)

A *Unified Robot Description Format* [ROS URDF] based description of the Meka M1 allows creating a virtual representation of the robot. Figure 4.2(a) shows a visualization of the internal representation of the robot's state. Additionally, the description in the ROS URDF allows specification of collision geometries that are attached to the links of the robot (see Fig. 4.2(b)). These can be used to calculate self, as well as collisions with the environment. The *Transform Library* [tf] is used to create a tree of all transformations generated by joints and links of the robot (see Fig. 4.2(c)). Instead of using the data of the robot, *Gazebo - Robot simulation made easy* [Gazebo] was integrated to simulate sensors and actuators to a certain degree. In the case of handover



(a) Virtual representation of the Meka M1, used for planning and simulation.

(b) The collision model of the Meka M1 constructed from primitive shapes for efficient collision checking.

(c) *Transform Library* [tf]-tree of the Meka M1.

Figure 4.2: The virtual representation humanoid Meka M1 (cf. Fig. 4.1) as visual and collision model as used for simulation and internal representation for motion planning. All images are taken from the same point-of-view.

the simulation of human interactants remains an open challenge and thus can only partially be employed.

4.3 REPRODUCIBILITY AND SOFTWARE DEPLOYMENT

To foster reuse and allow reproduction of my experiments, I integrated the software into the *[Research & Robotics] Development Toolkit (RDTK)* [RDTK]. It builds upon the concepts of the Cognitive Interaction Toolkit (CITK). It e.g. features software deployment for heterogeneous, component-based software systems, following the continuous integration paradigm [Lie+14]. It was shown that such a tool-chain enables documentation and reproduction of robotic experiments [Lie+17].

The base system of the robot utilized in this thesis already requires multiple software libraries and programs. Following the requirement of SR 5(c): Modularization I present additional components. The RDTK offers and requires description of artifacts in a specific repository in order to aggregate all required components for an experiment. There are two natures of such descriptions. One defines a single component, library, or configuration by stating where to get the required code, the required dependencies, versions, a description, and how to build it, if required. Building of software is supported via templates that already provide the most common build types and repositories for code acquisition. Besides the recipes and supporting templates there are distributions. A distribution is a collection of recipes and specified versions. Such a description defines a software system in regard of versions and requirements. With these information combined, jobs for *the Jenkins project* [Jenkins] continuous integration server are generated. This server can be run on the robot or an external machine.

For this thesis I added recipes for all integrated components as well as created a distribution.[1] I deployed a Jenkins with build jobs of all components to the robot as well as a dedicated server to make sure having the requested versions on the robot installed and providing them for workstations for e.g. simulation. The whole distribution was build each night to make sure the repositories are still reachable and compilable. Additionally, a catalog-style web view is generated that aggregates all required information user-friendly. For keeping hold of all the components running simultaneously on the distributed system, I integrated *VDemo* [VDemo] for controlled starting, stopping, logging and checking of components with a single graphical user interface.

4.4 ENVIRONMENT OF THE ROBOT AND EXPERIMENTS

The Meka M1 is embedded in the research project *Cognitive Service Robotics Apartment (CSRA)*. It is about a research apartment (see Fig. 4.3 located at the Cluster of Excellence Cognitive Interaction Technology (CITEC). It features 24-7 cognitive interaction technology for interaction research. Various sensors and actuators allow interacting with the environment. The robot and thus most of the experiments described in this thesis are located in the room called *Gym*. We conducted a study to investigate how users address such an environment or the robot to achieve goals. The resulting multi-modal data-set is also available upon request for research purpose [Ber+16; Hol+16].

Cognitive Service Robotics Apartment (CSRA)

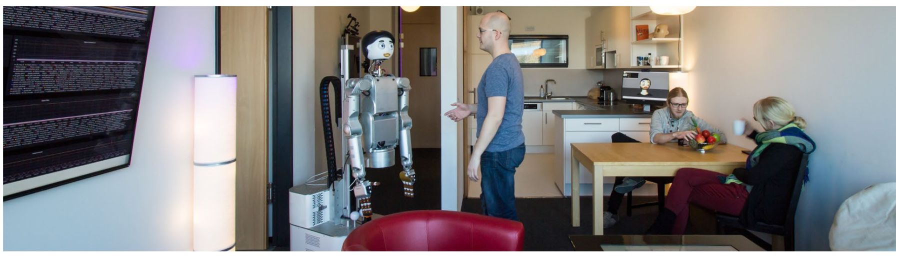


Figure 4.3: View of the living-room of the CSRA. Showing some guests and its inhabitant Floka [Wre+17].

1 https://opensource.cit-ec.de/projects/citk/repository/entry/distributions/meka-nightly.distribution

4.5 INCORPORATING MIXED REALITY

With "Improving Human-Robot Handover Research by Mixed Reality Techniques" we already presented methods to improve human-robot handover [Mey+18]. Parts of this chapter are based on the results and I discuss them in the context of this thesis. Close interaction scenarios like handovers, conducted in the real world, are limited in regard to safety concerns, a fixed body structure, as well as overall performance of the robot in the physical world. Another inherent problem with real robots is the availability of them for e.g. reproduction of experiments. Especially for handover there might be different behavior in different locations due to cultural effects. Thus finding such differences is complicated because of hard to transport hardware. Using entirely virtual environments for highly controlled experiments generates a new range of possibilities with respect to reproducible research. In such immersive experimentation environments even physical limits can be overcome.

We explored several mixed reality (MR) techniques to overcome such physical limitations. MR glasses, such as the *Microsoft HoloLens*, allow transferring information from the internal representation of the robot to a visualized counterpart displayed in the field of view (FoV) of the interacting person. This can help to better comprehend and anticipate the robot's behavior. Additionally, such devices allow augmenting real environments with simulated robotic parts like virtual robot heads or arms attached to a physical robot. This allows conducting HRI experiments between the worlds that currently can not be done in either the real physical nor the virtual world alone. Moreover, robots can be simulated in an immersive virtual reality (VR) environment to conduct highly controlled experiments.

4.5.1 *Mixed Reality Human-Robot Interactions*

There have been proposals for various applications of VR and augmented reality (AR) in the area of HRI. Robotic sensor data visualization with MR techniques are already applied in the area of robotics software development and debugging [Sti+05]. To this end, laser scans and point clouds from stereo cameras, navigation and path planning costmaps, as well as planning of footstep were visualized for a robot [Nis+08]. A tool to overcome the knowledge gap between experts and naive users was proposed by Renner et al. with a system to augment the user's environment with data visualizing the robot's perceptions and capabilities [Ren+18]

It was shown by Inoue et al. that presenting the same motion pattern with a real or virtual robot resulted in similar impressions by the participants [Ino+05]. On the contrary [Kam+11] found that a real robot was rated with "[..] higher scores for Utility, Possibility of communication, and Objective hardness and lower ones for Controllability as compared to a VR robot.” [Kam+11] Not only trained roboticists but also naive user’s interaction experience with robots profit from MR techniques. Dragone, Holz, and O’Hare proposed “the notion of a mixed reality agent, i.e. an agent consisting of a physical robotic body and a virtual avatar displayed upon it.” [DHO06] Such combinations allow for easy customization of the virtual component that matches the user’s expectations. In an industrial scenario AR can be used to show virtual robots in real environments or enrich the interaction with a real robot with visualizations like coordinate systems. A user survey showed that almost all people see an improvement of robotic training with AR and most of the participants see improvements for day-to-day work [BK06].

4.5.2 *Augmenting the User’s Perspective*

To explore this approach we integrated the *HoloLens* with our robot Floka humanoid (Floka). The data is provided over the ROS middleware. Figure 4.4 shows the augmented FoV of the interacting partner. The *Unity 3D* [Unity] game engine is used for the implementation of the visualization on the MR device. Communication between the MR device and the ROS topics available on the robot is realized using the *Message Queuing Telemetry Transport* [MQTT] stack.

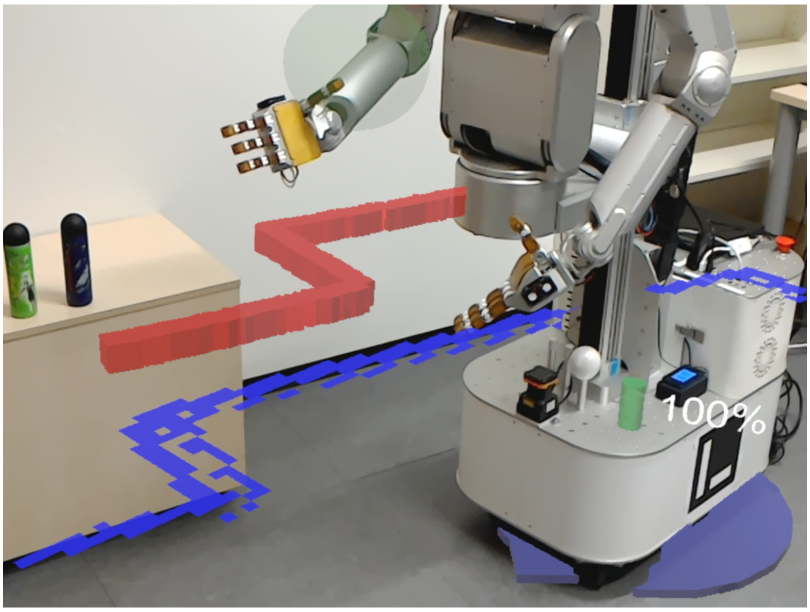


Figure 4.4: The sensory data visualized in the HoloLens: Map (blue), laser scans (red), the robot’s pose(purple) and battery status (battery symbol with white text). Additionally, the space where the robot can sense and receive objects for handover is highlighted with a green sphere above the hand [Mey+18].

As both systems have their own coordinate systems, we need to create a transformation between both entities. The room-scale tracking of the *HoloLens* is used to generate a position of the glasses and thus the user. The robot uses its navigation stack to localize itself in the room. A marker attached to Floka that is also integrated in its coordinate system is used to create an extrinsic mapping between both systems. This way the robot's information updates can be mapped into the FoV of the user and vice versa.

An occupancy grid map with the robot's position is shown as a grounding about the robot's knowledge about its surroundings. The plans on where to go for informing the user about the robot's next moves. In a handover situation this information might be useful to let the user know that he or she is approached and in which way. A virtual three-dimensional (3D) model of the robot can show the user the final position of the robot after the approach. A colored volume (see Fig. 4.4), can be used to show the predicted object transfer point (OTP) making it easier for the human to adopt. Additional, helpful visualizations could be imagined: the reachable workspace of the robot and planned arm trajectories could be visualized. The other way around the robot can use information from the glasses by means of accurate position and orientation. For debugging the system other data like the current forces and torques measured by the robot can be shown directly at the wrists. Tracking and prediction results could be added to the scene.

4.5.3 *Simulating Parts of the Robot*

Limitations of hardware could be overcome by simulating parts of the robot. Such limitations include hard movement limits of joints, speeds, and soft limits that are applied due to safety concerns. Combining a real robot with a virtual version allows to take the advantages of both worlds. Showing a different arm trajectory in VR but having the hand already in position to receive the object from the human. Such an approach could completely decouple the visual, nonverbal cues from the changes in the real world.

As stated in Section 3.3.2, gaze improves predictability of robots and their actions. In contrary to the arms, which are also perceived tactile and manipulate a real object, the head and eyes of the robot are primarily used to transfer information to the interaction partner. Thus, this visual cue can be used to create and evaluate gaze behavior in a highly controlled way. Where a real robot always produces motor noise and other artifacts a virtual pendant allows testing future functionalities.

An evaluation study of robot head designs for smart environments has been done with mockups in static pictures to study the impact of the appearance [BE17]. Such research would also greatly benefit from

MR where the different heads are shown on the real robot in 3D with movements. Fig. 4.5 shows the real robot torso and arms with two different virtual heads as perceived by the user through the *HoloLens*.

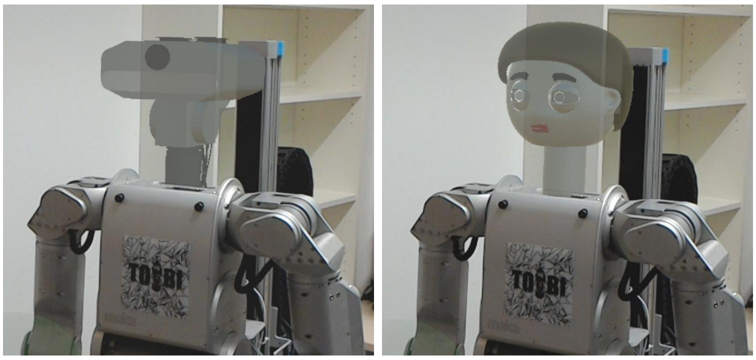


Figure 4.5: A mixed reality view from the HoloLens with two different heads on the Meka M1. The right picture shows a simulation of the newly developed Floka Head [Mey+18].

4.5.4 *Simulating the whole Robot*

By simulating the whole robot, in e.g. a CAVE Automatic Virtual Environment (CAVE), one gets rid of all stated hardware limitations. Fig. 4.6 shows the robot receiving a box from Patrick, in an L-shaped 3D stereo back-projection environment. Another advantage of such an environment is the integrated OptiTrack Prime 13W system, which is able to track the user's skeleton as well as rigid-bodies. Thus, here we can either copy movements recorded on a real robot for one-to-one reproduction or render any imaginable movement with the robot and present it to the interaction partner while always having the real position of objects and persons.

Figure 4.6: A virtual Meka M1, displayed in a CAVE, is receiving an object from Patrick. For the picture stereo vision was disabled and the perspective was corrected for the camera and not the interacting user [Mey+18].

Performing tasks with physical objects in a virtual environment is still an open task. Handing the object between a human and completely virtual robot might interfere with the immersion. However, it is possible to measure the reaction time and evaluate whether the users understand the robots intentions. In addition, the immersion can be improved with haptic gloves or a controller that has a virtual object attached. Showing a fully virtual robot in a head-mounted display (HMD) would allow to completely cover the view of the real world and thus separate the visual perception from the haptics.

4.5.5 *Mixed Reality Summary*

Though VR offers a lot of possibilities in regard to HRI, limitations like degree of realism or level of detail apply. Especially handover adds the requirement that a physical object is transferred. Here the haptics are still hard to simulate for immersive experiments and research. Strategies developed in this work could be deployed on robots that lack nonverbal communication (NVC) or human-like appearance by adding the needed parts virtually. On the other hand, by using a simulation in a completely virtual environment like in a CAVE, one could evaluate handover strategies in a completely safe and controllable environment while still utilizing the software stack of the real robot. This might especially of interest in the case of research of cultural influences on handover. Still, the problem remains that in the end there needs to be a real object transferred for a realistic handover which yet can not be done in a pure virtual setting. MR might pose a solution to the problems of pure virtual environments. Coming at the trade off, of hardware, like glasses that stand between a natural interaction of a human and robot. Limiting the possibilities of the robot in regard to face or gaze detection and might add a layer of unnaturalness that reduces focus on the actual task. The FoV is also quite limited and thus also negatively impacting the interaction.

Such tools can already help to either train the users or improve the user experience in some aspects. Additionally, they can be used as a debugging and development utility in the context of human-robot handover. While I target a system that incorporates nonverbal cues to overcome barriers between robots and inexperienced users, the presented VR and MR techniques still need development to fulfill the requirements.

5 HUMAN-ROBOT HANDOVER EXPERIMENT

In this chapter I present the design and evaluation of my human-robot handover experiment. It aims to validate and extend the aforementioned concepts in natural HRI. I added a distractor task that aims to prevent synthetic behavior of the participants. With this study I address hypothesis (H) 2: Second Arm Helps to Synchronize and H 3: Experience Changes Interaction. To test for H 2, I designed two different gestures for the arm, not directly involved in the handover, which I evaluate with a questionnaire and timing analysis.

From fairs and participation in *RoboCup@home* [ZW06; ZW07] competitions I gained the impression that there might be a significant difference in interaction from users that have no or only little experience with robots compared to experienced interactants [Wac+12; Zie+13; Zie+14; Zie+15; MKW16; Wac+17; WLM18]. Previous work discussed in Section 3.4 often solely evaluate interactions of participants that already have experience with robots or instruct them to test for a distinct behavior. Thus, one main goal of this experiment is to assess the behavior of users with different levels of experience (H 3: Experience Changes Interaction). Therefor the participants are grouped by their self-assesed level of experience and compared regarding timing and behavior.

The experiment also targets H 1: Handover Has a Distinct Pattern by putting the previously proposed structure to a real-world test. Parts of this chapter are based on “Hand in Hand with Robots: Differences Between Experienced and Naive Users in Human-Robot Handover Scenarios” [MBW17].

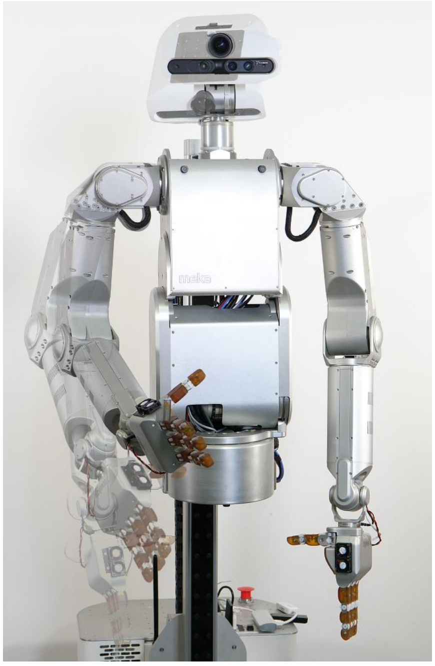
5.1 EXPERIMENT PROCEDURE AND DESIGN

The goal of this user study is to record the interactions with the Meka M1 as natural as possible. I decided to go without a tracking system that depends on markers or sensors attached to the participants to prevent interference with the interactants behavior. I post-annotated the user movements using automatic skeleton extraction from video recordings. In order to inhibit the emergence of artifacts by participants concentrating on the handover itself, a distractor task was placed. The participants were instructed to help the robot to learn to recognize a set of objects new to it. I chose the Meka M1 humanoid robot to study the interaction with the human. Its human-like torso with two arms allowed designing gesture-like motions that are equal to human ones and thus should be easily recognizable by an interactant. To solely analyze the effects of NVC, the robot did neither speak nor react to speech input during the study.

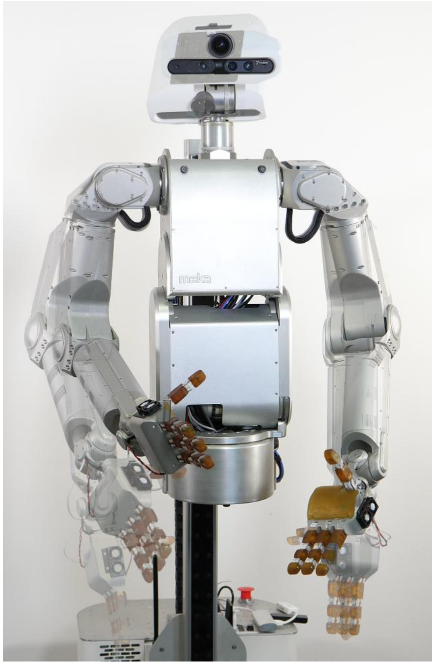
5.1.1 *Robot Behavior*

The behavior of the Meka M1 was programmed for handover type (HType) 3: I:H, G:R and HType 4: I:R, G:R, thus having the initiative in a giving and receiving handover. With these HTypes the robot did not require to recognize an intent of the interactant. It was proactively executing a reaching motion and waits for the human to react in both the receiving and giving case. The experiment always starts with the robot as a receiver and the human as a giver. For the robot, handover phase (HPhase) 0: Acquire is hence executed by a preceding handover. As human and robot start in a vis-à-vis configuration, no behavior for the HPhase 1: Approach was required. In the HPhase 2: Reach the EEF was always moved to the same fixed object transfer point (OTP_{fixed}). Reaching to such an OTP makes it easier to have a Full Joint Trajectory that follows the basic properties of a gesture-like handover motion to fulfill SR 2(b): Gesture Motion. This was achieved with prerecording a joint trajectory which can be seen in Fig. 5.1(a). For contact detection in the HPhase 3: Transfer I used an *ATI F/T Sensor: Mini40* [ATI:Mini40] in the robot's wrist to measure forces applied to its EEF, similar to related work discussed in Section 3.4.7. An experimentally determined threshold triggers opening and closing of the robot's hand to establish a power grasp (see Section 4.1). Thus, I can validate a first implementation of SR 4(e): Contact Detection. The robot is only able to detect contact after the arm trajectory finished as the motion itself applies larger force on the sensor than the interaction with a human. The whole implementation follows the paradigm of SR 1(a): Human-Like Pattern while being slowed down and having sharper isolated phases (see SR 1(b): Pattern Scalability). The behavior in the HPhase 4: Retreat was implemented, similar to the reaching, with a prerecorded joint trajectory.

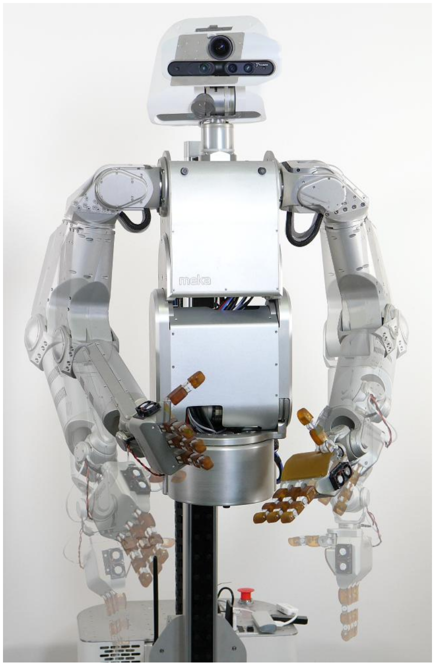
Due to safety reasons, the experimenter stayed next to the external camera with a wireless emergency stop. This e-stop device was also programmed to start each run of the experiment with an additional button on the e-stop remote control (see Section 4.1 on page 53). The experimenter started the next run when the participant had the object ready and was close enough to the robot to start the handover. Figure 5.1(d) shows the movements when the Meka M1 is learning an object as the mentioned distractor task. It moves the object in-front of its head, opens the hand except thumb and index finger, and nods to signal successful learning.



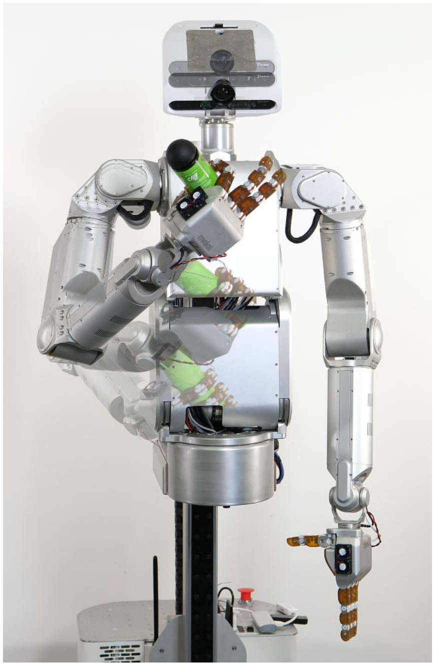
(a) No gesture ($C_{control}$)



(b) Low gesture (C_{low})



(c) High gesture (C_{high})



(d) Object learning distractor task

Figure 5.1: A collection of blended pictures of the Meka M1 during the interaction. Each figure is rendered out of three frames during the movement. Viewpoint is similar to the participant's. [MBW17]

5.1.2 *Gestures with the Second Arm*

A handover incorporates a lot of communication to synchronize between the interaction partners (see Section 3.3). Therefore, I designed an experiment to test in which way gestures with the second arm of the robot help to indicate the state of the robot (H 2). As of now robots are either not able to move and react with the speed and acceleration of humans or safety concerns lead to a limitation of those parameters. Hence, humans can not easily apply the same patterns and expectations they have from human-human handovers to the human-robot-case. The additional gestures with the second arm do not interfere with the reaching motion itself and hence can be easily added to existing handover systems.

I designed two different gestures for the left arm, which was not directly involved in the handover, for signaling the state of the robot. Figures 5.1(b) and 5.1(c) visualize the trajectory of the two different gestures. The first one (C_{low}) was moving the hand in a presenting manner below the object to signalize readiness. Figure 5.1(b) shows that this gesture made only use of small movements to be less intrusive. The second gesture (C_{high}) depicted in Fig. 5.1(c) started with a protecting movement of the object with the goal to signalize that the robot is not yet ready to hand the object. The trajectory ended as well in a presenting gesture but in a more obvious fashion. Both gestures where synchronized with the handover trajectory. As control condition, ($C_{control}$) the Meka M1 did not move the left arm at all during the handover. The arm was kept in a neutral posture as can be seen in Fig. 5.1(a). Each participant was assigned randomly to one of the three conditions, in which the gesture was activated for odd-numbered runs. Resulting in a control group that never saw a gesture, a second group that saw the C_{low} gesture in each second run and a third group that had the C_{high} condition activated in each second give and take. This allows analysis of between as well as within subject differences. It also prevents fatigue effects for the users seeing the gesture repeatedly. The interaction consisted of nine gives and receives each.

5.1.3 *Basic Gaze Strategy*

As discussed in Section 3.3.2, incorporating gaze in HRI improves the interaction and overall rating of the robot. Hence, I implemented a simple turn-taking gaze scheme for the Meka M1. The implementation of SR 2(c): Gaze for Predictability shows the advantage of a gaze pattern in interaction. In the HPhase 1: Approach the robot had the head facing straight ahead. The robot gazed at its EEF while executing the arm trajectory the participant (HPhase 2: Reach). After that, it gazed straight as an approximation of a face-gaze towards the participant when it was ready to hand/receive the object (HPhase 3: Transfer). These head movements are the same for all the interaction runs.

5.1.4 *Setup and Environment*

This experiment was setup in a room called *gym* in the CSRA. Figure 5.2 shows a schematic of the setup from top. The setup as seen from the view of the external camera is shown in Fig. 5.3. This camera is placed on a table on the other side of the room to have a complete view on the interaction. The Meka M1 is positioned such that the participant can freely chose a position in front of the robot. Three differently colored objects (Fig. 5.4) are placed on a small table near the interaction area.

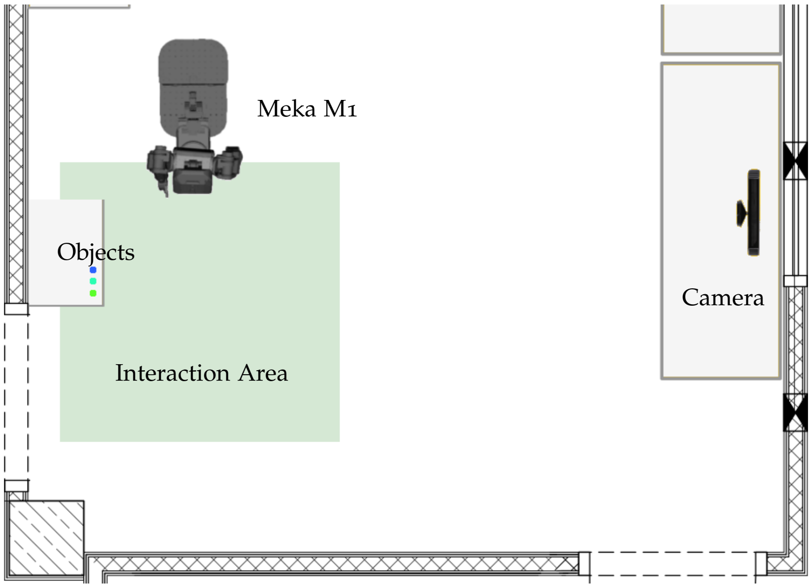


Figure 5.2: The setup of the experiment as a schematic top view. On the left is a small table with three objects. The participants interact with the Meka M1 in the interaction area (light green). An external camera is placed on the right [MBW17].

In total N=40 participants took part in the experiment. They were recruited by postings in social media groups and by distribution of flyers around the campus of Bielefeld University and Bielefeld University of Applied Sciences. Participants received 3€ as compensation for taking part in the study. Eight runs were not used in the following evaluation because of technical dropouts during the experiment, like the E-Stop losing connection, the robot turning off, or the recording being incomplete. Another three runs were not completely evaluable due to a timeout resetting the robot's arm after 20 s. Figure 5.5 visualizes the validity of runs. The remaining 29 participants (15 male and

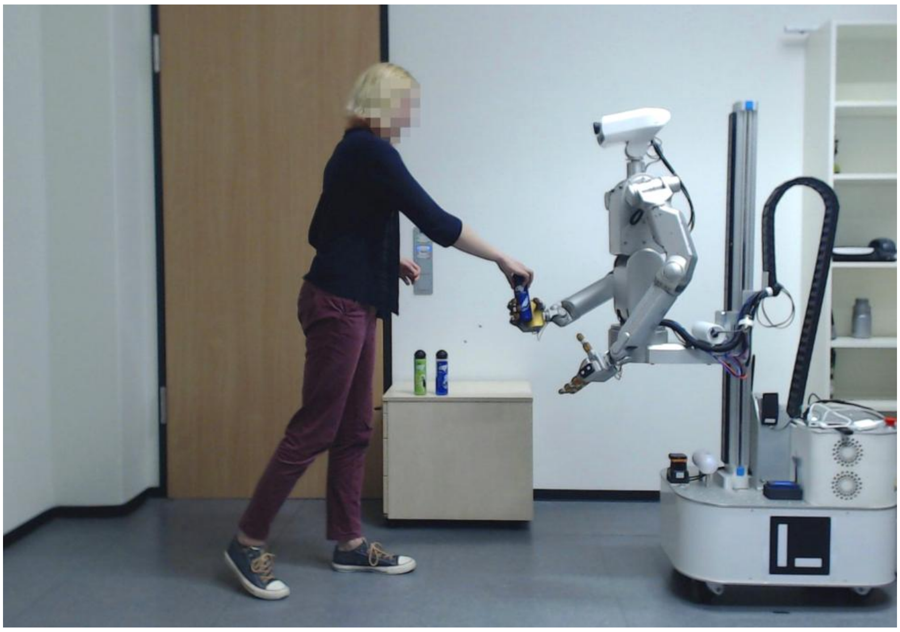


Figure 5.3: The experiment setup form the external camera perspective. The Meka M1 is receiving an object from one of the participants. The other two objects are still placed on a small table next to the interaction [MBW17].

(a) (b) (c)

Figure 5.4: The three different Objects as used in the experiments. All three objects have the same shape and weight. They are shampoo bottles with a baton like shape being ≈18 cm long and having a diameter of ≈5 cm with a weight of ≈200 g. They have different colors/textures to fulfill the requirements for the distractor task with 5.4(b) being mostly dark blue, 5.4(a) being mostly light blue/transparent and 5.4(c) being green.

14 female aged between 18-53 years) were distributed across condition as follows: 10 $C_{control}$, 10 C_{low}, and 9 C_{high}.

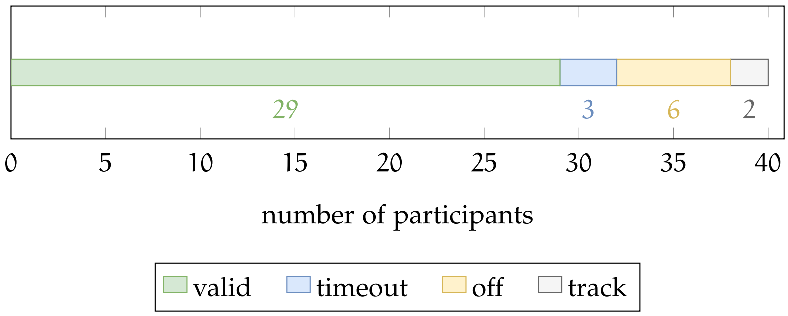


Figure 5.5: The outcome of the 40 participants with eight removed evaluations due to technical problems and three due to a too early timeout of 20 s.

For each participant the procedure was as followed:

- Enter the room
- Read and sign a consent-form on a designated table
- Walk to the *Interaction Area* (see Fig. 5.2)
- Receive an introduction to the Meka M1
- Get instructions on the experiment (see Appendix A.3.2)
- Interact with the robot for nine runs (each object (see Fig. 5.4) three times give and receive)
- Answer a survey (see Appendix A.2)

The participants were instructed to give and receive the objects each three times so that the robot is able to learn them. This way the participants gave and received the object nine times. See Appendix A.3.2: Examiner Instructions [translated] for a translation of the full instructions given. After the interaction, the participants had to answer a questionnaire (see Appendix A.2). Besides age and gender, they gave a self-assessment on experience on a Likert-scale of 1 (no experience) to 7 (a lot of experience) with technology like computers, robots, and the robot used in this interaction.

I divided the participants into three groups based on experience with robots, as stated in their self-assessment in the questionnaire, to test for H 3: Experience Changes Interaction. The group of naive users only contains participants that stated they have no experience on interaction with robots. This is the first group with 12 persons. The participants that answered to this question with 4-7 form the group of experts, containing eight persons. The remaining nine participants form the group of semi-experienced. I collected information on how the robot was perceived during the handovers with the "Godspeed-Questionnaires" [Bar+09] ($\alpha = .90$) to evaluate it in context of H 2: Second Arm Helps to Synchronize. In the conclusion of the survey the participants were asked whether they noticed different behavior patterns during the interaction. This data was collected by means of a free-text field. The attitude towards robots was investigated using the Negative Attitude toward Robots Scale (NARS) ($\alpha = .64$) question-set [Nom+06]. This information can be used to detect interactants that have a negative bias towards robots in general. Figure 5.6 gives an overview of the composition of participants in the experiment. Although the Meka M1 is bidextrous and my software takes care of using both hands/arms by mirroring joint values to both arms, I chose to run all experiments right-handed as this design decision takes care of having the same behavior and annotation technique in all runs (see Section 3.4.8: Laterality and Handedness).

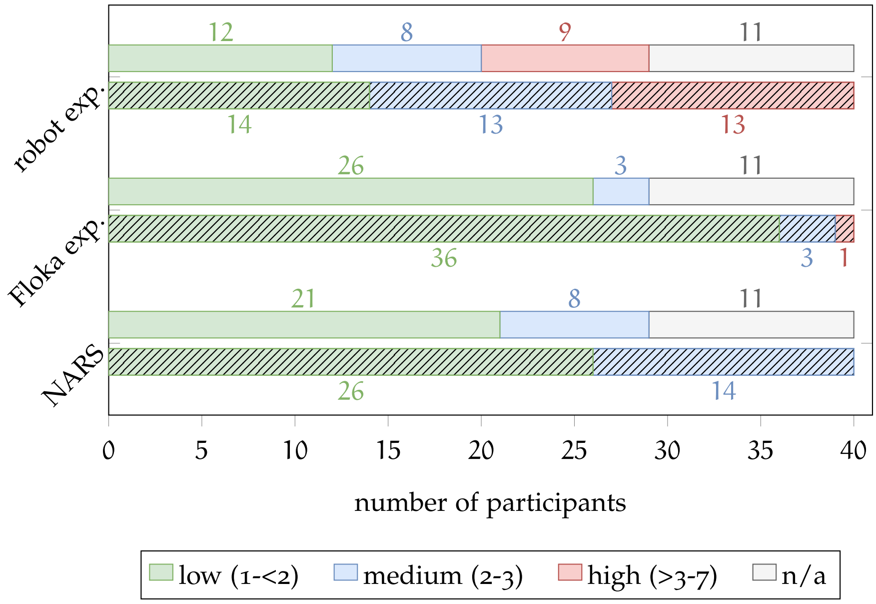


Figure 5.6: The structure of the participants taking part in the experiment. On top is the final group, after removal of runs with technical problems. Below is the distribution of all 40 participants.

5.2 EXPERIMENT ANNOTATION

I recorded 725 handovers during this experiment with the following data: the timing and state of the handover behavior control system, the control state of the Meka M1's joint_trajectory, calibration data of active cameras, the whole tf-tree, image-streams from and internal as well as an external camera, and the forces measured by the FTSs installed in the robot's wrists (see Section 4.1). A detailed list of the recorded topics during the experiments can be found in Appendix A.1.

A marker on the robot's base helped to exactly determine the position of the external camera in relation to the robot and thus add it to the tf. I used the *Bielefeld Augmented Reality Tracker* [BART] to detect and track the marker in the camera image. This marker position allows mapping internal robot data like forces, torques and positions into the video as depicted in Fig. 5.7 as well as transforming data in to the same coordinate system for an easier comparison.

I implemented a software-pipeline that loads and automatically annotates the recordings in order to extract the position of the human with *OpenPose* [OpenPose]. To precisely annotate the hands, I enhanced the pose detection with hand keypoint detection [Sim+17]. The resulting annotation can be seen in Fig. 5.7. I compared the extracted positions and velocities of the human hand with the recorded data of the robot. The processing-pipeline generated a log of the positions extracted by OpenPose and hand keypoint detection as well as a video with all data visualized for each recording. In addition, timestamps for the beginning of the following HPhases were logged: reaching (HPhase 2), transfer (HPhase 3), and retreat (HPhase 4).

Figure 5.8 shows the trajectories of the robot and participants during handover as extracted by the described pipeline. This novel annotation system which does not interfere with the participants was created to evaluate HRI by making use of state-of-the-art deep-learning techniques. This low-cost and easy deployable system allows fully automatic processing of human motion without time-consuming manual annotation of video data. Furthermore, it can replace other intrusive marker-based tracking-solutions.

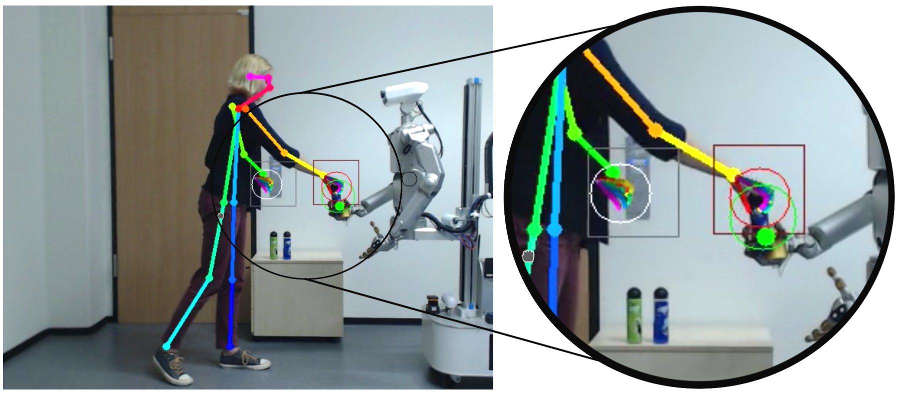


Figure 5.7: A visualization of the results from post-processing with help of *OpenPose* [OpenPose]. Each joint is visualized in a different color. The position of Meka M1's hand is marked with a green dot. Bounding boxes of possible participant hand positions are red for the right and gray for the left hand. Accordingly, the center of hand joints is surrounded by light red and light gray, respectively. When contact between human and robot is detected a green circle is drawn around the hands in contact. [MBW17]

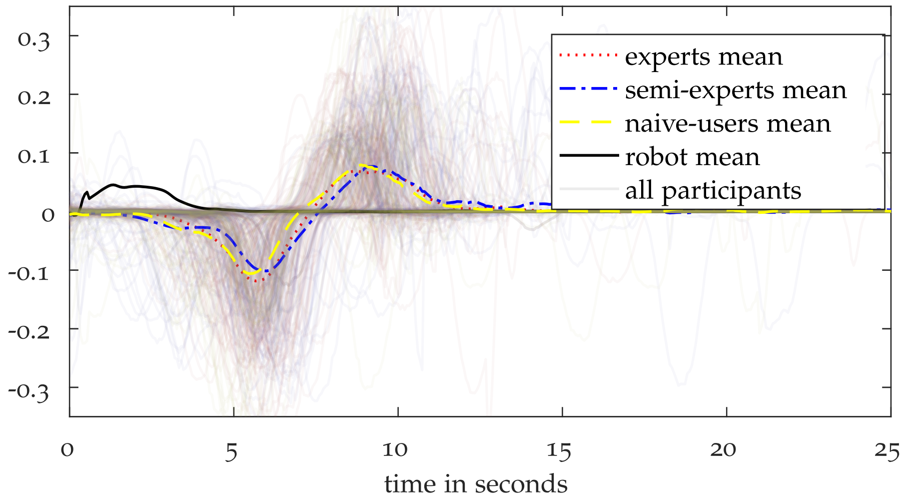


Figure 5.8: The velocity profile of the handover runs for right-handed interactions. Meka M1 moves its hand towards the participant. Some participants start to move right after the robot started. Most of them wait until it finished, then hand the object and move back. The colored lines represent the average for each of the three groups. [MBW17]

5.3 FINDINGS AND RESULTS

After the recordings were annotated with the before-mentioned processing pipeline, I analyzed the quantitative data regarding differences for experience as well as influence of gestures with the second arm. With the extracted timestamps I calculated a reaction and a transfer time for each handover. These were analyzed with an ANalysis Of VAriance (ANOVA) to test for significant deviations between the experience and condition based groups. I also compared the Godspeed results for the rating of different conditions. The object was dropped in a single run and timed out in three runs after the robot was waiting 20 seconds for the person to pull the object strong enough to trigger the force threshold for releasing the object.

5.3.1 *Timing Analysis*

One of the analyzed aspects was the reaction time to see how well the movements of the robot and participants aligned. The alignment was calculated as the difference between the time the robot was ready and the time the person's hand getting close to Meka M1's hand. A perfect alignment would be a 0.0 s result. At an overall average the participants took 0.29 s which means, that they gave a little more time for the robot to finish moving. Negative differences are cases in which the participants tried to hand the object while the Meka M1 was still executing the trajectory. Going on, I calculated the time needed to transfer the object by measuring the duration from the first contact until the force threshold is reached. This mainly tests, how well the force-based approach succeeds in detecting a stable handover.

5.3.2 *Influence of Gestures*

At first, I grouped the timing data by the condition group and calculated mean and standard deviation for each group. The results for the reaction time can be seen in Table 5.1a. Respectively, Table 5.1b shows the results for the transfer duration. Additional boxplots of the data can be seen in Appendix B.2. To compare the three mean values of the groups I calculated an ANOVA. Results did not show a statistically significant effect ($p > .05$) on both duration between any group.

	high	*low*	*control*
$\overline{x}_r$	0.00	0.44	0.33
σ_r	1.32	1.38	1.60

(a) Reaction times: the time between the robot was ready to receive or hand an object and the participant reaching for it.

	high	*low*	*control*
$\overline{x}_t$	2.44	1.67	1.70
σ_t	4.45	2.39	2.41

(b) Transfer times: the time between first contact between robot and human and contact detected by the robot.

Table 5.1: The mean and standard deviation for measured duration grouped by the gesture condition in seconds.

Further analysis of the ratings, from the Godspeed survey, with an ANOVA did not show a statistically significant effect ($p > .05$) either. Hence, H 2: Second Arm Helps to Synchronize could not be supported based on the data. The analysis of the free-response of the survey showed that in total 17 of the participants stated that they experienced differences in the behavior of the robot in between the runs. This includes answers like: "The robot nodded in the second run.", "One time the robot closed its hand faster". Only seven participants were able to describe the differences correctly in the way that the second arm did support the handover with a gesture. Here, correct answers included: "Gestures with the free hand as a sign to take the object; sometimes clear; sometimes inconclusive", "The left arm was moved to different postures during the handover.", or "The robot performed different gestures during handover (two arms vs. one arm).". Some participants stated in the free-response that movements with the gesture looked more natural.

For some participants the high gesture even looked like the Meka M1 was offering the left hand for receiving an object. This might have led to confusion and created delays until they continued to give the object into the right hand, as only the robot's right hand was able to detect and grasp objects.

5.3.3 *Influence of Experience*

To evaluate the influence of the interactant's experience on the timing in this handover experiment, I repeated the analysis based on the experience groups. Tables 5.2a and 5.2b show the resulting means and standard deviations. Here, bigger differences between the groups can be observed. These become more obvious in Fig. 5.9 which shows a boxplot of the reaction times grouped by experience. The ANOVA resulted in $F(2, 519) = 20.97, \eta^2 = 0.075, p < .001$ for the reaction time and $F(2, 519) = 11.10, \eta^2 = 0.041, p < .001$ for the transfer time.

	naive	*semi*	*expert*
$\bar{x}_r$	0.36	-0.30	0.76
σ_r	1.09	1.38	1.86

(a) Reaction times: the time between the robot was ready to receive or hand an object and the participant reaching for it.

	naive	*semi*	*expert*
$\bar{x}_t$	1.60	2.68	1.27
σ_t	2.45	3.89	1.62

(b) Transfer times: the time between first contact between robot and human and contact detected by the robot.

Table 5.2: The mean and standard deviation for measured duration grouped by the experience with robots given in seconds.

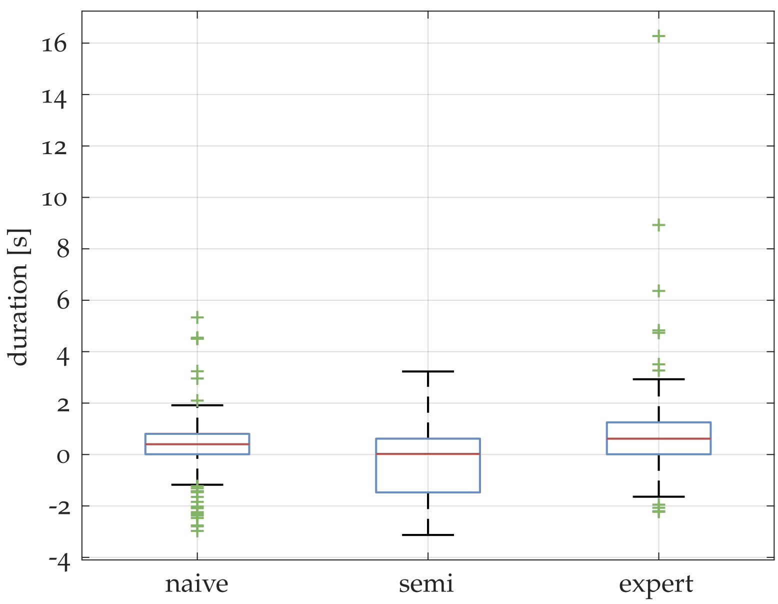


Figure 5.9: The reaction times of my handover study grouped by the three experience groups. The extracted times are visualized as a boxplot. The central red line shows the median. Top and bottom edges of the blue box mark the 25th and 75th percentiles. The whiskers reach to minimum and maximum, not considering outliers, which are plotted individually using a green + symbol.

Table 5.2b shows that experts have lower mean and standard deviation of time when exchanging the object with the Meka M1. However, naive users show the least standard deviation. In the analysis of recordings I could observe the experts and also some semi-experienced participants actively challenging the robot, which caused most of the outliers. They purposely introduced delays to see how the robot would react to them. The testing went as far as giving the object in the hand of the robot and pulling it away as the robot closes its hand. Based on these results I can confirm H 3: Experience Changes Interaction.

5.3.4 *Discussion*

I conducted a study on natural human-robot handover with the Meka M1. Therefor, I used an implementation of wrist-force based handover detection in the HPhase 3: Transfer. There was no artificial tracking system and only minimal instructions for the participants to observe the interaction without interference.

A gesture with the second arm did not show a statistically significant effect on the rating of the robot (H 2: Second Arm Helps to Synchronize). This might be explained with the fact that the participants which experienced the gesture only saw it in each second run but rated the interaction as a whole in the questionnaire. While this design was chosen to reduce fatigue effects by not always seeing the same behavior, the mixed design aggravated the individual analysis. Nevertheless, participants that consciously perceived the gesture, stated in the survey that they experienced the robot more human-like when the gesture was part of the interaction. Further analysis of the gesture condition based on timing data also did not show statistically significant results. As the gestures with the second arm were handcrafted, it might be that they just did not meet the expectations of the interactants. Also, the interaction with the robot itself might have diverted the attention of users to other aspects of the interaction. While it might be interesting to address the topic again in the future, I progress with a focus on the primary hand during handover.

Effects of other phenomena appear stronger in the data and addressing them appears promising for smooth handover, e.g. the delay of the interaction because the arm trajectory needs to finish before the robot is able to sense the interaction, as well as the limitation to depend on force. The former observation in form of participants reaching for the object or EEF of the robot before it finishes the trajectory motivates SR 1(c): Reactive Pattern and SR 3(b): Shortcuts for Experts along with revealing the desire to allow in motion handover. The pure force-based detection approach revealed to be problematic as some participants did not apply force at all on the first tries and expected the robot to see that they hand the object. This was especially happening for handing an object to the robot compared to receiving it. It even led to timeouts and thus to incomplete handover interaction. This shows the necessity to not only rely on force measurements to prevent a social gap between users of service robots but also shows the importance of SR 4(f): Visual Transfer. Nevertheless, when giving an object to the robot, applying pressure seems to be less intuitive than pulling on the object when taking it from the robot.

The influence of different user behaviors and their needs became even more apparent after dividing the interactants into groups based on their stated experience with robots. Therefor, I proposed H 3: Experience Changes Interaction and evaluated the timing data accordingly. I could show statistical significant differences for both, reaction and transfer, duration with varying levels of prior knowledge in regard to the H 3) with robots. I chose to split the users into three groups as it showed to give the best explanation for participant in the resulting groups. Only two groups or a direct correlation of experience and duration would not haven been able to explain spikes for experts. Here a descriptive analysis has the best effect to address the needs of all the groups (SR 3(a): Understandable by Everyone). While the results are limited to Western, Educated, Industrialized, Rich, and Democratic (WEIRD) [HHN10] participants, they still show that robots are required to adapt to the experience level. Experts for example actively challenged the robot by removing the object after it tried to grasp it. They also seem to be already used to trigger force thresholds to make a robot react. In this experiment naive users tried to align well with the system and seemed to help it to fulfill the task of learning objects most efficiently. However, while naive users expect the robot to visually perceive the environment and react accordingly, experienced users know that they need to pull and push objects to let the robot perceive their intention. Especially with the elderly and disabled people in mind, handover of robots needs to be more adaptive to cope with the observed variance of handovers and to adapt better to the human expectations.

Summing up, this experiment revealed differences of human handover behavior based on the experience with robots. It also hinted the validity of H 1: Handover Has a Distinct Pattern as most participants were able to accomplish the stated task without further explanation. Nevertheless, the experiment showed the importance of a more reactive pattern to better fulfill everyone's needs, which I present in the following chapters.

6 USING ROBOT GAZE FOR PREDICTABILITY

Based on the related work on robot gaze (see Section 3.3.2) and the experience made in the previous study, I aimed of improving the gaze capabilities of the Meka M1 (SR 2(c): Gaze for Predictability). As the current sensor-focused head of the Meka M1 moves rather slow and its appearance might not even create the impression of looking at something for human interactants, it is possible that the interactant is oblivious of the robot's gaze behavior. Thus, I first discuss the physiological improvements I introduced to the platform and then discuss the behavioral additions for the enhanced gaze behavior, addressing the integration of both hard- and software to fulfill SR 5(a): Robot Integration. Attending to SR 1(a): Human-Like Pattern the gaze needs to be aligned and scaled (SR 1(b)) to the basic pattern of a handover. For the SR 5(c): Modularization it needs to be taken into account that the created module and addition do not interfere negatively with the existing system.

6.1 FLOKA – THE INTERACTIVE HEAD

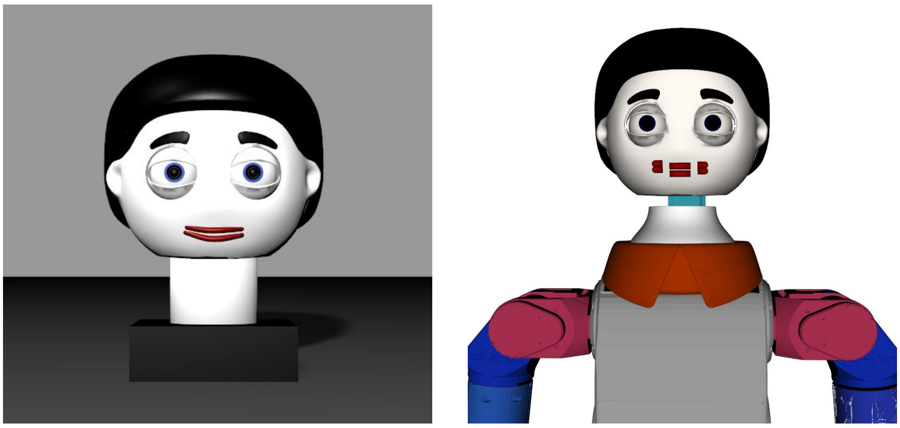
The base for employing gaze as means of NVC for communicative purposes is an embodied agent with human-like properties in terms of appearance and performance. The most important part take the eyes as they are the body part known by humans to visually perceive the environment. With "See and Be Seen – Rapid and Likeable High-Definition Camera-Eye for Anthropomorphic Robots" [SMW19] we presented a robotic eye that reaches human-like speed and acceleration. While their appearance is inspired by a comic-like character, they are very expressive and thus easy to read by interactants. A high resolution camera inside they eye provides a view from the robot's perspective. These eyes are integrated in the *Floka Head*, which is a subsequent development, based on the humanoid head *Flobi* [Lue+10]. In combination with the Meka M1 we named it *Floka humanoid (Floka)*. Figure 6.1 shows the Floka Head mounted to the Meka M1 making it a mshs-robot. It was also evaluated to be a good fit to the whole robot [BE17]. For testing and development a virtual version was created based on the concepts of Lier, Schulz, and Wachsmuth [LSW14] (see Fig. 6.2).

Floka Head

Floka humanoid (Floka)

The low-level gaze-control employed as base of this work is build on the concepts of Schulz et al. and implemented in the *[hu]man [motion] low-level robot control library [Sch+16]* [humotion]. This control framework allows human-like control of robotic heads. For the Floka

Figure 6.1: The Floka Head for enhanced gaze behaviors features two eyeballs with human-like capabilities in terms of acceleration, velocity, range of motion, and perception [SMW19]. In addition to the eyes, the brows, the lids, and the mouth are movable. The comic-character-like look increases expressiveness and allows humans to easily read emotions and gaze directions. It can be mounted on the Meka M1 to create Floka.



(a) Floka Head simulation in Blender (b) Floka Head visualization in ROS-rviz

Figure 6.2: The virtual Floka Head for interaction and internal representation in the ROS environment. These can be used for development purpose or a replacement of the physical robot for direct interaction as a virtual agent on a computer screen or embedded mixed or pure virtual scenarios (see Section 4.5).

Head it coordinates head (neck) motions with the eyes'. The naturalness is further increased by actuating the eye-lids of the robot to follow the motions of the eye-ball. As the actuating principle resembles behaviors learned from humans, interactants can easily follow the generated gaze [Sch+16]. It also allows scaling of accelerations and velocities of the neck and eyes while maintaining human-like properties of the overall motions. Here, I focus on producing selected gaze-targets for humotion which then generate the motor commands.

6.2 ALIGNING HUMAN-LIKE GAZE WITH HANDOVER

Based on the literature we designed pattern that fit the model of handover. The goal is to communicate the internal state of the robot to the human. It has also to be kept in mind, that staring needs to be prevented as this might upset interactants or might be perceived as impolite. On the other hand the gazes needs to be long enough to be correctly perceived and do not let the robot appear nervous.

6.2.1 *Aligning to Handover Phases*

I present two types of gaze, namely *main*- and *sub*-gazes, which are combined in pattern. The main-gazes have the distinct task of communicating the state or intent of the robot and the sub-gazes give the possibility to interrupt the main-gaze to prevent staring or add smaller pieces of information to the interaction. Some approaches chose to model gaze of robot's implicitly by setting targets based on an own agent that decides where to look at based on the current sensing. This has the advantage of being applicable to different kinds of interactions, as well es being less tightly coupled to the overall system, in favor of a stronger modularization (see SR 5(c)). As discussed in SR 1(a), handover features a well researched pattern which allows to explicitly set targets that feature the best information gain for the interactant. Hence, the gaze should be modeled according to HPhase 1: Approach, HPhase 2: Reach, HPhase 3: Transfer, and HPhase 4: Retreat as listed in Table 6.1 based on the findings discussed in Section 3.3.2.

Handover phase	**Main gaze**	**Sub gaze**
Acquire	left; right	-
Approach	face $\rightarrow$ object	hand; human chest
Reach	face	object; predicted OTP
Reach done	face $\rightarrow$ object	hand
Transfer	object	hand; face
Retreat	EEF $\rightarrow$ straight	-

Table 6.1: The gaze targets in the handover phases.

PATTERN IN THE IDLE PHASE This pattern is active, when the robot does not sense anyone around him, aiming to signal looking for an interactant to appear lively and ready to exchange objects. It might happen that the interactant does not witness this behavior as it transitions to the next phase as soon as someone approaches the robot.

GAZING WHILE APPROACHING This has the target of creating a joint attention on the cooperative task of handover. It begins as soon as a human is detected, fixating the face and then the object aims to produce a common reference, while the occasional looks at the hand intend to inform about exchanging the object. It can also be used to check whether the possible handover carries something to estimate the intent based on a giving or receiving situation. Looking at the chest can be incorporated to signal interest in interaction and focus at the approaching person without staring into the face.

LOOKING WHILE REACHING The robot needs to inform the interactant about the state of reaching out. Another look into the face acknowledges the addressee of the handover. Additionally, a look at the final OTP should increase the predictability of the arm's motion. As here, a recorded joint-trajectory is used and in motion handover is not yet considered, another goal is to inform the user when the robot is ready to release or grasp the object. A gaze from face to object tells "give it" respectively "take it".

TRANSFERRING THE OBJECT This gaze pattern is likely to happen when joint attention is already established. Here it is important to signal that the robot is focusing on the object, to safely transfer it from hand to EEF or vice versa. On the other hand, looking occasionally at the face showed to increase the likability, thus it might be added for a better interaction.

6.2.2 *Timing and Duration*

There are a number of factors influencing the timing and duration of the selected gaze targets. For the timing, they should be triggered as soon as another phase becomes active. This allows to inform the interactant about the internal state of the robot immediately. While smaller deviations might not harm the interaction, bigger delays might negatively interfere with the alignment [Adm+14].

Deciding on a duration is more complex as not only the informative part of the gaze needs to be taken into account, but social norms as well. Especially the impression of staring needs to be prevented while looking long enough to be perceived by the interaction partner. The minimum and maximum duration for *main*- and *sub*-gazes is based on the findings of Mutlu et al. It was found that main-gazes mostly last 1.40 ± 1.30 s, while the shorter glances where found to last about half as long with 0.77 ± 0.58 s [Mut+09a]. As the transition between targets was not discussed, an extra of 0.5 s serve as a transition time between targets. This parameter is also robot dependent and needs to be decided according to SR 5(a): Robot Integration and SR 1(b): Pattern Scalability. The gaze cues need to be scaled according to the overall handover behavior the robot shows. Yet, the distinction of two gaze types automatically scales, if the interaction takes longer, by repeating suitable gazes. For a more natural behavior the duration is drawn with a Gaussian distribution with the given mean and standard deviation. An additional upper and lower boundary prevents outliers.

6.3 Gaze Integration with Floka

Addressing SR 5(a): Robot Integration, the developed strategies need to be synchronized to the existing handover behavior. By explicit modeling of pattern for the phases, the gazing is tighter coupled to the handover. The importance of precise alignment asked for embedding into a combined finite-state machine (FSM). We chose to transfer the existing behavior (cf. Section 5.1.1) to the *FlexBE - The flexible behavior engine* [FlexBE] software to allow for easier reuse of gaze pattern in the phases [Poh18]. It is build on the *executive smach* [SMACH] library, that helps to create robust robot behavior. Also, the visualization of current states makes it easier to test and extend the models. As it does not provide a functionality to run states for a minimum duration, which was required to prevent too frequent gaze shifts, we added blocking states that wait for the duration to expire before transitioning to a new gaze target [Poh18]. On the downside this also retards the rest of the handover behavior but was accepted in favor of a better gaze experience.

6.3.1 *Input Stimuli*

In contrast to the gaze strategy discussed in Section 5.1.3, in which all head motions where prerecorded, I integrated a module that provides such targets based on perception. With the previously used sensorhead, smaller motions are hard to recognize and a straight orientation might be perceived as a face gaze, with the precise and fast Floka Head a straight gaze does not look lively and does not create the impression of being looked into the face either. Thus, SR 4(d): Face Tracking needs to be addressed here. An OpenPose based software was used to satisfy this requirement. This module did not only determine the position of the interactant's face for mutual gaze, but also estimates the proxemics to transition from the idle state (referring

SR 4(c)). It also provides the interactant's hand (addressing SR 4(a)) and torso positions. The stimuli are provided with ≈18 Hz while the data of the Meka M1's EEF was given by the tf system at 50 Hz. As the gaze targets are set directly from the FlexBE states, the updates are forwarded to humotion at 10 Hz.

6.3.2 *Comparison with Non-Interactive Gaze*

For an assessment of the improvement by this strategy compared to the basic gaze strategy (cf. Section 5.1.3) we recreated the same setup as in the previous study [Poh18]. The original behavior was adopted for the Floka Head and FlexBE. Eight persons were asked to exchange the three objects two times with Floka. One time in the enhanced condition and once in the original, more simplistic, scenario. After each condition, the participants had to answer a questionnaire.

Analysis showed that half of the participants liked one condition better than the other and vice versa. Some people stated to feel more addressed by a straight gaze than by the incorporated perception-based face gaze. A later analysis revealed an offset in the face tracking component which might have caused the confusion. This shows the importance of precision when employing enhanced gaze strategies. Another stated downside of the enhanced strategy was that some participants reported switches of targets as too frequent. Some participants reported that Floka appeared less focused on the handover task. The glances at the torso were reported as misaligned face-gazes. Here, at least for some participants the eye and head target shifts should have been less frequent. Possibly, if the whole robot would move faster, the saccades would have been accepted better, which shows the complexity of SR 1(b): Pattern Scalability. Participants that moved their head to test whether the robot smoothly follows their motion reported the gaze to be jerky sometimes. Here, the stated 100 ms between target updates shows to not suffice for smooth transitions and needs to be addressed. On the other hand it was stated that especially the behavior of following the own movements created a feeling of being addressed. The overall behavior was also described as more human-like and less repetitive, which might get tiring over multiple interactions [Poh18].

Overall, even the baseline condition yielded good results as it was already in line with previous implementations found in related work as discussed in Section 3.3.2. Nevertheless, the enhancement showed the potential offered by a human-like head with fast and predictable eye movements. The number of information added to the gaze could be increased and the interaction can be more pleasant.

7 FUNCTIONAL GESTURE MOTIONS

In this chapter I present how to create predictable as well as adaptive reaching motions with a humanoid. As learned from the previous experiment, a robot should be able to adapt to the human (see Section 5.3.4). At least reaching for a single $\mathrm{OTP}_{\mathrm{fixed}}$ is not sufficient for smooth object exchanges with a naive person as it requires the human to do the adaption alone. Based on previous work (see Section 3.4.6) I found that ideally a robot should be able to adjust its movement to the humans continuously as there is no perfect prediction of a $\mathrm{OTP}_{\mathrm{static}}$. While there are multiple approaches to generate inverse kinematics (IK) for a robot to move its EEF to an updated target position, most lack the capabilities to retain human-like motions that are predictable and thus keep signaling properties.

In line with SR 5(c): Modularization I targeted to build a module for trajectory generation during the HPhase 2: Reach, aiming to have an adaptable OTP, predictable motions, collision prevention, and reduced effort for the human.

7.1 COMBINING RECORDED AND DYNAMIC MOTIONS

It was already observed by Kajikawa and Ishikawa, in one of the first handover studies [KI00], that the human reaching motion consists of two parts (see Section 3.4.6.1 on page 41). In this chapter I present the adoption of this concept for humanoids. Splitting the motion into two parts (HPhase 2(a): Base Reach, HPhase 2(b): Adapt) is not only in line with findings on human handover, it also has the advantage of being able to tackle each of them with the ideal solution. Nevertheless, one has to make sure that both phases blend smoothly into each-other. For the first, less precise, phase I propose a database of recorded motions and for the second, more refining, an inverse instantaneous kinematics (IIK) approach. This concept allows reaching functional to a moving target while maintaining the initial signaling effect of the motion.

7.1.1 *Handover Motion Database*

Using a database of motions is in line with previous work that emphasized the importance of a gesture-like handover Reach motion (see Section 3.4.6). Therefor, not only the EEF trajectory is taken into account, but the whole arm (see Section 3.4.6.2: Full Joint Trajectory, SR 2(b): Gesture Motion). A database allows taking full-control over the appearance and expressiveness of the arms motion. Another advantage is that trajectories can be selected from the database with nearly zero delay, allowing a fast reaction of the robot. I term such a first part of the motion read from a database a *base trajectory (BT)*. On the other hand it is easy to record a database by animating the robot in a real or virtual environment. Nevertheless, each BT requires some manual work and it needs adjustment for different hardware. Thus, I suggest keeping the number of recorded trajectories to a minimum by spreading their final position in the task space of the robot. As the adaption takes over before reaching the final goal, this approach is feasible to then cover the whole interaction space. Other options to fill the database include sampling the goal space offline and generating BTs based e.g. on the joint motion model (JMM) proposed by Rasch, Wachsmuth, and König [RWK18]. This would allow filling the database without manual work by an operator, while having the results without delaying the interaction due to online computation.

base trajectory (BT)

The database D ($D = \{\vec{t}_i, ..., \vec{t}_N\}$) consists of a set of trajectories $\vec{t}$ ($\vec{t} = (p_k, k \in 1..N_{\text{traj-points}})$), which consist of point tuples p ($p = \{\vec{q}, d\}$) with joint positions $\vec{q}$ ($\vec{q} = (q_j, j \in 1..N_{DOF})$) and duration d defining a trajectory segment. On reading the database a final position (p_{final}) for each motion is computed by forward kinematics (FK).

For a humanoid with two arms that are symmetric the trajectory in the database can be mirrored/inverted to allow ambidextrous handover. Thus, for the Meka M1 the training and recording needs to be done only for one arm and can be directly applied to both.

The best trajectory is selected based on the euclidean distance ϵ (Eq. (7.1)) to an offset goal ($p_{\text{goal}} + p_{\text{offset}}$). Offsetting the goal with p_{offset}, allows to have an approaching direction as well as taking EEF and object properties into account for the Transfer.

$$\epsilon = \|p_{\text{final}} - (p_{\text{goal}} + p_{\text{offset}})\| \tag{7.1}$$

The trajectory with the smallest ϵ is selected as base for the reaching motion. This approach already enhances the capabilities of the robot compared to the first implementation (cf. Section 5.1.1) by having multiple possible motions and goals.

7.1.2 *Adaptive Motions*

Now that the robot is able to start an initial motion of the arm, it needs to be enabled to update the motion during execution if, e.g. the goal changes or if the selected BT does not finish close enough to the goal position (SR 1(c): Reactive Pattern). I term such an adaption of the robot's motion *adaptive trajectory (AT)*. While calculating the FK is quite simple, calculating the IK can cause problems, like redundancy resolution of high DOF manipulators. One solution to the *inverse instantaneous kinematics (IIK)* approach, that deals with online calcula-

adaptive trajectory (AT)

inverse instantaneous kinematics (IIK) tion of IK, is the inversion of the manipulator Jacobian [SK08]. An approach tackling both the challenges of bringing the EEF closer to the target position and maintaining a natural posture (SR 2(b): Gesture Motion) is required. As the regular inverse of the Jacobian (J^{-1}) is not always computable, different approaches have been developed to achieve approximations. The Jacobian Transpose, the Pseudoinverse, and the damped least squares (DLS) methods offer solutions to the problem of non-inversible Jacobians, each with different advantages as well as issues [Bus09]. While the mathematical approaches have been established for a long time, I briefly discuss them in the context of robotics.

All the approaches aim to give a transformation X that states how to modify the joints $\vec{q}$ to come closer to the desired Cartesian position ($\vec{p}_{\text{goal}}$):

$$\Delta\vec{q} = X\vec{e}, \ \vec{e} = \vec{p}_{\text{goal}} - \vec{p}_{\text{current}} \tag{7.2}$$

JACOBIAN TRANSPOSE The Jacobian transpose method was first used by Wolovich and Elliott [WE84] and Balestrino, De Maria, and Sciavicco [BDS84] to solve the stated IIK problem. They replaced the inverse of J by the transpose of J and introduced an α used for damping, resulting in the following equation:

$$\Delta\vec{q} = \alpha J^{T}\vec{e} \tag{7.3}$$

This has the advantage of being fast to compute but the results may not be good and α needs to be set correctly.

PSEUDOINVERSE METHOD The Pseudoinverse method makes use of the Moore-Penrose inverse of J to solve the equation.

$$\Delta\vec{q} = \alpha J^{+}\vec{e}, \text{ if J full row rank: } J^{+} = J^{T}(JJ^{T})^{-1} \tag{7.4}$$

This approach is also fast to compute, but the Pseudoinverse Method might become unstable near singularities [Bus09], which is an effect to be avoided, especially during HRI.

DAMPED LEAST SQUARES METHOD The *damped least squares (DLS)* method, also known as the Levenberg–Marquardt Algorithm eliminates many of the problems with singularities of the Jacobian Transpose and Pseudoinverse Method. Still, it maintains real-time compatibility [Bus09]. According to Buss, this approach was first used for IK by Wampler [Wam86] and Nakamura and Hanafusa [NH86].

damped least squares (DLS)

$$\Delta\vec{q} = J^{T}(JJ^{T} + \lambda^{2}I)^{-1}\vec{e} \tag{7.5}$$

The damping λ should be small enough to get stable results for $\Delta\vec{q}$ but big enough to prevent slow conversion to $\vec{p}_{\text{goal}}$ [Bus09].

We also tested and evaluated implementations of the three discussed approaches in simulation and found DLS to be the best performing solution which still meets the real-time requirement on the onboard computers on the Meka M1 [Six18] (SR 5(a)). Thus, I selected this approach for generating the ATs for the second part of the reaching motion.

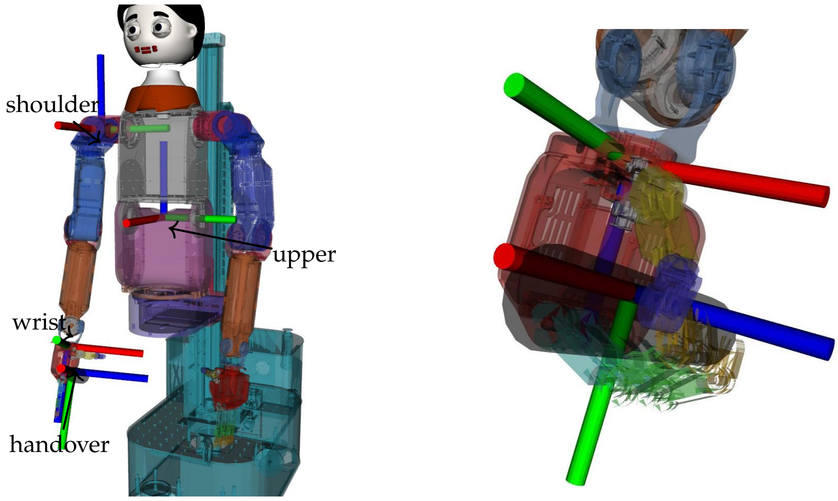
7.2 INTEGRATION WITH FLOKA

In this Chapter I discuss the concepts of a database containing BTs and IIK with DLS for precise adaption to the goal in the context of SR 5(a): Robot Integration as well as SR 5(c): Modularization. As smooth transition between the prerecorded and the dynamic movement is required, both concepts are combined into a single module that handles the arm motions in the HPhase 2: Reach. The trajectories for the first phase are currently manually handcrafted. To show the validity of such an approach I added three BTs to the database that fit the Meka M1. I also made sure that the joints are not close to hardware limits and that the links have some distance to other body parts, so that the DLS method can smoothly take over. Right now the trajectories cover the front area of the robot. For sideways handover the database can be extended in the future. The motion for the HPhase 4: Retreat is also handled by this module by making use of MoveIt's IK calculation, as during the retreat timing and looks are less important.

As discussed in Section 4.2, the Meka M1 offers a JTC interface embedded in ros_control. Thus, inputs have to be provided to the low-level control system as joint trajectories. For the trajectories in the database this format is already maintained. For the generated outputs of the IIK, an appropriate conversion of the data is required. Figure 7.1 visualizes the tf-frames for the right arm, the kinematic chain is constructed of. Below the wrist frame I added another frame that is located at the position were the object resides after grasping (see Fig. 7.1(b)). I also made sure that this frame is close to zero rotation in the roll pitch yaw (RPY) notation when transitioning from the BT to the AT. This helps to prevent running into overflows of rotations and thus a more stable rotation when handing over.

7.2.1 *Reaction to Input*

The proposed concept is based on a continuous input of EEF goals. As reactivity to a user behavior is only possible if his or her actions are perceived, such input is required. Thus, I integrated the *Nuitrack Full Body Skeletal Tracking Software* [Nuitrack] to generate such inputs on the real robot to fulfill SR 4(a): Hand Tracking, SR 5(b): Onboard Sensing and Processing, and SR 4(g): Markerless Perception.



(a) The upper, shoulder, wrist, and handover frames

(b) A detailed view of handover frame, wrist frame, and the virtual object

Figure 7.1: A visualization of the frames for handover reaching and retreat motion generation in *Transform Library* [tf].

Figure 7.2 gives an overview of the proposed reaching motion generation. Based on the frames described in Fig. 7.1 a *Kinematics and Dynamics Library* [KDL] chain is constructed from the shoulder to the handover frame. As a first sanity check, the inputs are validated with this chain by searching for a possible IK solution with KDL. This generates only little computational demand while making sure only valid goals are further processed in the module. If a valid target is received, the BT is selected with Eq. (7.1). The best trajectory is sent to the JTC which starts the motion of the arm. Right before finishing the BT (0.2 s), the adaption algorithm takes over by generating the first adaption that moves the handover frame closer to the current target.

7.2.2 *Blending of Trajectories*

The transition from the BT to the AT motion is done by the smooth trajectory replacement of the JTC. It offers multiple modes of blending over to a currently executed trajectory. These modes depend on the start time send with the trajectory. Here I chose to send a zero trajectory time, as this removes any delay while still blending over to the new trajectory smoothly. Here, it is important to send a new trajectory before the old trajectory finishes. This holds true for both the BT in the database, and the later-on generated motions generated by the DLS method. Otherwise, the JTC would stop the robot between two trajectories, which would result in jerky motions that are generally slower and without gesture characteristics. Figure 7.3 visualizes the replacement of a currently executed trajectory by a new one. It can be

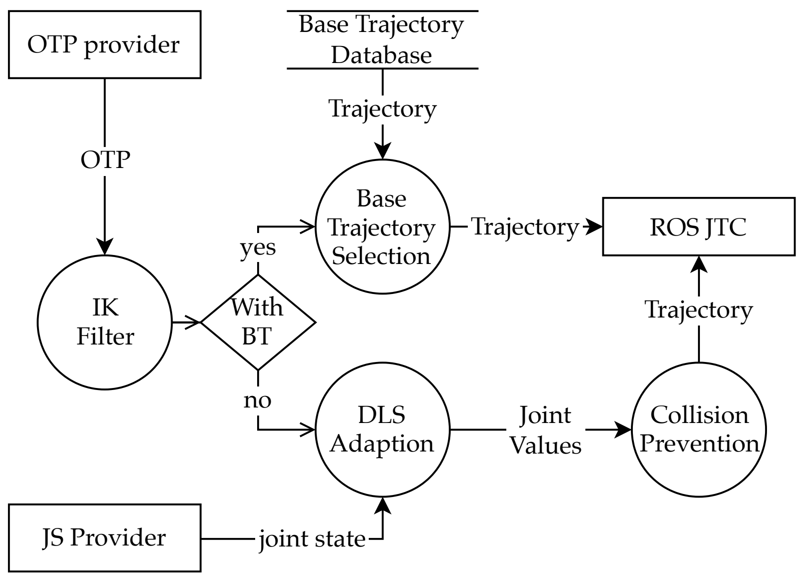


Figure 7.2: The reaching trajectory generation visualized as data-flow diagram (DFD). A continuous stream of OTPs are used to either select a BT or AT based on the phase controlled by the behavior engine. The result is received by the JTC that runs in the real-time loop of the robot.

seen that the transition is smooth on the joint level. As soon as a new trajectory is received, the JTC blends over to the new way-point.

7.2.3 *Adaption Motion Generation*

Besides the target position, the adaption fetches the current joint positions from ros_control. With a rate of 100 Hz the DLS is calculated based on the current inputs. Before sending a goal to the JTC, MoveIt is queried whether the new joint positions are free of collisions. I added the object to the ROS URDF as seen in Fig. 7.1(b) to also ensure that the carried object is collision-free as well. This check is enabled/disabled based on the carrying state of the robot. We tested multiple ways of calculating the orientation of the hand during approach and exchange (HPhase 3: Transfer). The first was keeping the orientation of the handover frame as it was after finishing the BT. Another one was based on the relative position of the handover frame in regard to the robot's shoulder. The algorithm turned the hand to always face away from the shoulder. While this approach produced reasonable results especially for sideways handovers, it was in some cases less predictable than keeping a fixed orientation during the process [Six18]. To control the speed of the AT I introduced two means of damping. The first damps the goal and sets priorities for the positional and rotational parts of the goal. The other damping can be set at runtime to scale the velocity of the whole motion. Another fac-

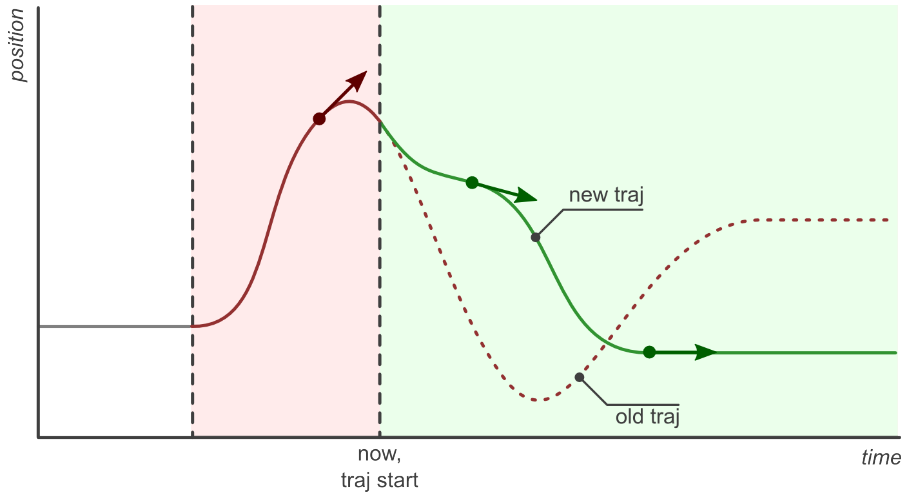


Figure 7.3: Trajectory replacement in the joint trajectory controller of *ROS Control* [ros_control] with zero trajectory start time [t.r.].

tor influencing the velocity is the trajectory duration parameter that needs to be provided to the JTC. Together with the newly calculated target joints it provides a velocity measure. I also added a per joint speed limit as an additional safety and smoothness measure. Here it needs to be taken into account that limiting the per joint motion too much, might result in bad convergence to the actual goal. This speed should be controlled with the damping and the limit only be employed for safety/sanity checks.

7.3 RESULTING MOTIONS

In this chapter I discuss the practical implications of the presented approach by showing results from execution in a simulated environment as well as on the robot's hardware. This includes verification of reachability of different goals distributed in the robot's workspace and checks for self collisions during execution.

7.3.1 *Reaching in Simulation*

To show the feasibility of the stated approach, I ran the system with a virtual version of the Meka M1 in Gazebo. This provided the same interface as the real robot. I generated input targets by sampling the space in front of the (simulated) robot in a X=0.2 m to 0.7 m, Y=−0.6 m to 0.4 m, and Z=−0.2 m to 0.8 m workspace, originating at the upper frame (Fig. 7.1) I selected a sampling resolution of r=0.05 m, resulting in 4851 sampled targets in the $0.5 \times 1 \times 1$ m box for the right arm. During execution, the position data, goals, collisions, and the duration of each run were recorded for analysis and visualization. The timeout was set to 6 s, to proceed with the next goal if the robot is stuck. The target goal distance was set to 0.045 m. After reaching one of those thresholds, the arm is reset to a zero position and the next target is given to the module. A limitation of this approach is that the goal was kept constant during execution which would be only the case with a perfectly predicted OTP. As a simulation of the human behavior is a challenge itself, changing goals are evaluated on the real robot.

Figure 7.4 visualizes the reachability of the robot's EEF. The targets that caused timeouts are mostly due to being either too far from the end-points of the BT, which could be improved by adding more points to the database, or by the torso blocking the motion of the arm in the positive Y direction (left of the robot). Here one could also add more BTs that keep bigger distance between torso and elbow to allow for a bigger motion to the left. On the other hand one could reposition the robot or use the other arm for handover in such situations. Due to the human-like physiology of the Meka M1, the reachable spots are in a region where a human would also reach. These targets were also reached fastest.

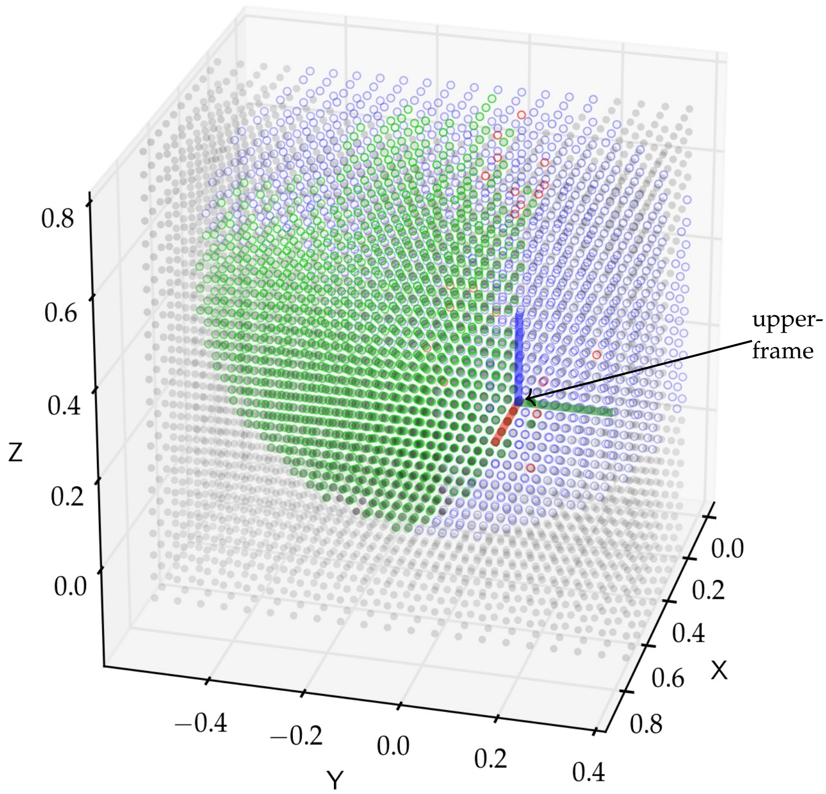


Figure 7.4: Visualization of the analysis of the workspace of the Meka M1 during the adaption. Gray dots mark targets rejected right from the beginning. The darkness of the points is mapped to the reaching time, were darker means shorter duration. Green edges mark successful execution. Red edges mark cases where a collision with the internal model was detected. Blue edges mark timeouts.

Figure 7.5 shows the end positions of the handover frame at the end of the AT. It highlights again the limiting introduced by the torso during left directed motions. Thus, clusters originate from joint limits and the implemented self-collision prevention that stops the robot before touching itself.

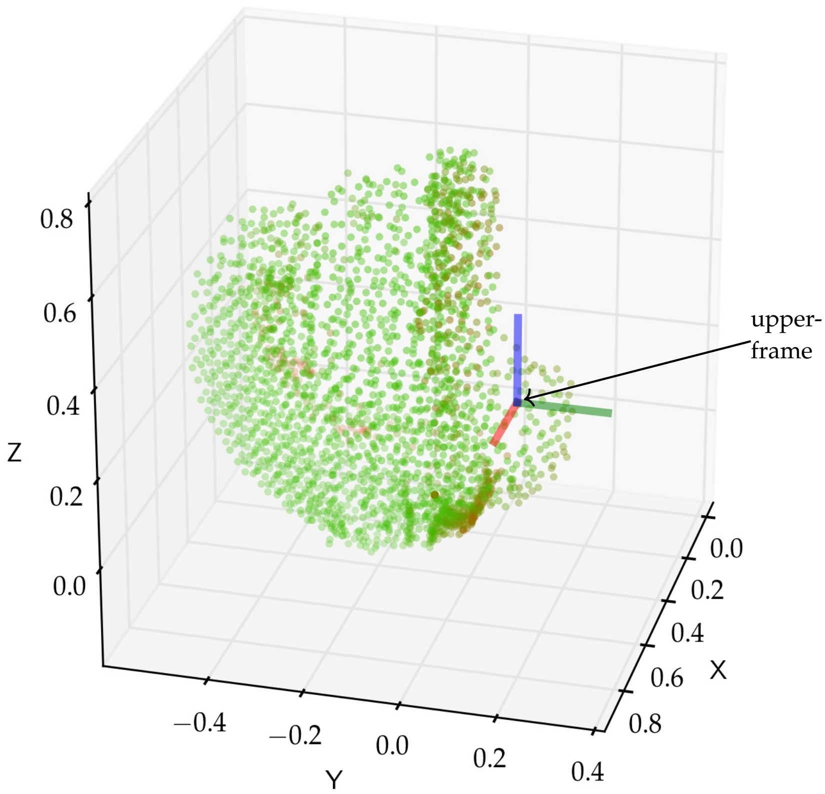


Figure 7.5: Reachable points by the AT. Green to red color encodes the final distance to the desired target. Red means bigger offset to the goal. For different perspectives on this data refer to Fig. C.2.

The resulting ATs are shown in Fig. 7.6. As the plotting starts after the three different BTs in the database, the motions are clustered into three groups. Most of the trajectories look mostly smooth. Only the longer, as well as ATs on the left side show some minor oscillations before reaching the goal. This highlights the importance of starting with good and close BTs for predictable reaching motions. Also, with an appropriate stopping distance and timeout, oscillations can be successfully prevented.

7.3.2 *Reaching Motions on the Real Robot*

While the results from the simulated environment look promising, they can not be directly transferred to the robot. Challenges include a moving target and imperfections in the robot's motions. Especially

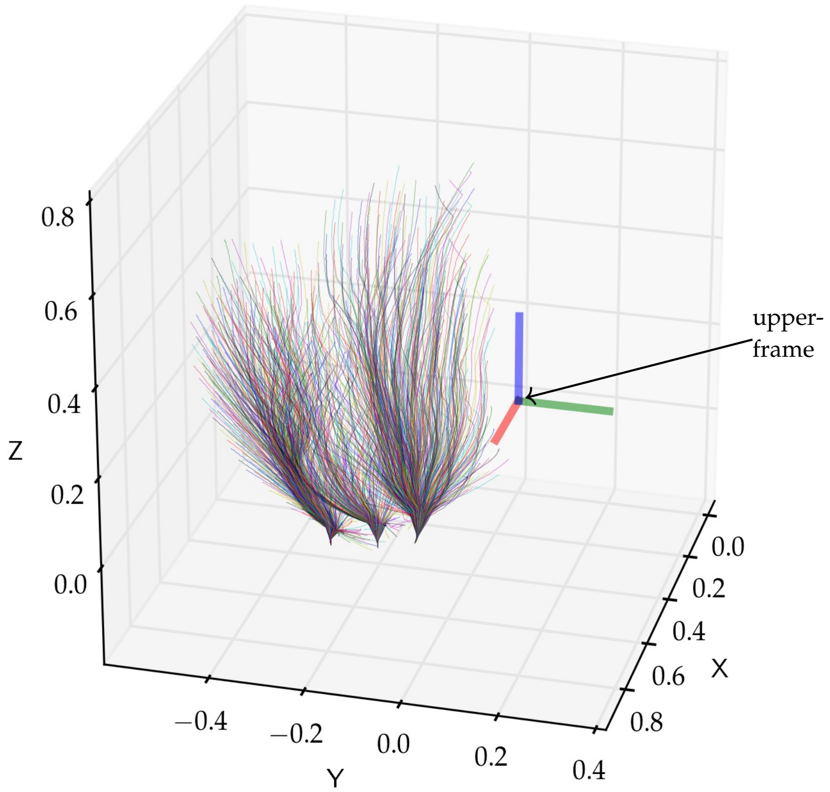


Figure 7.6: Visualization of the generated ATs. Each trajectory with an individual color. The selection of different BT based on the target can be seen by the three clusters of starting points of the AT. For different perspectives on this data refer to Fig. C.3.

the combination with an approach that dynamically calculates new joint goals and thus relies on reading the current joints as input is a complex control problem. While SEA have many advantages, their compliance needs to be taken into account when setting new joint targets compared to the precise simulation. Also in the real world external forces applied by the human and the exchanged object require higher control inputs to the joint drives for a successful AT execution. It showed that the damping needs to be above a given margin, so that the robot starts moving at all [Six18]. Also, the joints need to be calibrated precisely in terms of torque and positions. Otherwise, errors in the IIK control add up quickly and cause the arm to drift away from the target.

With a horizontally mounted *Intel RealSense D435*[Int17] camera in the Meka M1's torso, Nuitrack provided 3D positions of the person standing in front of the robot. While it provided the human hand position with 30 Hz and only little deviation, the narrow horizontal FoV limited the interaction space of human and robot. Also, it only provided the current and not the future position of the hand. Thus, the target is changing all along. Especially for selecting the correct BT this poses a problem. Furthermore, it provides only the ability to reactively follow the human hand [Six18]. Thus, SR 4(b): Object Transfer Point Prediction becomes important to select the correct BT early and reduce the amount of adjustment by the AT. Therefore, I present improved hand tracking along with prediction in the following chapter.

Summarizing, the combination of BTs and ATs showed to be an improvement compared to a static reaching motion as it increases the interaction space already with only few recorded trajectories, while maintaining gesture-like characteristics.

8 PERCEPTION AND PREDICTION OF OTPS

As the initial handover experiment revealed, a strategy that relies on a fixed gesture motion and forces to detect the intention to transfer is not enough. A robot needs visual perceptive capabilities for handover. The requirements SR 4(a): Hand Tracking and SR 4(g): Markerless Perception call for a camera-based hand tracking system. Additionally, the tracking system needs to be integrated with an existing robot (SR 5(a): Robot Integration). I aimed to replace the previously integrated Nuitrack module to have full control over the tracking and to be able to apply optimizations for a fast and predictive handover. Another module that fulfills SR 4(b): Object Transfer Point Prediction can be added on top of the output of the tracking module. Therefor, we developed tracking and prediction modules, optimized for human-robot handover and integrated it with the Meka M1 [Sim+19].

8.1 HAND TRACKING

Existing approaches can be distinguished by several factors. These include the number of cameras used, whether color, depth images or a combination of both is used and if the output is only one point describing the hand center position or if the hand's articulation is tracked.

Approaches making use of multiple cameras to detect and track hands proved to be accurate and reliable [SOT13; UO99]. However, for a self-contained robot such approaches are infeasible as they require multiple cameras from different viewpoints in contrast to egocentric vision provided by a handing over mobile robot (see SR 5(b)) Detecting hands on single red, green, blue color model (RGB) images has also been proposed for human-computer interaction [DB09]. Also, tracking on RGB images from a moving camera has shown good results [CR00]. Even on single gray-scale images hands can be detected [KT04]. Unfortunately, these approaches lack 3D information which is mandatory for interactions that take place in the 3D space. Some approaches extend these RGB based algorithms on two cameras to generate stereo images for a depth representation [Ngu+18a; Man+08].

The proposed system should enable a robot to react to a handover with a human-like timing (see Section 3.4.6). But not only the average handover time should be considered, but also the reaction time which was found to be ≈0.43 s [NDL18]. Therefore, the processing of the hand tracking and the planning of the robot's motion should only induce a minimal delay below 0.4 s. Approaches with a high processing time (e.g. Bray, Koller-Meier, and Van Gool [BKV04]) do not fulfill these requirements. Hackenberg, McCall, and Broll report a rate of more than 50 frames per second (FPS) and rely only on depth images [HMB11]. However, they assume strong visibility constraints, e.g., the palm and fingers have to be always facing the camera. Same visibility constraints apply for techniques that focus on hand detection in which the hand has to be close to the sensor for input gesture recognition [Qia+14; MES18]. While the proposed algorithms give precise posture information of the users hand from depth images they require close and completely visible hands, which is not always the case in handover scenarios with e.g. occluding objects and longer approach distances. Another approach is to detect or track the whole body posture in point clouds. Ehlers and Brama incorporate the people detection algorithm by Munaro, Basso, and Menegatti [MBM12] and fit a self-organizing map (SOM) to the detections. This approach has the advantage of solely requiring a depth camera without color information and yields the whole body posture of the human [EB16]. Here again the visibility constraints come into play. The authors report problems like hands being detected as heads when held to far up and detection problems with rotational movements of the camera which can occur when the robot orients itself towards the user during handover (see SR 2(a): Body Orientation).

8.2 TRANSFER POINT PREDICTION

For a fluent handover, detecting and tracking the human hand is only half of the story as this provides only data of the current world but not the future. This would only allow following the human hand but not synchronized meeting in space at the OTP and at the same time. In contrast, a prediction allows to send the needed actuator commands as soon as possible (see SR 4(b)). In most existing prediction approaches human-human interaction was analyzed to generate models and validate them in post processing. This enables us to later transfer these models to robots [Shi+97; Hub+09b; PAN17; Str+12; Pan+17; MHB15; Han+17]. Some even use the data, recorded during human-human handover, to train models or incorporate imitation learning. This requires many demonstrations or interactions with a system to gather the required samples [PS15; Mae+17b; MBR17]. Properties of the giver, like kinematics and body weight of the human, need to be determined beforehand to adjust such models [PAN17]. To generalize from individuals or specific scenarios the training process can become very demanding. Also, the transfer from the human physiology to a robot with different properties is another challenging task. [Har+15] stress the importance of prediction and present a first draft of a predictor that takes skeleton data as input to predict timing and position of the OTP. Their solution predicts only very little of the future, meaning that the prediction becomes accurate only after already seeing almost the complete human motion. The authors also state that, to be of use, the predictions need to be available earlier [Har+15]. Li and Hauser propose different models for predicting reaching motions [LH15]. However, only a limited number is evaluated.

Pan et al. make use of a precise tracking system with markers on multiple body positions to determine the important kinematic features to detect and predict handover. It was found that the right-hand velocity is one of the most important features [Pan+17]. A similar experiment by Hansen et al. showed that the object mass had only little influence on the height of the OTP. The duration is influenced by the distance of the interactants [Han+17].

The prediction module I present here, builds upon the idea of Nemlekar, Dutia, and Li and is combined with the approach of Li and Hauser [LH15; NDL18; NDL19]. It incorporates the minimum jerk model as proposed by Flash and Hogan [FH85] to predict an integrated object transfer point ($\mathrm{OTP}_{\mathrm{integrated}}$).

8.3 integration with floka

To address SR 5(a): Robot Integration, the provided inputs, and required outputs need to be taken into account, like providing targets for the reaching module (see Chapter 7). The concept of the tracking module consists of detecting the human in front of the robot once with a state-of-the-art detector and use this information to track the human hand in a point cloud. This can be done fast and resource efficient in a big interaction area. Obtained position and velocity is forwarded to a prediction component that provides an $\mathrm{OTP}_{\mathrm{static}}$ and $\mathrm{OTP}_{\mathrm{dynamic}}$, which are than fused to a single $\mathrm{OTP}_{\mathrm{integrated}}$ that can be used as input for the reaching motion module. Figure 8.1 gives an overview over the proposed pipeline. It starts by receiving an external trigger of the behavior coordination, e.g. when the person comes into the Space 2: Personal of the robot. At first, the posture of the person is extracted from the RGB image with OpenPose. Important keypoints like the wrist, torso and hand position are combined with depth information to extract 3D positions. The hand position is used to initialize the newly developed pointcloud based hand tracking. Torso and wrist positions serve as inputs for the $\mathrm{OTP}_{\mathrm{static}}$ computation. The tracking results are passed to the prediction for estimating an $\mathrm{OTP}_{\mathrm{dynamic}}$. These are then combined to a single $\mathrm{OTP}_{\mathrm{integrated}}$.

8.3.1 *Optimal Sensor Mounting*

As I aim to fulfill SR 5(b): Onboard Sensing and Processing a sensor is required that is small enough to be mounted inside, or on the

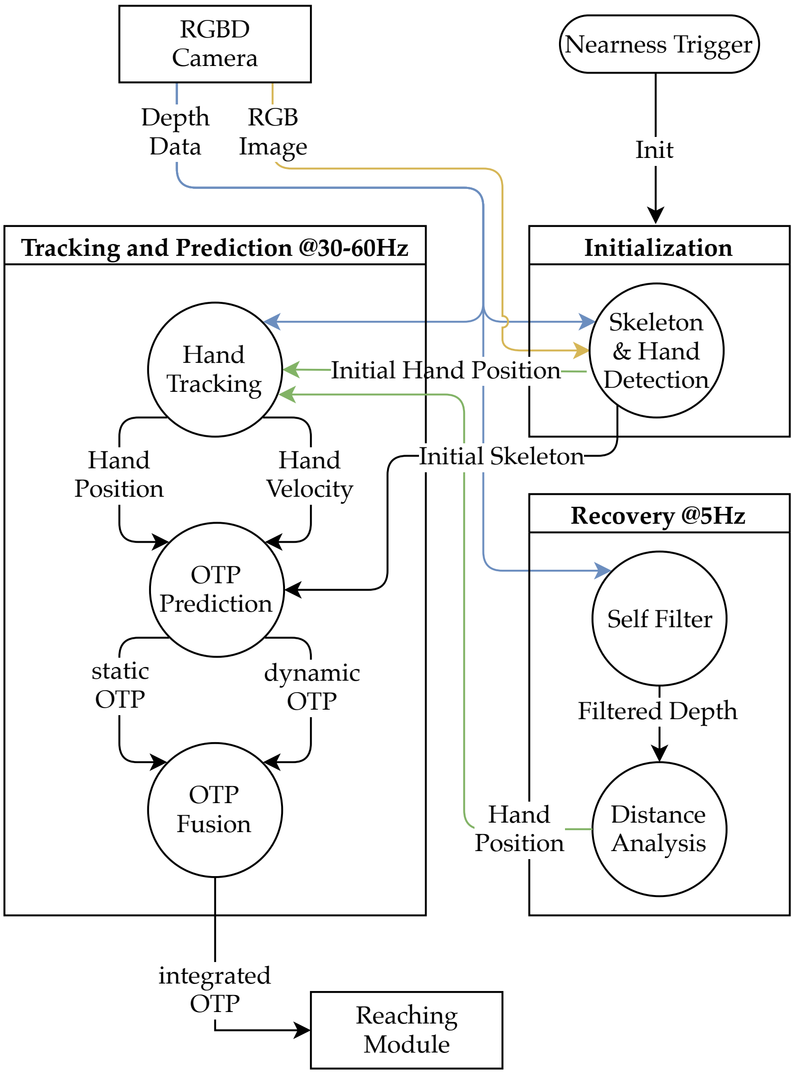


Figure 8.1: The proposed hand detection, hand tracking, and OTP prediction pipeline visualized as DFD. An external signal, e.g. when the person comes into the Space 2: Personal of the robot, starts the processing pipeline. Then the skeleton is extracted from the RGB image, transferred into 3D space and given to the tracking as initialization and to the prediction for computing the $\text{OTP}_{\text{static}}$. The tracking results are passed to the prediction for estimating an $\text{OTP}_{\text{dynamic}}$. These are then fused to a single $\text{OTP}_{\text{integrated}}$.

Meka M1. I decided to take the Intel RealSense D435 stereo camera, as this sensor provides a depth perception with a large FoV and high FPS rate. Providing 91° horizontally and 65° vertically, resulting in 100° diagonally FoV. The calibrated color-image provides a narrower FoV (HxVxD: $59\,^{\circ} \times 42\,^{\circ} \times 77\,^{\circ}$) and thus is not as good for close interaction. By incorporating dual global shutter sensors, this sensor provides undistorted data on fast motions of the interactant [Int17]

For the mounting we need to find a position that covers the interaction area during handover. Especially situations where the human hand is close to the robot need to be considered when deciding on a position. Other limitations include that cables do not block joints of the robot, and parts of the Meka M1 do not occlude the optics. Thus, the resulting position is at the top of its torso with an angle of 18° [Sim19]. Figure 8.2 visualizes the position of *Camera 1* which is optimized for handover and *Camera 2* for face detection during interaction.

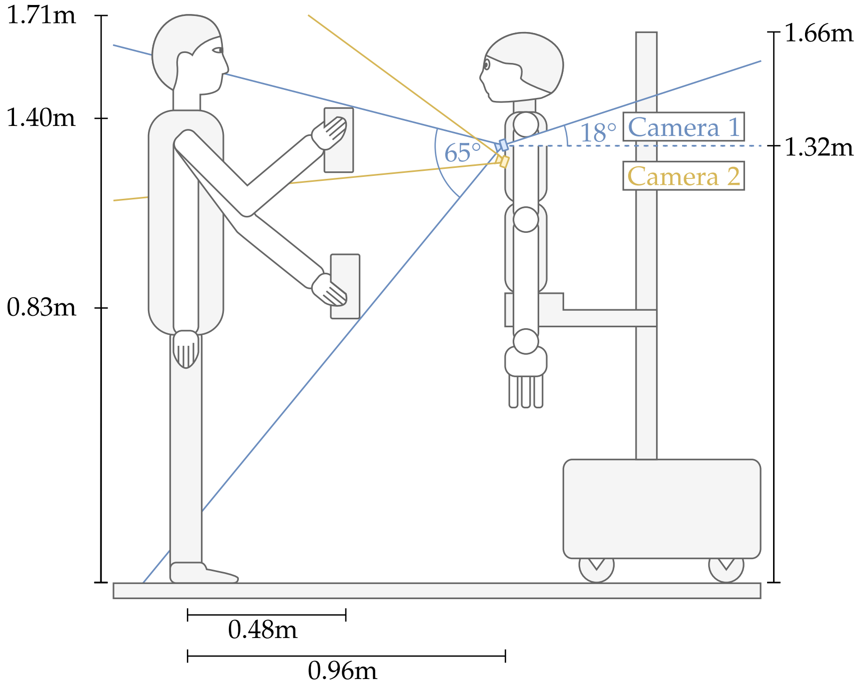


Figure 8.2: Integration of the RealSense D435 [Int17] cameras into Floka for best visibility of the workspace for hand tracking during handover. Camera 1 is mounted at 1.32 m, when the z-lift is at the default position of 0.3 m, and angled at 18°. Camera 2, for face detection, is attached below Camera 1. The human is modeled based on the averages found in literature.

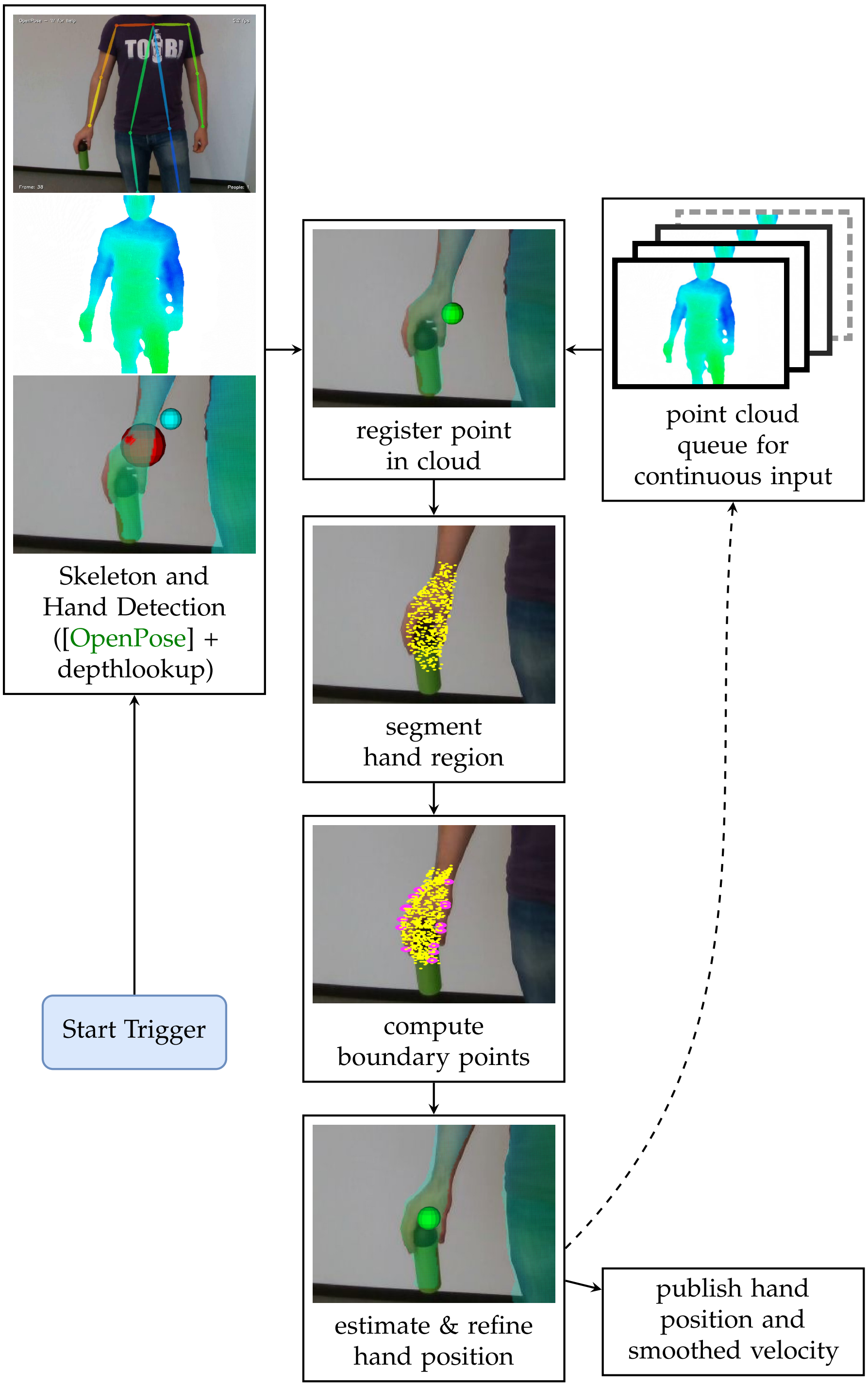


Figure 8.3: The hand tracking pipeline visualized as a flow-chart. *OpenPose* [OpenPose] serves as initialization, during which point clouds are buffered for a continuous tracking. Region growing and boundary estimation refine the hand position in each run. The colored images in the back are only added to give context in the chart. A four times zoom at the hand region is applied for a closer look at the relevant area.

8.3.2 *Hand Tracking Pipeline*

The module to track human hands purely relies on point clouds obtained by a depth sensor. Combined with the *RealSense D435* [Int17] this allows a big interaction area. By not relying on color information during the tracking, the system becomes robust to lighting changes and e.g. wearing gloves. The approach extends on the ideas of Chen et al. by combining region growing with edge detection and position refinement strategies [Che+11]. While the authors propose an artificial hand movement to initialize the tracker, I propose to combine the tracking with OpenPose, which gives precise and reliable results, but still takes time to compute. Extended with recovery strategies for situations where color information is not available, create a robust hand tracking module. This combination results in a fast and reliable input provider for the OTP prediction.

Figure 8.3 shows the essential processing steps of the detection and tracking pipeline. When the pipeline is started, OpenPose's hand detection module returns the person's hand position [Sim+17]. While this computation is running on the RGB image, the perceived depth is stored in a queue. As soon as the initial hand position is available, the tracking algorithm takes a cloud from the queue and matches the result to a point in the cloud. This region is grown with the received point as a seed. Boundary points are computed from this region as limbs have more border points like e.g. the torso. These points are used to calculate a new center position of the hand. which is the output of the current tracking result and input for the next tracking step. The pipeline was integrated as a *ROS nodelet* [nodelet] to receive the sensor data with zero copy. This reduces I/O-load and thus results in faster processing.

8.3.2.1 *Tracking Initialization*

For a successful tracking, a good initial position is required. Chen et al. proposed to use a special clicking gesture to initialize the tracker. While in some scenarios such an approach might be reasonable and has the advantage of being completely independent of color information, here, for natural handover initial, artificial gestures are not feasible.

Deep learning techniques for image processing become increasingly popular in computer science, as well as robotics. While some algorithms are already well-developed, they mostly require huge processing power, like GPU are other specialized hardware to run fast. Even if higher FPS rates (>30) are achieved, delays for uploading the data to the GPU and pipelining the processing need to be taken into account. Especially two phase processing, which first detects the whole-body pose and then searches a hand near the wrist create a significant delay (>200 ms). Other downsides consist of visibility problems in close interaction, where such approaches heavily rely on context for body parts. As the quality of the detection in controlled environments is currently unmatched, I propose to integrate such techniques for initialization of a tracking system. To overcome the visibility constraints, the initialization should take place during the HPhase 1: Approach. Here, usually, human and robot are facing each other and the algorithm can start when a good visibility is achieved, as well as having time to process the queued depth data until the handover is taking place.

Scenarios that prevent retrieving a good initialization with the interactant's position, like when having the arms crossed or the hands in the trouser pockets, require the tracker to recover such cases. I present strategies for such cases in Section 8.3.2.3.

8.3.2.2 *Continuous Tracking*

The goal of tracking module is to achieve a low delay by updating the track continuously. The approach is based on the assumption that the hand position changes only a certain amount between two consecutive depth inputs. Thus, it requires only once a prior which is afterwards generated by maintaining the track. Together with the velocity, calculated based on the last two frames and the current position, a new seed for the next frame is calculated. Proceeding from that point, the goal is to find points belonging to the hand of the interactant. It calculates and marks the boundary of the hand in a radius r. This is done by computing the point normals for each of the point in the region. An angle criterion classifies the points as either boundary or non-boundary point. While normal computation is computational heavy, it showed to be more stable than a simple neighborhood distance criterion. As limbs and especially fingers cause lots of boundary points to be found, it is a fast and stable approach to track a hand. As the aforementioned cropping generates a smaller point cloud with boundary points at the limits of the cloud, a second cropping step with a smaller radius is required to remove the boundary points that solely arose due to the first crop. A mean-shift step refines the hand position [Che+11]:

$$x_{\text{hand}} = \frac{\sum_{p \in r_{\text{hand}}} p * \mathbb{1}_{d(p, x_{\text{approx}}) < d_{\text{ms}}}}{\sum_{p \in r_{\text{hand}}} \mathbb{1}_{d(p, x_{\text{approx}}) < d_{\text{ms}}}} \tag{8.1}$$

The calculated position x_{hand} is the result of the current hand tracking and serves as the input for the next tracking step.

8.3.2.3 *Tracking Recovery*

As there might be some scenarios were the tracking gets lost, like the hand being completely covered or being so close to the body that no boundary can be estimated, a reliable module requires means of recovery. I implemented a strategy that regularly checks the point cloud for points close to the robot that have hand-like properties. This heuristic allows the robot to reestablish a track when the interactant reaches for a handover. It is executed in a bounding-box covering the handover area. The closest point to the mid-point between giver and receiver is used as an input to the tracking explained in Section 8.3.2.2 on the facing page. To not interfere with the performance of the main-loop of the tracking, this heuristic is limited to be executed each fifth frame.

This approach showed to require protection against the following scenarios: The human torso being really close, noisy data from the depth-sensor, and data from the Meka M1's own hand being inside the FoV of the sensor. The first problem, of the upper body of the interactant stepping into the Space 1: Intimate, where only the hand would be expected, was resolved by setting an appropriate radius for the boundary detection. Thus, no boundary points can be estimated in the torso area and the track does not become incorrectly assigned to the upper body. As current stereo vision based depth-sensors tend to exhibit noise in certain scenarios, resetting the tracker to such inputs needs to be prevented. Therefor, requiring a minimum number of points at the new location showed to be a successful measure in preventing a random jump.

Another challenge was to keep tracking the human hand, while the robot's hand was also in the sensor's FoV. Both, the tracking itself and especially the recovery strategy, tend to be attracted by the Meka M1's hand the same way as by the human hand. Thus, I adopted self filtering of body parts of the robot from the point cloud. The approach is based on the *Robot Self Filtering* [self-filter]. I extended it to take a list of links to filter based on the collision primitives (refer to Fig. 4.2(b)). This way only links that extend into the FoV of the sensor, like hands, fingers, and the forearm, can be selected for filtering for minimal impact on performance. The same applies for the object carried by the robot, depending on the shape, it might attract the tracking and recovery and thus needs to be filtered for maintaining stability. I added the capability to activate filtering of an object if the robot is carrying it (see Fig. 7.1(b)). Especially in the HPhase 3: Transfer, it is vital to keep a stable track of the human hand, to allow for a visual handover as required by SR 4(f): Visual Transfer. As in this case the hand and EEF come really close, the resolution of the sensor might not be able to provide a clear boundary between them. This could cause the track move between human and robot, which is prevented by applying a self-filter.

8.3.3 *OTP Prediction*

Addressing SR 4(b): Object Transfer Point Prediction the goal is to have an estimate of the position where the handover will occur. It is crucial to have the best prediction as early as possible, especially when an early and predictable motion of the robot is required. Hence, Section 3.4.5 already introduced the basic idea of an online and offline component of such a prediction.

$\text{OTP}_{\text{STATIC}}$ ESTIMATION The static object transfer point is calculated once when the handover is initiated. It aims to give a first estimate to select an appropriate starting motion for the robot's arm. The position is based on findings from previous human-human handover-studies. Basili et al. e.g. found that the distance between head and hand during handover are correlated [Bas+09]. As we optimized the FoV for hand-visibility, head positions are replaced by the upper torso position. As mentioned in Section 3.4.5 on page 38, the OTP is roughly at the midpoint between giver and receiver.

Based on these findings I propose the following equations for an $\text{OTP}_{\text{static}}$ estimation:

$$\begin{aligned} \text{OTPs}_{\text{height}} &= 0.65 * \text{wrist}_{\text{height}} + 0.35 * \text{torso}_{\text{height}} \\ \text{OTPs}_{\text{distance}} &= 0.4 * \text{hand}_{\text{dist}} \\ \text{OTPs}_{\text{offset}} &= \text{hand}_{\text{offset}} + (\text{torso}_{\text{offset}} - \text{wrist}_{\text{offset}}) \end{aligned} \tag{8.2}$$

The height depends on the person's torso height which is the same as the shoulder position. Offsetting it by the current wrist position brings the arm length in. The distance is roughly the midpoint with a tendency towards the robot. Lateral displacement is based on the current hand position in relation to wrist and torso distance. Whilst wrist and torso position are only estimated once during the HPhase 1: Approach, the hand position is based on the most recent track when the initiation of handover is detected.

$\text{OTP}_{\text{DYNAMIC}}$ PREDICTION The dynamic object transfer point prediction is based on the current data of hand tracking. Based on the most recent hand position and velocity the current $\text{OTP}_{\text{dynamic}}$ is calculated. We decided to modify the minimum jerk model of Flash and Hogan as this is in line with findings about human motion (see Section 3.4.6) [Sim+19]. This approach has the advantage of not requiring any additional training data (cf. [NDL19]).

The original work by Flash and Hogan is based on the assumption that a major goal of motor coordination is keeping the rate of change at a minimum. If duration and start position are known, the following equation models the hand trajectory [FH85]:

$$x_c(t) = x_0 + (x_0 - x_f) * (15\tau^4 - 6\tau^5 - 10\tau^3) \tag{8.3}$$

with $x_c(t)$ being the position at a given time t with $\tau = (t - t_0)/(t_f - t_0)$ and t_f the final time, x_0 is the start position at $t = t_0$ and x_f the final position. The goal is to solve the equation for x_f as this should be the OTP. To make use of Eq. (8.3) to predict the final position the duration is assumed with $1.24 \pm 0.28\,\mathrm{s}$ as found to be the average duration of the HPhase 2: Reach [Bas+09]. As this is only an average estimate, t_f can be updated during the observed motion. Taking the bell shape of the velocity profile into account, updates are possible at maxima of acceleration and velocity.

The point of maximum velocity is reached when half of the movement is executed, meaning if a decrease in velocity is detected, t_f can be updated with the following equation:

$$t_f = t_0 + 2 * (t - t_0) \tag{8.4}$$

The extrema of the acceleration can be found by solving the derivative of the acceleration for zero:

$$(x_f - x_0) * [360 * \tau^2 - 360 * \tau + 60] = 0 \tag{8.5}$$

$$\Leftrightarrow \begin{cases} \tau = \frac{3-\sqrt{3}}{6} \approx 0.21 \\ \tau = \frac{3+\sqrt{3}}{6} \approx 0.79 \end{cases} \tag{8.6}$$

The resulting equations can be solved for t_f to predict the end time of the handover:

$$\frac{t - t_0}{t_f - t_0} = \frac{3 \pm \sqrt{3}}{6} \Leftrightarrow t_f = \frac{t - t_0}{\frac{3 \pm \sqrt{3}}{6}} - t_0 \tag{8.7}$$

Thus, the minimum jerk model for x_f, which is the predicted end position of the movement and therefore the $\mathrm{OTP}_{\mathrm{dynamic}}$, is:

$$\mathrm{OTP}_d = \frac{h_c - h_0}{10 * \tau^3 - 15 * \tau^4 + 6 * \tau^5} + h_0 \tag{8.8}$$

The error of the model should decrease over the observation time.

FUSING AN $\mathrm{OTP}_{\mathrm{INTEGRATED}}$ The $\mathrm{OTP}_{\mathrm{integrated}}$ is the combination of an $\mathrm{OTP}_{\mathrm{static}}$ and $\mathrm{OTP}_{\mathrm{dynamic}}$. I discuss an approach on how to blend between both predictions. Ideally one would want to blend to the $\mathrm{OTP}_{\mathrm{dynamic}}$ as soon as it is a better prediction of the actual OTP. As this point in time is not computable, an estimate is required. The interpolation can be written as [NDL19]:

$$\mathrm{OTP}_{\mathrm{integrated}} = W * \mathrm{OTP}_{\mathrm{dynamic}} + (1 - W) * \mathrm{OTP}_{\mathrm{static}} \tag{8.9}$$

To estimate the weight I suggest linear blending, based on the duration. We introduce t_s as a time when the dynamic prediction gives reasonable results, depending on the accuracy and resolution of the tracking [Sim+19]:

$$W_{lin} = \frac{t - t_s}{t_f - t_s} \tag{8.10}$$

I incorporated an approach based on three phases. In the beginning it fixates the $OTP_{integrated}$ at the static prediction, then it applies the blending, and in the last phase, when the $OTP_{integrated}$ is already close to the hand position, the $OTP_{integrated}$ gets locked to the current hand position. Figure 8.4 visualizes the process of generating an $OTP_{integrated}$ prediction and update it during the interaction, based on the current observations.

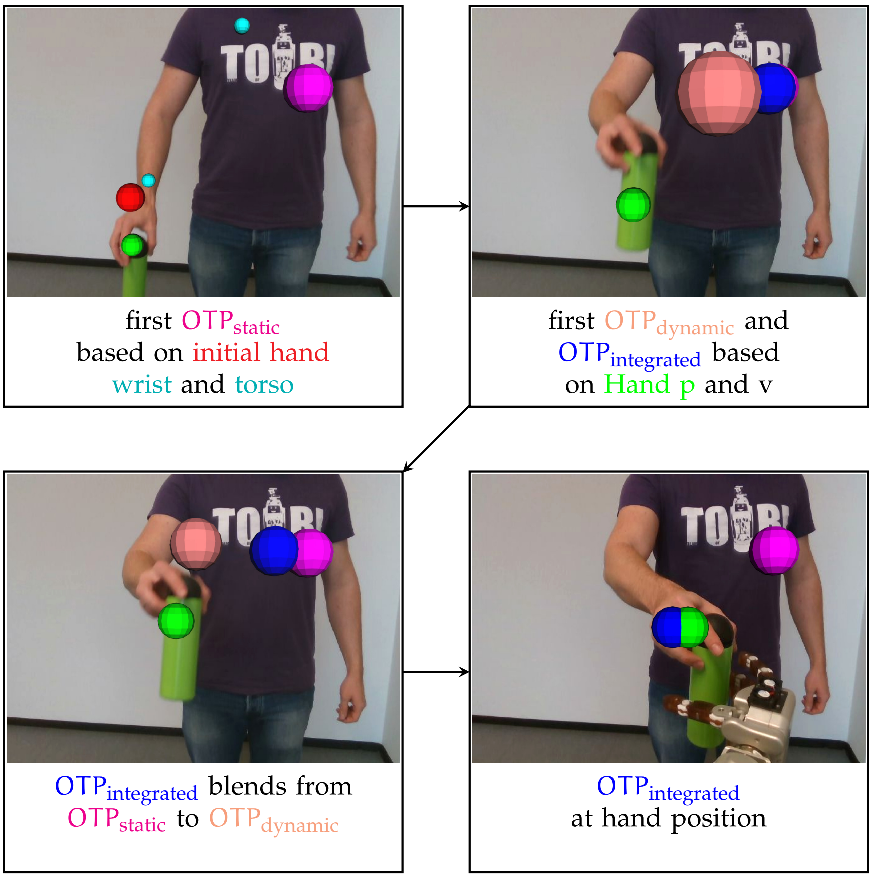


Figure 8.4: Visualization of the prediction during the interaction. Torso and hand position (turquoise) are used to predict an initial OTP_{static} (magenta), the $OTP_{dynamic}$ (light red) is predicted based on the hand start (red) and current position (green), OTP_{static} and $OTP_{dynamic}$ are fused to an $OTP_{integrated}$ (blue) until the hand is close to the prediction.

8.4 TRACKING AND PREDICTION RESULTS

To assess the performance of the proposed tracking and prediction pipeline we tested it in various handover configurations [Sim+19]. The position was varied in terms of approaching angle with three different angles ($-35°$,$0°$,$35°$). Several object types and sizes, as well as, persons performed handover motions in front of the robot. Additionally, the hand start-position was varied in terms of distance to the body, by passing from one hand to the other, or having the arm's folded. The resulted in a total of 21 recordings of a variety of possible interactions.

As the algorithm's outcome might vary over runs because of e.g. random downsampling, skipped frames, etc., each recording is tested five times to rule out random results. For this evaluation the recovery strategy (see Section 8.3.2.3) was not active to analyze the tracking and prediction performance without interference. Due to the huge variety of assessed cases, a general statement of the tracking performance is not possible. The cases where the hand was visible and separated from the body generally gave a stable tracking. In these scenarios the tracking algorithm only failed when the initialization got no results because of the person being not in the view of the camera as expected. Poorest performed scenarios where the hand was too close to other body parts and thus gave no separable border points. Also, situations where the hands are hidden because of the arm's folded allowed no correct initialization. These are the cases motivating the recovery approach, although crossing arms might not be likely when planning to give or receive an object. Besides the general tracking capability, the delay was analyzed. Due to the OpenPose based initialization which takes 0.3 s to 0.4 s until the tracking takes over, the initial delay is at least 0.3 s. After about 0.5 s, most of the queue is processed at the delay drops to below 0.1 s. After the complete queue is processed the algorithm is able to keep up with the camera rate of 30 fps, resulting in a delay of $\approx$0.05 s. The general idea of initialization during the HPhase 1: Approach for a low-latency tracking in the HPhase 2: Reach shows to be a valid approach. Summarizing, it fulfills the SR 4(a): Hand Tracking by providing a module in terms of SR 5(c): Modularization that can be integrated in robots (SR 5(b)) while being independent of any artificial markers or environment modifications (SR 4(g)).

We took 19 of the recordings to evaluate the OTP prediction, leaving out the runs, were tracking initialization failed [Sim+19]. Two different sized boxes serve as a quality classification, for which the prediction was compared to the actual final position. A coarse box with the dimensions of: $0.2 \times 0.35 \times 0.35$ m and a smaller one with: $0.1 \times 0.1 \times 0.1$ m. Only two of the runs were outside the coarse box, where the final OTP has been estimated outside the box. In all other runs the OTP prediction stayed in the defined area over the whole motion from start to finish. After about 75 % eleven cases are within the smaller box, which is only about the size of a hand. Overall, the concept and implementation of a predicting module (SR 5(c): Modularization) for SR 4(b): Object Transfer Point Prediction with a pipeline that at first gives a rough estimate of the OTP, allows to select a trajectory from the database as proposed in Section 7.1.1. Over time the prediction gets more precise allowing to precisely adapt to the latest provided target.

9 COMBINED HANDOVER SYSTEM

In this chapter I go into details of the final system, addressing all the requirements by presenting behavior strategies for robots as well as recovery methods for reliable handovers that follow the paradigm of SR 1(a): Human-Like Pattern and SR 1(c): Reactive Pattern. Based on the results explained in the previous chapters I combine the modules (SR 5(c): Modularization) to a coherent system. I describe how these modules interact and how the gaze cues are integrated in the Floka humanoid (SR 5(a): Robot Integration). In contrast to the first implementation where the system had the initiative (cf. Chapter 5), expressed by starting the motion before the user, I aimed to have the behavior more user focused by reacting to its initiative. The behavior includes giving as well as receiving objects. Based on the classification given in Section 3.4.1, this behavior falls into categories HType 1: I:H, G:H and HType 3: I:H, G:R.

9.1 THE FINAL FLOKA ROBOT

For the final object handing robot, the Meka M1 was again combined with the Floka Head to the coherent Floka humanoid (see Fig. 9.1). I added two Intel RealSense cameras [Int17] for ideal view of interactant's face and hands to combine the perceptive capabilities as described in Chapter 6 and Chapter 8. The camera angles were chosen as shown in Figure 8.2. Resulting again in a mshs-robot, with all required perceptive and actuatoric capabilities.

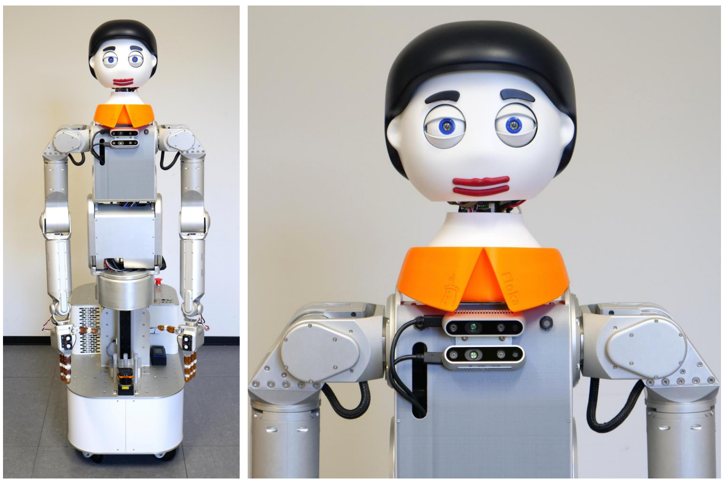


Figure 9.1: Pictures of the final Floka humanoid as seen by an interactant. Two chest-mounted wide-angle depth cameras allow to observe the interactant's face as well as hands and arms in a large area.

9.2 COMBINED BEHAVIOR

The robot's behavior is geared on the established HPhases 1 to 4 following H 1: Handover Has a Distinct Pattern. It is build of 120 software packages that help to produce the required perceptive capabilities as well as the control of the Floka Head and Meka M1 [CK:m-n]. Figure 9.2 gives an overview of the modules and involved data flows in the system. The general stages of the behavior and how the inputs are used in the different phases can be seen in Fig. 9.3. Additionally, the outputs of each module are shown with green arrows. The HPhase 0: Acquire is executed by the previous receive of the robot. I describe the subsequent HPhases in the following chapters.

As recognizing the intent, to exchange an object, is out of scope of this work, I presume that the intent of exchanging an object is given and the interaction starts as soon as a person is around the robot or an operator starts the interaction.

9.2.1 *Enhanced Gaze*

Before I go into details of the individual behaviors according to the HPhases, I explain the embedding of the gaze behavior to address SR 2(c): Gaze for Predictability. Similar to the integration explained in Chapter 6, the targets are selected by states in FlexBE to be able to synchronize the targets with the rest of the handover behavior. To enable smooth gaze, I addressed the two problems of the previous implementation. The first was the low update rate of moving targets and the other was the FSM being blocked while looking at a target with a minimum duration. Therefore, I created an external module (*gaze forwarding with minimum duration* [gaze_relay]) that is responsible for forwarding percepts of other modules to humotion. It receives a target of the type: neutral; face; left/right-hand; $\mathrm{OTP}_{\mathrm{integrated}}$; left/right-EEF. System requirement 4(d): Face Tracking already proved to be important for mutual gaze, such targets are generated with *OpenFace 2.2.0* [OpenFace]. Combined with a minimum gaze duration it allows guaranteeing a minimum gaze duration without blocking the interaction as well as near zero delay forwarding of targets for smooth following of moving objects, e.g. the face or the hand of the interactant or the robot's EEF. Looking at the human hand to signal the exchange intend (SR 4(a): Hand Tracking) is done with the module described in Section 8.1. To increase the predictability, looking at the goal of the robot's EEF is implemented with the previously presented tracking module, which motivates again SR 4(b): Object Transfer Point Prediction.

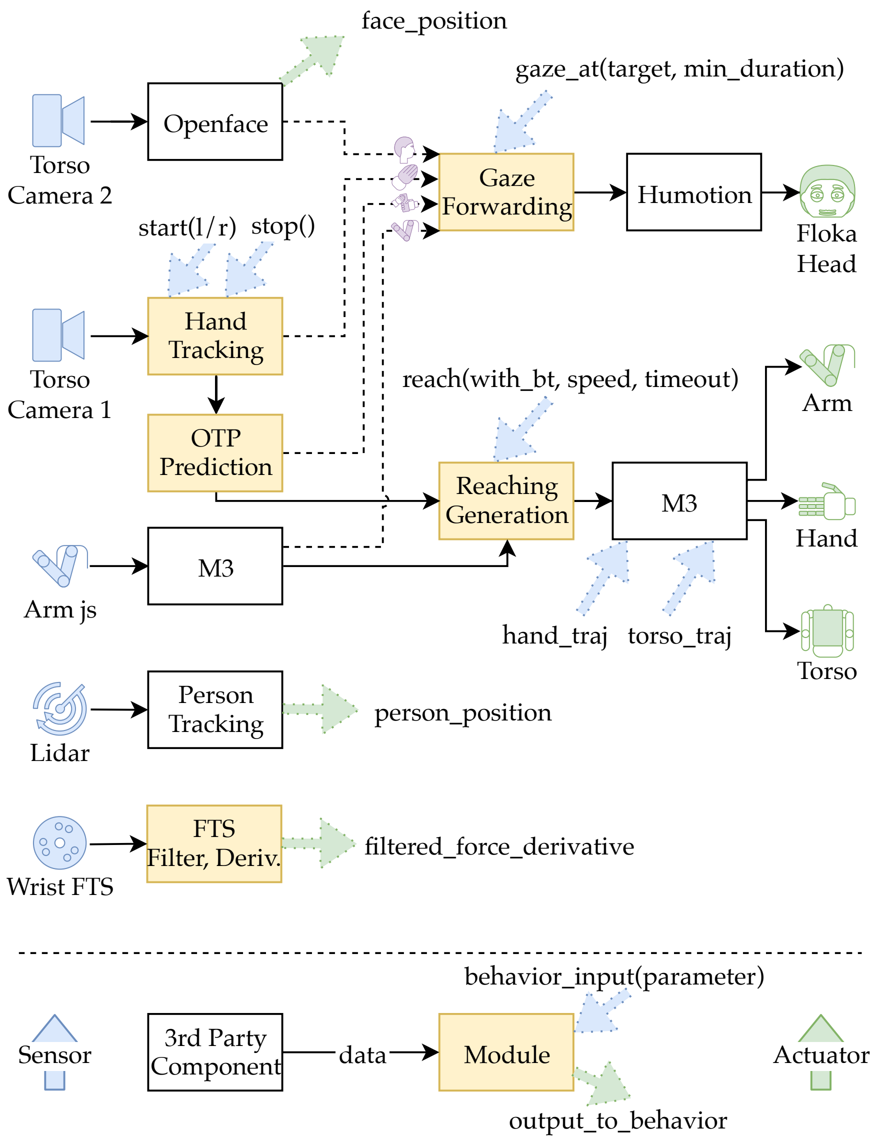


Figure 9.2: An overview of system components and how they interact. Dashed lines mark data flow of gaze targets shown in purple, big dotted arrows visualize interfaces to the behavior layer. Blue symbol mark sensors and green mark actuators. Lines highlight data flow between components. Modules created in the context of this thesis are marked yellow.

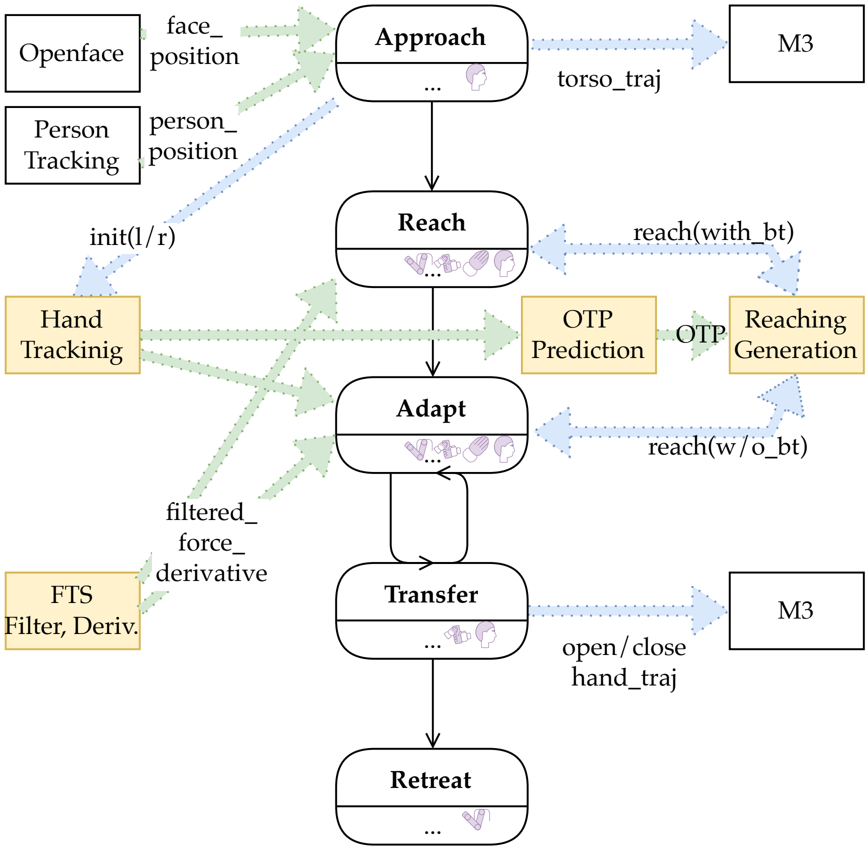


Figure 9.3: A schematic of the full handover behavior and the interaction with in- and outputs. The double boxed nodes are behaviors themselves which include the gaze behavior and additional control schemes, that are explained in the following chapter. The gaze targets in each phase are shown in purple. Dashed lines represent data-flow and solid ones show state transitions.

9.2.2 *Approaching - Orient Towards Interactant*

In this phase the behavior of the robot pursues two goals. The first is attracting the interactant with an open posture and an addressing gaze. Concerning SR 2(a): Body Orientation, Floka adjusts itself by turning it's gaze and upper body towards the interaction partner. This establishes and maintains a vis-à-vis configuration for the interaction. This directly helps with the second goal of getting a good view of the interactant. With this behavior the robot is enabled to initialize the hand tracking on the interactant. Figure 9.4 shows the actions in this phase. Incorporating multi-modal perception with the *STRANDS people perception pipeline* [PPerception], fulfilling SR 4(c): Person Tracking, based on Floka's LiDAR, the panning can start early and does not require the interactant to be in the FoV of the camera. As the interactant comes into the view of the camera, it can also adjust the tilt based on the person's height, which is optained with OpenFace for SR 4(d): Face Tracking. During the approach, the gaze target is the face of the interactant. This behavior is only active after reaching the Space 3: Social ($\leqslant$3 m) until Space 2: Personal ($\leqslant$1.2 m) when reaching the interaction space. In this phase Floka can signal readiness to interactants by establishing joint attention. For a longer approach distance or attraction of interactants from Space 4: Public more complex behaviors could be integrated like the addition of random gazes (cf. [Hol14]. When human and robot get close, initialization of the tracking is triggered and a small motion with the arm is executed to signal readiness to handover.

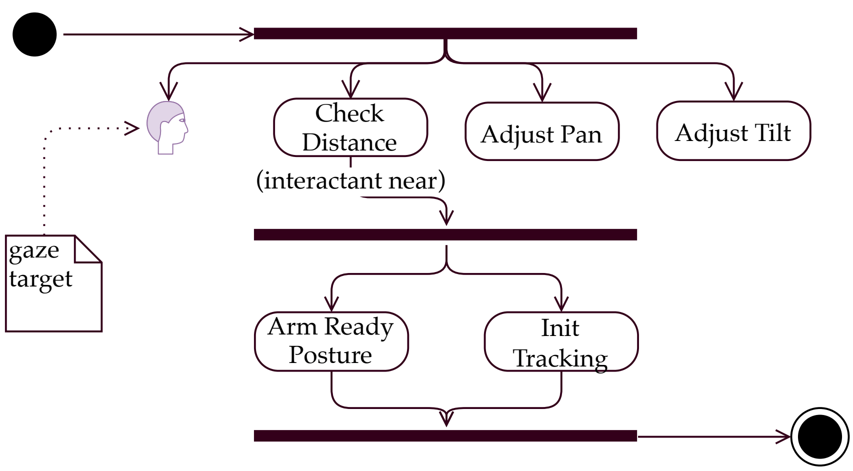


Figure 9.4: Designed behavior for the HPhase 1: Approach as Unified Modeling Language (UML) Activity Diagram. The robot orients itself towards the participant, based on the persons position. The face position is used to estimate the height and thus effects the tilt of the torso.

9.2.3 *Reaching Behavior*

After the approach, the robot waits for the human starting to reach out to react accordingly to the human's motion. However, for a proactive behavior only a minor change would be required by providing an $\text{OTP}_{\text{static}}$, that lets the robot reach before a human motion is detected. In this phase, mainly the reaching module acts on the data by the OTP prediction pipeline to create a gesture that conforms to SR 2(b): Gesture Motion. The FSM observes the force (see Section 9.2.3.2) and controls the gaze pattern. The HPhase 2: Reach is split into the two proposed sub-phases.

Figure 9.5 shows the gaze pattern and transitions during reaching behavior in the HPhase 2(a): Base Reach. When the robot is the giver it looks from its EEF to the human's hand and into the face. This aims to create joint attention on transferring the object from the robot to the human. For the receiving case, the gaze pattern is reversed. As soon as an OTP is available, which means the prediction detected initiation of handover, the robot looks at this point to inform the interactant. Besides the signaling with gaze, I added a grabbing gesture with the EEF for the receiving case, as suggested by Lee et al. [Lee+11] to signal readiness to take an object. It is synchronized to the arm motion which is controlled by the adaption module.

The second part of the Reach phase (HPhase 2(b)) implements the *mode2* behavior by employing execution of ATs. Figure 9.6 shows this part of the reaching behavior. Here the finer adjustments of the EEF position take place until being close to the human hand. Also, the visual handover can take place as both are already close and in good view of the camera. The force check can be more sensitive as the robot's arm motions have less velocity and acceleration, which lowers the impact on the FTS. It also serves as a fallback when the transfer was not successful by adapting again to the human hand without executing a BT. This phase incorporates the same gaze behavior as during the first part of HPhase 2(a): Base Reach. If the phase gets prolonged due to unexpected behavior, the robot repeats the gaze behavior until transitioning to the next phase.

9.2.3.1 *Prediction and Adaption Interaction*

After triggering the initialization of the OTP prediction module, the adaption modules waits for an $\text{OTP}_{\text{integrated}}$ to be published. As soon as a reaching motion is detected, the prediction module outputs an $\text{OTP}_{\text{integrated}}$. This way, there is no delay between handover detection and starting the robot's reaching motion (SR 1(c): Reactive Pattern). The adaption module selects a suitable BT from the database based on the first $\text{OTP}_{\text{static}}$ and smoothly adapts to the latest prediction until it comes close to the $\text{OTP}_{\text{integrated}}$, which triggers the transition to the second phase (mode 2 reaching) where the adaption module is called

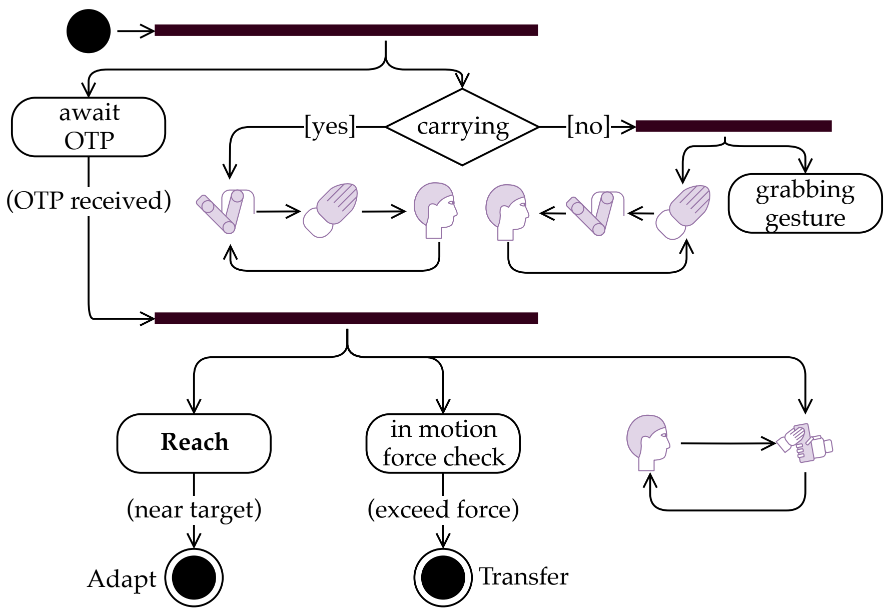


Figure 9.5: The handover reaching behavior (HPhase 2(a): Base Reach) as UML Activity Diagram with the according gaze pattern. Here the robot waits for the OTP of the prediction module and incorporates gaze to create shared attention on the task. As soon as the motion starts it regularly looks at the predicted OTP for a good predictability. Force derivative is constantly monitored to allow in motion handover

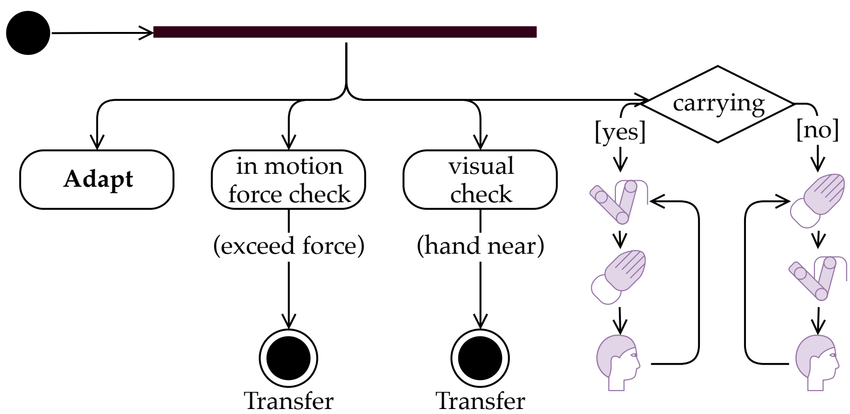


Figure 9.6: The behavior during the Adapt phase (HPhase 2(b)) as UML Activity Diagram. In general this phase is comparably short as it only bridges a small distance. Here, visual handover is possible. If this phase takes longer, the robot repeatedly signals the goal of exchanging the object with appropriate gaze pattern.

without executing a BT, with lower velocities, and a closer target distance. This way it converges to the human hand carefully. When a contact is detected as explained in the following section, it directly transitions to the HPhase 3: Transfer.

9.2.3.2 *Improved Transfer Detection*

As discussed in Section 3.4.7, using an FTS in a robot's wrist is a common approach for contact detection (SR 4(e)). Nevertheless, it was found that not all people actually apply force when handing over. Hence, I complement the contact detection with visual handover (SR 4(f)). I also present an improved contact detection allowing in motion handover fulfilling SR 3(b): Shortcuts for Experts.

To allow visual handover, I compare the tracked hand position with the robot's EEF location. First, an offset is applied as the human hand is expected to approach from the EEF's opening side. Then, the distance d is calculated and filtered for a duration t, to prevent noise from triggering the transfer. If the distance falls below a threshold for a certain duration, the transfer can safely be triggered to fulfill SR 4(f): Visual Transfer. This is supported due to the added self-filtering of robot and object as explained in Section 8.3.2.3.

Nevertheless, the visual transfer detection also has downsides, like being only applicable in the sensors FoV, failing if the human hand approaches from a position where the robot's EEF, or the object obscures the view on it. Additionally, the filtering introduces delay and a distance threshold gives no guarantee that the human is really in contact. Thus, SR 4(e): Contact Detection is still required and should be combined with vision. The biggest challenge is to differentiate between forces applied by the robot itself and those introduced by the human. Depending on the position of measuring, like in the elbow, the wrist, or even on the surface of the EEF with tactile sensors, changes the amount of each type of forces. While a sensor in the humanoid's "elbow" is exposed to all forces before and after that joint in the kinematic chain, a surface sensor on the EEF only measures forces between an object and the EEF. Estimating the dynamic load introduced by all the actuators, gravity, and an object is an open challenge and would require an exact model of the robot, including friction, link inertias, and other hard to estimate factors. Thus, some approaches stop the robot to rule out it's influence on the measurements, then they bias the sensor and apply a force threshold. As discussed in Chapter 5, I could observe that even stopping the robot after the reaching motion, it is not free of introducing forces or torques on the FTS due to post-pulse oscillations and the SEA compensating gravity generate a permanent change in signal. Thus, one can not simply acquire forces applied by the human interactant. Delays and filters prevent false positives that would cause unwanted drops of the object but prolongate the handover with a negative impact on the interaction performance. For in motion handover this approach is not applicable at all.

I designed an additional module that aims to address the discussed issues by applying a suitable filter chain, incorporating the robot's motions, and a rate of change based decision process. The input to that process is the data of the wrist mounted FTS, which is sampled at 1 kHz, filtered with a third order Butterworth filter [But+30] with a cutoff frequency of 20 Hz. This data is forwarded and processed at 100 Hz. Rate of change estimation and smoothing is performed in a single step by applying a Savitzky and Golay filter [SG64], which is similar to differentiating with respect to time for streamed data. A filter length with a history of thirteen is applied, having the center of the filter delayed to an average of 50 ms [Der17]. This way the filter delay is kept to an acceptable amount while removing noise introduced by the actuators, vibration, and other factors. The norm of the x and y component of the force data are added to have a one dimensional data stream. It is damped by dividing it with the total requested arm motions smoothed with a sliding average. This prevents false positives on robot induced forces, nevertheless requires a higher interaction force applied by the human for in motion handover. Another optimization is to increase the threshold by a factor of 1.5 when the robot carries an object. As the object adds inertia to the EEF, motion forces on the sensor increase and need to be addressed. On the other hand, in the giving case, Floka is responsible of the object and a false positive would cause a drop. For the receiving case, a false-positive would only trigger a re-grasp.

9.2.4 *Object Transfer*

When the contact is detected, this phase is about grasping, respectively releasing of the object, depending on the robot being giver or receiver. In the giver scenario Floka opens the hand such the human can take it. When the robot is tasked to take an object, the SEA that actuate the fingers are set to a higher compliance for closing the hand, to establish a power grasp. The reduced stiffness allows for a soft grasp, being safe for the interactant. The resistance is checked to verify that an object is in the EEF. If this is not the case, it transitions back to the adaption phase, making sure, human and robot hand are close together for the next try. After a successful transfer, the retreat phase becomes active. Gaze is incorporated for signaling focusing on the transfer by looking at the own hand.

9.2.5 *Retreat*

The primary task of the HPhase 4: Retreat is moving back to a ready state. Executing a motion from a database is not possible as the start state of the motion is not known, because the AT moved the arm to a new configuration. Thus, MoveIt is tasked to plan a collision free motion from the transfer position back to the ready state. During that, hand-tracking is stopped and the robot's gaze follows the EEF until it moves back to a neutral gaze for signaling readiness for the next interaction.

9.3 PARAMETERS

A designed behavior, as presented here, always includes choosing appropriate parameters that help to shape the robot's actions. These parameters also concern the SR 1(b): Pattern Scalability. They should be selected to make the interaction safe, fast, and predictable. While some of these parameters were already discussed with the modules containing them, I address the most important parameters in the context of the whole integrated behavior in order of their occurrence. Starting with the Approach, the speed and acceleration of the torso adjustment can be modified. While it shapes the first impression the interactant gets of the robot, sane parameters are easy to find. Transitioning out of this phase is done with a distance threshold parameter, which impacts not only the behavior change but also the initialization of the tracking pipeline. As this pipeline only requires a good view of the person, we need to ensure that the person is not too close when the initialization takes place. Distances found in human-human handover literature can be employed as a starting point. For a gesture-like motion in the Reach phase, the BT in the database play an important role and can be seen as parameters themselves as they define velocity and motion profile. They are also specific to the robot. During the BT, the DLS parameters, like alpha and damping as well as the velocity clamping impact the motion but mostly affect the speed. However, if chosen too high overshooting and thus oscillations might occur which needs to be prevented. For the transition to the HPhase 3, parameters of visual and force handover need to be chosen. Regarding the visual case, the distance and a duration need to be set for a stable recognition of closeness of hand/object and EEF. The force handover mainly needs adjustment of a threshold and the filter setup. Here the choice is about balancing between false positives and negatives, by having either a faster recognition or being safer by preventing dropping of objects. In terms of gaze parameters mostly the duration is implemented as parameter. Giving the option of sending more information with the robot's gaze while making sure not looking nervous and thus more natural.

9.4 DISCUSSION

In this chapter, I presented a holistic approach to human-robot object handover. I transferred the proposed and identified phases of human-human handover to a humanoid robot. As gaze is not a static process where a single target is set and kept for a duration a new module keeps track of such targets. This module forwards them to a humanoid robot head that is able to communicate goals and intentions. A module that preprocesses FTS data for easier contact detection helps reduction of false-positives and even allows for in motion handover. This way, experts can take a shortcut by exchanging the object while the robot is still moving its arm (SR 3(b)). In general, I presented a concept that integrates the previously introduced modules and behaviors to interact well together. This enables a humanoid to naturally exchange objects with humans without external support in terms of sensing or processing.

10 FINAL EVALUATION

To evaluate the previously introduced and discussed concepts and their implementation, I designed and conducted a user study. The design builds on my first study (Chapter 5) and was modified in consideration of the introduced modules. The study targets to test the fully integrated and autonomous system with naive users and different tasks. Ethical approval for this study was obtained from Ethics Commission of Bielefeld University (2019-217) with the guidelines of the Deutsche Gesellschaft für Psychologie (DGPs; German Psychological Society). Written informed consent was obtained from all subjects before the study. No monetary or other compensations were given to the participants.

10.1 INTERACTION DESIGN AND PROCEDURE

I crafted ten different tasks that test the new capabilities and recovery strategies. To accommodate for the proactivity of the newly designed behavior I changed the spatial layout by increasing the distance of robot and object location. This experiment consists of a questionnaire, handover task (HTask), and an interview to find out how people perceive the handover interaction with the improved Floka. The pre-interaction questionnaire (see Appendix D.3) collects personal data, including the experience with robots and other technical systems as well as the pet ownership as inspired by Mutlu [Mut09]. The NARS items aim to identify interactants who have a negative bias on robots. Though the interactants were asked to perform the tasks with the right hand, the handedness is collected with the hand-items of Ehrenstein and Arnold-Schulz-Gahmen to determine whether they have to use their non-primary hand [laterality]. The instructions (Appendix D.4.2) for the interaction consist of reading a task, attending to it, and repeating the procedure until all tasks are done. The FlexBE behavior (Appendix D.2) was extended with states to control the external recording of the experiment. Similar to the first study, a trigger of the e-stop is used by the experimenter to start the next run when the participant is reading a task, so that the robot is ready when they start approaching. A post-interaction interview (Appendix D.5.2) assesses how the interactants perceived Floka's behavior in terms of gaze, motions, and overall handover performance.

10.1.1 *Study Setup*

This evaluation was set in the same room of the CSRA as the first study. The original setup was modified mainly in terms of the object positions by moving the table from left of the robot to the opposite wall. Figure 10.1 shows the spatial rearrangement, to obtain an approach of ≈2.5 m. As the HTasks was apparently about handover, there was no point in having a distractor task. Thus, only one object, which can be seen in Fig. 5.4(b), was used. It weighs about 200 g and was placed on the table before the experiment started.

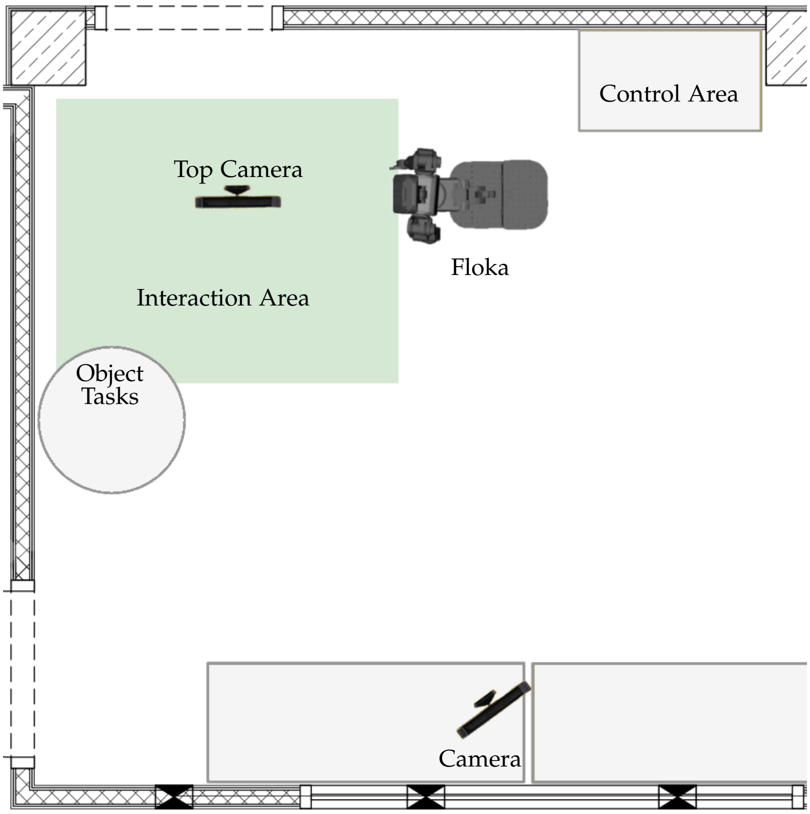
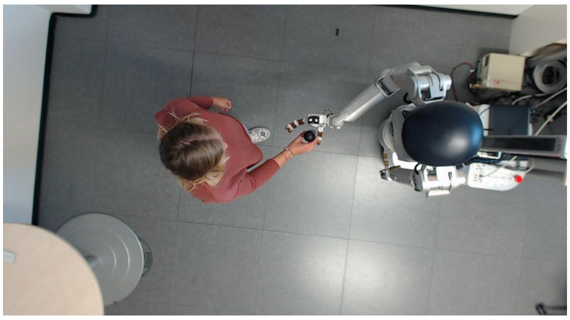


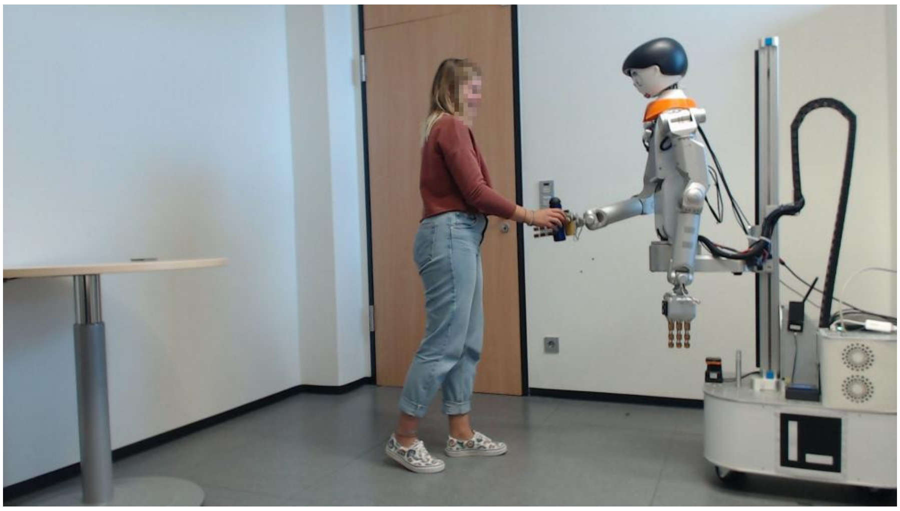
Figure 10.1: The setup of the experiment as a schematic top view. On the left is a small table with three objects. The participants interact with the robot Floka from the interaction area (light green). An external camera, as shown on the bottom of the schematic, is placed for a side view of the experiment.

Two external cameras were added to the setup for to record the experiment to allow a later analysis. One of them was mounted above the interaction area and the other one was in a similar position as in the original experiment. It was tilted to also have a view of the object picking area. Figures 10.2(b) and 10.2(c) show a view of those cameras during the experiment. The view of Floka's eye camera records where the robot is looking during the experiment (see Fig. 10.2(a)). As the whole system was build on ROS, all data could be recorded with the corresponding tools. To interfere as little as possible with the system

(a) Floka's eye camera looking at its hand



(b) View from the top-mounted camera aligned to the interaction area



(c) Side view on the handover from an external camera

Figure 10.2: The experiment from view of the recorded cameras. Floka is receiving an object from one of the participants.

running on Floka, the recording of data streams was outsourced to an external PC, connected via Ethernet to the robot. Both cameras were calibrated to the robot's tf with help of BART to allow mapping system data in the video for later annotation and analysis. As the hand tracking was running as a nodelet (see Section 8.3.2 on page 99) in the same process as the sensor providing data of the hand region, this sensor data could not be recorded without having interfered with the performance. Appendix D.1 lists all recorded topics.

10.1.2 *Handover Interaction Tasks*

I designed specific handover tasks that aim to evaluate different behaviors addressed in this thesis. Therefor, I created tasks that each address human behavior known from previous human-robot handover experiments. Starting with a random give and take where no restrictions were made (HTasks 1 and 2), a natural handover can be observed and experienced by the interactants. Where possible, the giving and taking scenario are designed with similar tasks to equally test both cases. This way, the HPhase 0: Acquire is executed by the previous handover. The HTasks were written on cards with 0.09 m × 0.05 m size and placed on the table next to the object (Appendix D.4). They are placed upside down, so that the participant can only see one task. The order was kept constant over all handovers.

Handover Task 1: Random Give
Instruction: Give Floka the object.

Handover Task 2: Random Take
Instruction: Take the object from Floka.

Handover tasks 3 and 4 aim to address the finding, that not all interactants tend to actually apply force during the interaction. For SR 4(f): Visual Transfer I added a visual handover that is tested with the associated tasks which ask the person to not apply force to the robot.

Handover Task 3: Visual Give
Instruction: Give the object by holding it out to the robot.

Handover Task 4: Visual Take
Instruction: Take the object from Floka without pulling on it.

In contrast to the previous tasks, HTasks 5 and 6 are about creating contact with the object or robot. Thus, for SR 4(e): Contact Detection it can be tested how the contact is actually established.

Handover Task 5: Pushing Give
Instruction: Give the object by pushing it into the hand.

Handover Task 6: Pulling Take
Instruction: Take the object by pulling it out of the hand.

Addressing SR 3(b): Shortcuts for Experts and the, to some degree, linked SR 1(b): Pattern Scalability, HTasks 7 and 8 target the performance of the system when interacting with fast people. The robot's capability of in motion handover aims to be tested by the users. At least if the interactant does not wait until the robot stops because they assume that this is the earliest time of handover.

Handover Task 7: Early Give
Instruction: Give the object as early as possible.

Handover Task 8: Early Take
Instruction: Take the object as early as possible.

Handover task 9 is the first non-symmetric handover task, in terms of giving and receiving. It aims at testing the recovery functionality that regrasps the object as introduced in Section 9.2.4. As stated, this recovery function is only possible when the robot is receiving an object. For the last interaction (HTask 10), I chose to give the same task as HTask 2 to observe whether the previous tasks might have changed the participant's behavior as they can freely decide again how to take the object from Floka.

Handover Task 9: Regrasp Give
Instruction: Give Floka the object but pull it away as the robot closes the hand.
Afterwards give it at a different position.

Handover Task 10: Final Random Take
Instruction: Take the object from Floka.

10.2 APPRAISAL

Together with my research assistant, I recruited 16 participants (F=9, M=7) from the campus (Bielefeld University and University of Applied Sciences), aged 22.75 ± 2.86. Three participants were not able to finish the interaction due to technical reasons. Thus, they did not take part in the post-interaction interview. This resulted in thirteen fully evaluable participants, aged 22.23 ± 1.88 (F=7, M=6). Figure 10.3 shows the composition for all participants, and the sub-group that took part in post-interaction interview.

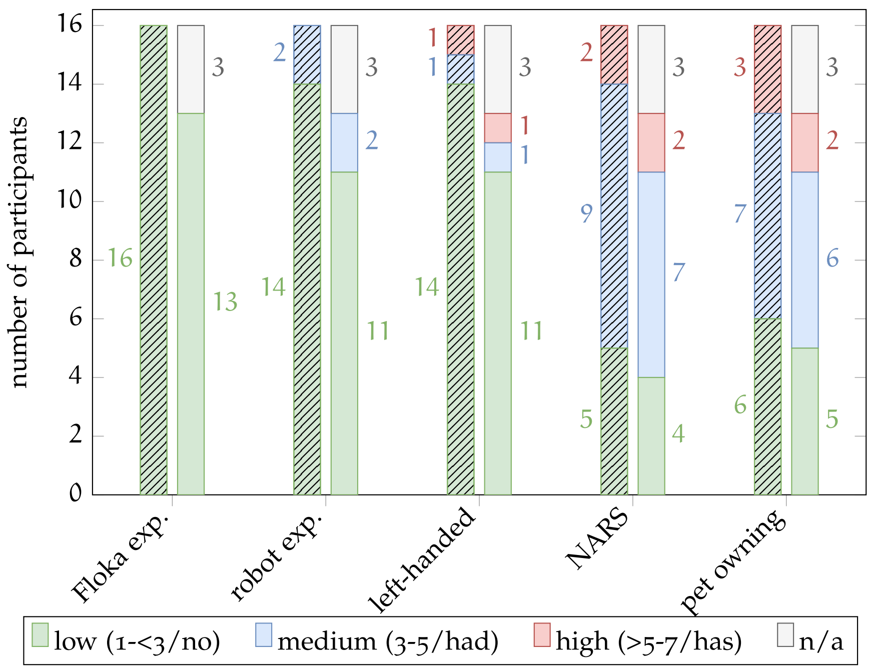


Figure 10.3: The structure of the participants taking part in the evaluation. Bars on the left represent all participants (lines pattern), the right bars represent only the ones that also took part in the post-interaction interview as three runs had to be canceled.

Only one participant had a little experience with Floka (2 on a scale of 1 to 7). Everyone else had no experience at all with the used robot. The average experience with robots was 1.50 ± 1.03. The subgroup had an experience with robots of 1.54 ± 1.13. Hence, it can be stated that the final evaluation was done with naive users in regard of experience with HRI. Only one participant stated in the questionnaire that the primary hand is the left hand as it was stated to use it for drawing and throwing.

I recorded a total of 138 object exchanges. In only one of them the object was dropped. The participant picked up the object and continued the interaction. In the two stuck and one missed case the interaction ended after its occurrence, resulting in 21 lost runs. In the other 66 gives and 69 receives the interaction was successful without interruptions. The system ran for about 8 hours without a restart, which shows the achieved robustness. Figure 10.4 visualizes the outcomes of the runs.

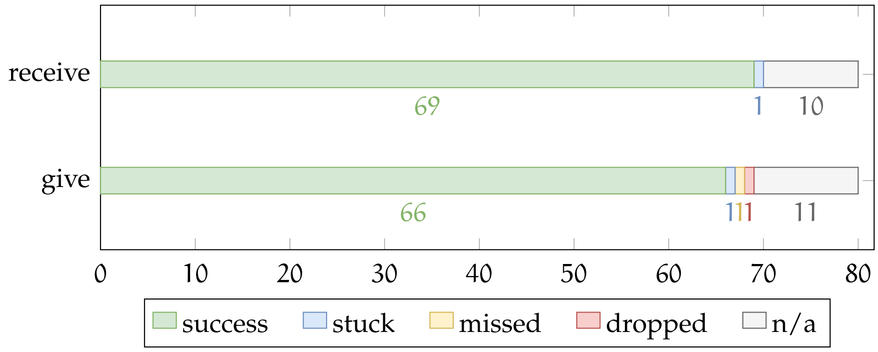
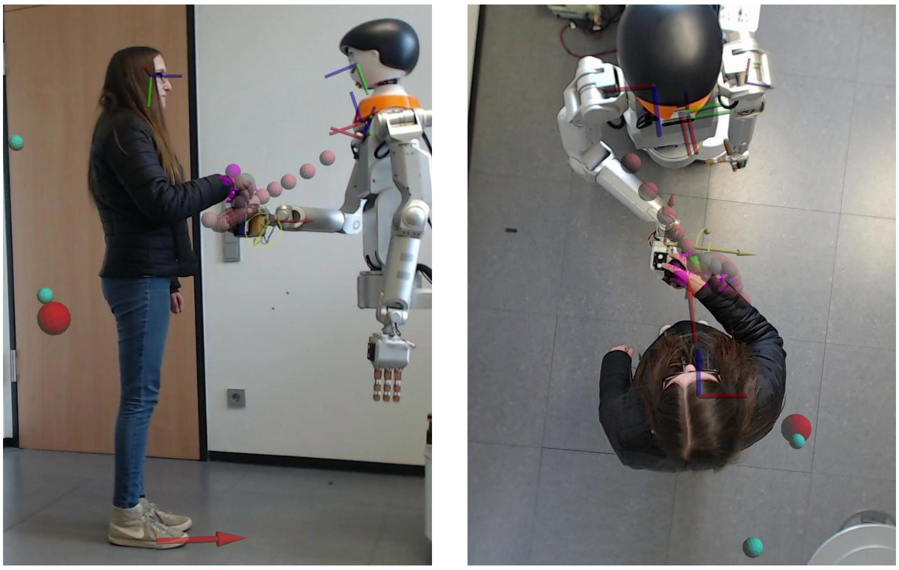


Figure 10.4: The results of the 16 participants with 5 give and 5 receive tasks each. One handover where the object was dropped but the experiment could continue. Two interactions got the robot stuck and one where the transfer was not recognized properly resulted in stopping the experiment.

For two participants the experiment had to be stopped as MoveIt did not find a solution for the retreat of the arm of Floka after some runs. Although the reaching module described in Chapter 7 queries MoveIt before execution of each command whether the goal is collision free, noise or imperfections in control might lead to being "virtually" in collision in the next step. To free the robot from that immobile state, the robot was stopped by the experimenter and it was proceeded with the next participant. While a future implementation should take care of the retreat in all situations, it was not the focus of this thesis. In another run the experimenter started the experiment a little delayed which resulted in a confused participant that in the second task removed the object by twisting the object out of the hand, which forced the experimenter to stop the experiment to prevent damaging the robot and guarantee safety of the participant. As safety and proper experiment execution is always the highest priority in HRI, stopping the experiment was comprehensible.

Figure 10.5 shows tracking and prediction results during a typical handover in the study. It visualizes most of the recorded tracking and prediction data. The recorded gaze directions also allow an analysis of such cues. It shows how the first predictions of the $\mathrm{OTP}_{\mathrm{dynamic}}$ overshoot but the technique described in Chapter 8 generates stable results. In the pictured situation Floka gives the object to a participant. EEF and hand meet at the midpoint between human and robot.



(a) sideview

(b) birds-eye view

Figure 10.5: A ROS-rviz visualization of the tracking and prediction results added to the view of external cameras. The big red sphere is where the hand tracking was initialized, the pink sphere marks the predicted $\mathrm{OTP}_{\mathrm{static}}$. The smaller reddish spheres visualize the updates of the $\mathrm{OTP}_{\mathrm{dynamic}}$ predictions. The red arrow shows the detected person position. The coordinate system represent *Transform Library* [tf]-frames of Floka's cameras, as well as the interactant's and robot's gaze direction. At the wrist, the measured forces and torques of the corresponding sensor is displayed with a yellow and a red arrow. A bigger red arrow marks the foot position of the person.

I analyzed whether the transition to HPhase 3: Transfer occurred due to the visual or force based trigger. Therefor I obtained the trigger from the recorded robot data. Figure 10.6 shows the results grouped by the HTasks. Overall, a little over half (0.51) of the exchanges were triggered visually. This highlights again the equal importance of both modalities. I therefore advise not to neglect SR 4(f): Visual Transfer in favor of SR 4(e): Contact Detection for behavior of robots that is understandable by everyone (SR 3(a)). The handover tasks Visual Give and Visual Take aim to exchange the object without applying force. It showed that in those tasks most of the time visual was actually the

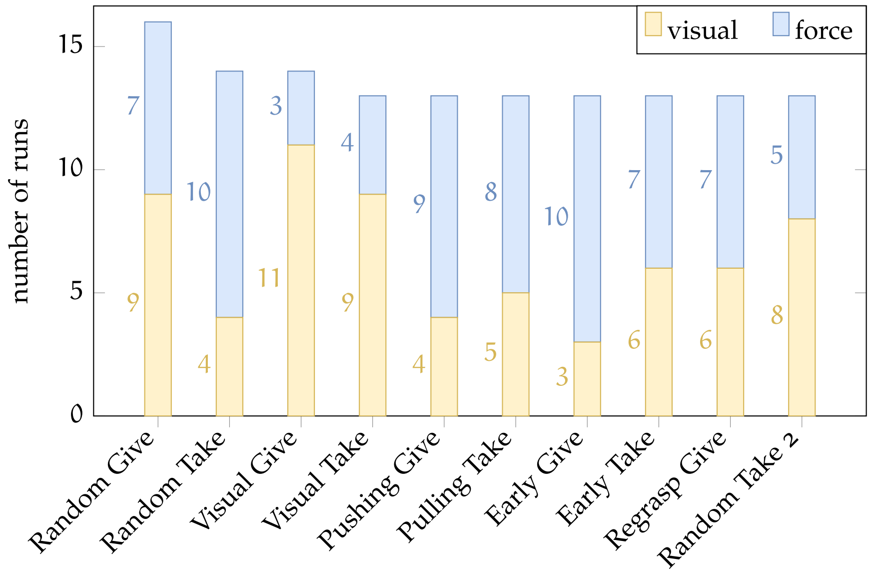


Figure 10.6: The number of runs per task that triggered the transfer by visual, as well as force sensing. Data was obtained by analysis of the recorded *FlexBE - The flexible behavior engine* [FlexBE] log.

trigger with 20 out of 27 runs. The exceptions might be caused by not understanding the task or by falling back to the behavior used before. In contrast, the subsequent Pushing Give and Pulling Take aim to bring the user to apply force. For this task, 9 out of 26 runs did not follow these instructions. Possible reasons include the task not being strongly worded enough, or the participants sticking to their natural behavior or do not dare to apply even minor forces against Floka. The early give task was the task with the most force-based triggers. Reaching into the robot's motion seems to be most intuitive to applying force. While pulling on the object while the robot moves it towards the interactant is not as easy. For the equal random take tasks (HTasks 2 and 10), there was an increase of visual handover from four to eight.

Besides the analysis of force and visual trigger, I extracted the duration of each phase in the behavior. Figure 10.7 plots the duration for the three phases HPhase 2: Reach, HPhase 3: Transfer, and HPhase 4: Retreat. The measurement of the HPhase 2: Reach started when the arm began to move and ended when it was close to the human hand. The sub-phases HPhases 2(a) and 2(b) are combined, as the transition is smooth and the duration of the BT is constant for all trajectories in the database with a duration of 1.5 s. When the distance between human hand and robot's EEF is less than 4.5 cm the measurement for HPhase 3: Transfer begins. It ends when Floka fully closed respectively opened its hand, transferring the object. At last, HPhase 4: Retreat took over and ended when the robot was back at the ready position. In the plot the times are shown as median as this allows visu-

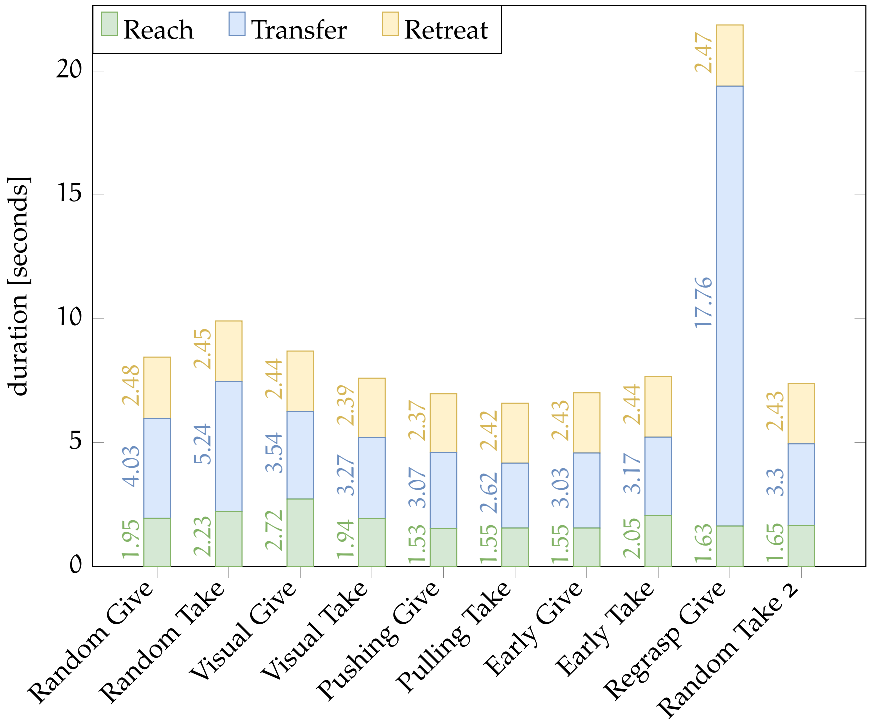


Figure 10.7: The duration of the successful handovers for the different tasks in three phases. All times are the median duration.

alizing the prototypical duration of each phase for the stated HTasks. While the overall Transfer median is at 3.59 s the HTask 9: Regrasp Give is at 17.76 s, as it contains at least two grasps and an additional reaching until the object is finally handed to Floka. This indicates that most participant followed the instruction of relocating the object after the robot tried to grasp it.

Figure 10.8 shows a histogram of duration of the three phases over all interactions. For another view on the duration for each task as a boxplot refer to Fig. E.1. Most of the Reach motions took between 1 and 4 seconds with 114 of 135. The fastest reach took only 0.46 s in the HTask 7: Early Give. For the transfer there is a lower bound of 2.4 s, as this is the duration the robot's EEF needs to fully open and close. In the giving case it can be reduced to 2.21 s due to the object stopping the fingers before fully closing the hand. Thus, most of the Transfer took about that amount of time. By analysis of videos, I found out that some runs took longer due to visibility reasons in the visual condition or are based on the regrasp in HTask 9: Regrasp Give. The longest transfer happened in a HTask 4: Visual Take were the view of participant's hand and object was covered by the robot's own EEF and the interactant waited patiently to correctly fulfill the task. All Retreat took 1 s to 4 s, except one retreat that was not completely recorded and thus could not be evaluated. The fastest retreat was in

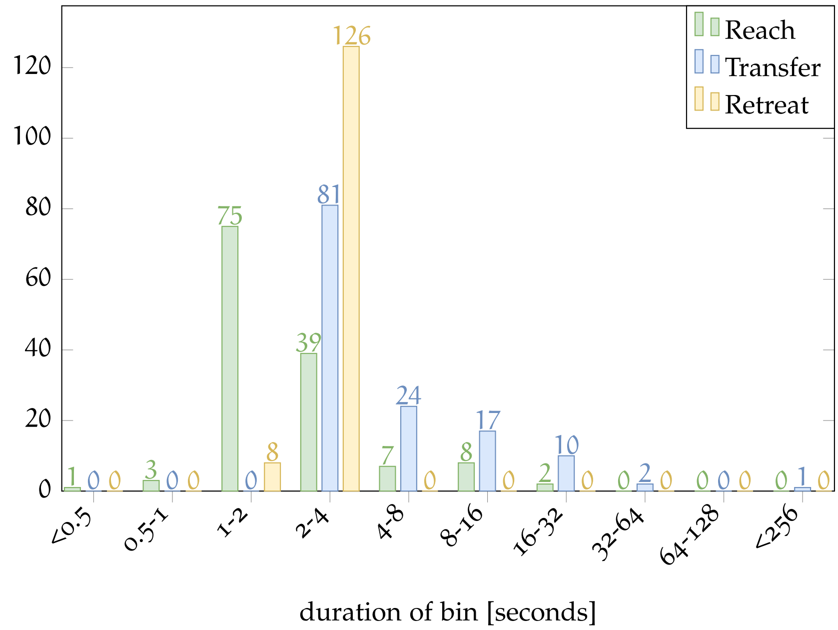


Figure 10.8: Visualization of the frequency of the duration in the three phases over all 135 successful handovers. They are defined as described in Fig. 10.7.

HTask 8: Early Take as here the travel distance was shorter and thus took 1.33 s.

Figure 10.9 shows the results compiled from the answers in the post-interaction interview. In the interview eleven participants used words that directly or indirectly describe Floka's behavior as human-like, which hints to a fulfillment of SR 1(a): Human-Like Pattern. There was only one participant that described the behavior as not human-like at all. This exception came from the participant with a high NARS rating and thus might be referable to a negative bias. Also, eleven participants described the interaction as slower or too slow but still nine of them would work with the robot. Here again the participant with a high NARS score was one of the exceptions and said the robot was fast enough. Ten participants rated the handover as generally safe. Three minor reservations in regard of safety apply for the case where the object was dropped, the person with a high NARS score, and someone that rated an occurred deferral as slightly unsafe. Half of the participants stated having recognized the difference in the give/receive force threshold. Only two participants were not aware of the gaze behavior Floka exhibited during the interaction. Everyone else could, at least roughly, describe the gaze pattern integrated in the behavior. Answers describing the gaze targets included: "The robot's hands, the object, in my face.", "first on the object, then into my face, more often on the object", and "Its gaze switched between our hands.". Some participants even stated that the gaze helped to understand what the robot wanted. All participants stated that they

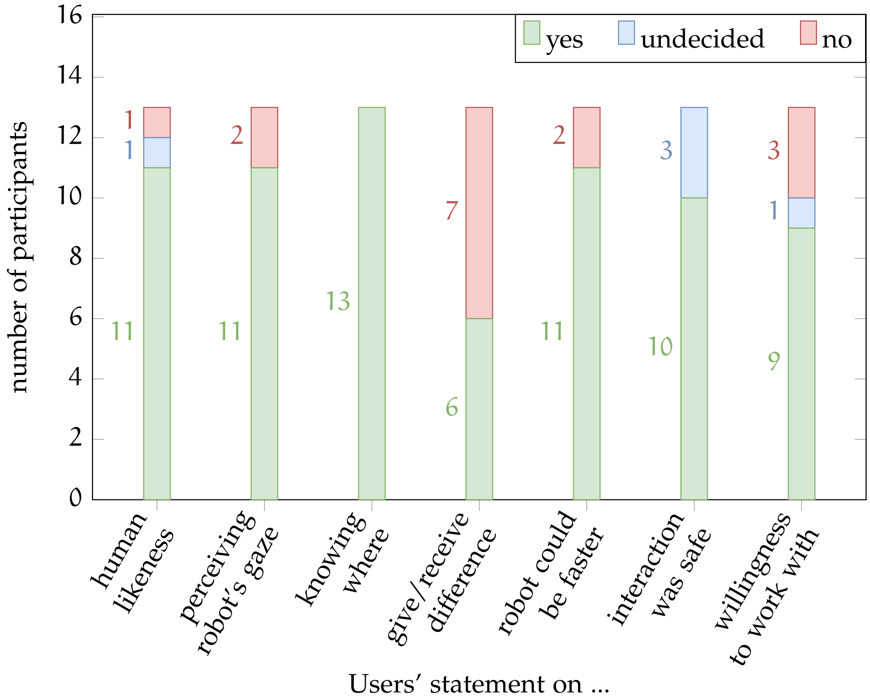


Figure 10.9: The results of the handover interaction with Floka compiled by the answers given in the post-interaction interview.

either consciously or unconsciously knew where to move their hand to exchange the object.

10.3 DISCUSSION

The presented results show that most of the participants where able to successfully exchange objects with Floka without an explanation (SR 3(a)). Even the different tasks that involved types of handover that caused problems before, like pure visual handover, in motion handover, and different positions, might have introduced delay in some cases but did not break the interaction.

I analyzed the situation in which Floka dropped the object while giving it to a participant and found it being caused due to a fault in the hand tracking (see Section 8.1). The object was not properly filtered from the point cloud and hence recognized as part of the human hand. This triggered a visual handover which released the object a little too early to be grasped by the participant. Nevertheless, it was a single occurrence which can be addressed with a sensor that has less noise in the data or an adjustment of filter parameters. The opposite occurred when the human hand was filtered because of being too close to the robot's EEF. Also, the hand was not recognized properly when being hidden behind the robot's hand or the object from being visible in Camera 1. In both of these situations the pure visual handover did not trigger and caused a delayed interaction. In this situation, also a camera with less noise or a second one from a different perspective could help to overcome these issues.

Another cause of slowdowns was occasionally introduced by too much filtering of the FTS data in the recognition with the improved transfer-detection (see Section 9.2.3.2). This module allows in motion handover and safe overall handover. However, by smoothing and damping measurements during motion as well as applying a more conservative threshold when the robot has the object, it was sometimes not sensitive enough to sense the small forces applied by some participants. Depending on the scenario and the object, a more progressive threshold could be applied for a smoother transfer. Additionally, an adaptive approach could modify such a parameter over multiple handovers. A tactile hand-surface could also improve the general transfer as it would not measure the forces between the robot's hand and arm but directly between EEF and the object. Still, it remains a challenge to precisely distinguish internal forces applied by the robot itself and external forces applied by the interactant.

Overall, the duration of the HPhase 2: Reach of Floka with a median over all runs of 1.8 s is a little over the typical human reaching duration of 1.24 ± 0.28 s (cf. Section 3.4.6). Due to some outliers, the average duration was 2.95 ± 3.21 s. This might have been caused by the more complex tasks and the aforementioned visibility complications. While the duration of the HPhase 3: Transfer is not as clearly stated in the literature, Mason and Mackenzie report a duration of 0.275 ± 0.009 s for the time from first contact to peak grip force in their human-human handover experiment [MM05]. In this experiment, Floka achieved a median over all tasks of 3.59 s. Considering the limitation of its hand which already takes $\approx$3 s to fully open and close when the stiffness is reduced, the reaction time is not too far away from the human's. As the interactants experienced the full duration, it is comprehensible that they mostly stated in the interview that the handover speed could be improved.

Comparing the current results to the transfer times of my first experiment which averaged at 1.77 ± 2.74 s combined with the $\approx$3 s EEF closing time, the average result is just below the original number with 4.51 ± 3.11 s for similar force based tasks like HTask 5: Pushing Give and HTask 6: Pulling Take. It is important to note, that this only considers a single phase of the whole process and that the initial implementation forced users to interact in a single way. Also, the first study contained experts and semi-experts whose results are part of the average result. The new design gives the choice on how and where to hand the object considering different levels of experience with robots. The importance of SR 4(f): Visual Transfer could be shown and an especially low average transfer duration of 3.64 ± 0.53 s could be achieved in HTask 3: Visual Give, while in the first experiment the interactants were required to introduce force on the sensor by shaking the object, which is also the case in many systems found in literature (see Section 3.4.7). Some researchers approach this by explaining it to the participants or by stopping the experiment if the participants do not get the right solution by themselves. The multimodal detection approach proposed by me proved to be successful relating to SR 3(a): Understandable by Everyone. Regarding SR 3(b): Shortcuts for Experts, interactants can now make use of in motion handover to skip the HPhase 2: Reach for an overall faster interaction. Here a transition to the Transfer could be achieved after only 0.46 s which is 2.49 s faster than the average reaching duration in this experiment and 4.43 s faster compared to the first experiment.

All participants stated to know where to hand the object. This indicates that gaze and gesture transmitted the *where*, leading to the assumption that the deictic gesture character of the motion stayed intact.

The HTask 9: Regrasp Give showed to be more complicated than the other tasks. The participants had to relocate the object when the robot started closing its hand to trigger a regrasp of Floka (see Section 9.2.4). Not all participants understood what was meant with "another position". Some tried it really far away and some even tried to give the object in the left hand of Floka. The integration of gaze seemed to have helped a lot by repeated looking from the participant to Floka's hand and thus communicating where to hand the object. Interview answers like "the gaze helped me to understand that I should come closer as Floka can not reach to where I was standing", hint at the importance of NVC, especially in more complex situations. Along with the data of only one participant not noticing the gaze it shows the usefulness of SR 2(c): Gaze for Predictability. One participant that rated the interaction negatively and as not human-like in the interview, showed to be the only left-handed participant (based on self-assessment) that also had the highest score in the NARS question-set with a score of 5.62 compared to the average of 3.36 ± 1.22 which revealed reluctance towards robots in general. During the interaction another participant sometimes used it's left hand. This participant is likely to have experienced delays, as the hand tracking was parameterized to initialize on the right hand. As stated earlier, the experiment was limited to the right arm due to a broken FTS in the left wrist. Still, both participants successfully exchanged objects with the designed behavior.

10.4 SUMMARY

The evaluation proved that the developed modules and the integrated behavior interplay well together and facilitate a smooth handover. I did not take shortcuts in regard of external computation, tracking, or environment modification and I could fulfill all proposed system requirements. The database approach selected different BT during the interaction based on the prediction of the users motion. Continued by the AT the human hand was reached by the robot. Different user needs and levels were addressed and the success could be shown by means of different tasks. Though the visual handover leaves room for improvement in some cases, the combination with gaze cues and a fall back to another reaching, enabled successful transfer of the object. The detection of completing the receiving, combined with a repetition of the reaching, recovered every time in the HTask 9: Regrasp Give. An analysis of the handover duration showed that while the motion in the HPhase 2: Reach almost reaches human performance, the Transfer is still limited by the speed of the used hardware. The in motion handover was evaluated such that naive interactants, as well as experts, are able to skip trajectory execution if their speed excels the robot's.

Part III

PERSPECTIVES

The last chapters summarize my work and look out on future research opportunities and applications in the field of human-robot object handover along with nonverbal communication. This part ends with a conclusion on the research questions and contributions.

11 DISCUSSION & OUTLOOK

In this chapter I am summing up the findings and my approaches to the topic along with their implications, followed by current limitations and possible continuation. With the work presented in this thesis, I researched the topic of human-robot object handover enhanced by features of nonverbal communication (NVC) cues.

I started with a thorough analysis of the human object handover (handover) phases based on the current literature and my own findings in order to address the hypothesis (H) 1: Handover Has a Distinct Pattern. Establishing five distinct handover phases (HPhases) from HPhase 0: Acquire to HPhase 4: Retreat that smoothly blend over to the subsequent phase showed to be valid also for human-robot handover. My analysis of related work revealed that the HPhase 2: Reach can be split into two parts of reaching motion. Where the first phase is an early estimation of the motion and second the part is the finer adjustment until both interactants meet. I showed that this concept can be transferred to a humanoid robot (humanoid) for predictable as well as adaptable motions. I presented a concept of four handover types that distinguish not only between who gives and receives but also discerns who takes the initiative. This allows a more precise categorization that influences the procedure of object exchange. These concepts were transferred to a humanoid to show the validity of this structure in human-robot handover.

In a first study I tested the proposed structure and research hypotheses. While added gestures with the second arm (H 2: Second Arm Helps to Synchronize), not directly involved in the transfer, did not show a significant effect on the perceived quality of handover, the experience of users with robots did (H 3: Experience Changes Interaction) have an effect on the interaction. This points out the fact that for researchers it is essential to take different needs and expectations into account, those of experts (system requirement (SR) 3(b): Shortcuts for Experts) as well as those of naive users, when researching human-robot interaction (HRI) topics. While addressing naive users becomes more and more relevant in the research community, systems and studies that deal with both are still rare. As the naive users become more experienced through the interaction, both are equally important (SR 3(a): Understandable by Everyone). The experiment also showed the requirement of additional robot functionalities that enable the robot to smoothly hand over objects to a human interactant. This finding results in two requirements that directly address the consideration of different levels of experience, six regarding motions and behavior of an object handing robot, seven concerning its perceptive capabilities, and an additional three addressing the integration into a robot. I discussed these system requirements and presented solutions for each of them.

Adding gaze behavior according to the handover phases allows the robot to communicate intentions and plans during the interaction for an overall improved experience in regard to user comfort and acceptance. As related work already validated the usefulness of gaze in HRI, I incorporated gaze cues for natural and human-like behavior from the beginning of my work. Additionally, I presented concepts for improved gaze in terms of reactivity and handover phase synchronization. One of the most important aspects is the reaching motion of the interactants during a handover. Thus, I developed a novel strategy that combines a database of prerecorded motions with inverse instantaneous kinematics (IIK) for generating predictable gesture-like motions that converge to the target. Combined with a fast hand tracking and object transfer point (OTP) prediction pipeline the system proved to be a good approach to the stated problem. Also, it was incorporated to allow transferring the object without applying noticeable force to the robot, which shows to be an important improvement, compared to other approaches. This method showed to be valid and the modules worked well together. The early prediction allowing the selection of a base trajectory (BT) from the database, to maintain the gesture characteristics of the motion, to inform the interactant about the *what*, *when* and *where*. An improved transfer detection module helps to reduce failures in the HPhase 3: Transfer. By an integrated behavior, I presented an approach to how to interleave the created modules for a near zero-delay interaction while maintaining modularity and interchangeability of components. I presented recovery strategies that help to resume the interaction, without interference of an operator. On occurrence of unexpected events, like the robot not having an object the designed behavior triggers a regrasp or a repetition of reaching when the distance to the human's hand is still to big.

In the knowledge, that the experience of interactants with robots influences the behavior, I evaluated the system with users that did not have experience and thus expectations concerning that specific robot. Specifically crafted tasks for typical behaviors discovered from different users allowed to show the validity of the overall approach to handover. By allowing visual and in motion handover, adding gaze cues, and adapting to the human I integrated features that shift the effort from the human to the robot. With the presented concept and implementation Floka humanoid (Floka) was able to successfully hand over the object 135 times while dropping it only once. Therefore, the system operated a whole day without requiring a restart, highlighting its stability. Compared to the first study this can be considered as a good result for such a highly complex system that operated fully autonomously and fully integrated into a mobile robot without any external sensing, processing, or modifications of the environment (SR 5(b): Onboard Sensing and Processing and SR 4(g): Markerless Perception). While the speed was rated as improvable, it was also rated as human-like and predictable. Here, the transfer did not match the expectations due to the low opening and closing speed of the robot's hand by hardware limitations. Nevertheless, the reaching motion comes already close to human performance.

Overall, I presented and evaluated a system that is able to accept shortcuts by users and does not take shortcuts itself in overcoming challenges in human-robot handover. By this work I contribute towards robots that are able to interact and help in a more acceptable and predictable way.

11.1 CURRENT LIMITATIONS

While, in the individual evaluation chapters, I focused on developing the system that can be understood by everyone (SR 3(a)), the participants were mostly Western, Educated, Industrialized, Rich, and Democratic (WEIRD). As this group of people accounts for only 12 % of the world's population, it limits the scope of this work [HHN10]. Especially the meaning of gestures are highly influenced by culture and social norms. While visiting Tsukuba, Japan, to present the results of my first study on the ICSR 2017, I experienced such differences myself. In the Japanese culture, one-handed handovers are really uncommon and considered disrespectful. This highlights gesture as a function of handover. While the functionality of transferring control over an object is always the same, the way it is done and the meaning might be different. It would be compelling to develop concepts that help to embed such cultural differences in humanoids. While adding such different motions to the proposed database is possible, the correct selection and synchronization of two arms is still an open challenge. As my approach aims to generalize well over different types of people, some might prefer options to customize the system to their needs. While there are already many parameters to adjust and tune the system, not all are easily exposable to the interactant and influence each other, making it even harder to tune them. The velocity and especially acceleration impact forces acting on the force/torque-sensor (FTS) and would thus require a parameter adopted for the transfer function. However, the thresholds for visual and force transfer could also be adjusted by a non-expert by reducing it until object drops occur and then backing off the parameter again. Still, it needs to be taken into account, that at least some understanding of the perceptive capabilities of the system would help. Incorporating the proposed mixed reality (MR) techniques could be used to show such percepts to the user and thus allow an easier adjustment. Nevertheless, an automatism for finding suitable parameter sets could help to overcome manual parameter search.

Another limitation poses the objects chosen for the handover. The cylindrical object only differed in color. In the presented tracking module there was no distinction between hand and object. The approach of considering them a single entity that needs to be reached for, showed good results for the selected object. However, with a smaller data set we were able to indicate the validity of the technique for more complex objects like boxes or bowls [Sim19]. Nevertheless, where precision becomes crucial, like for smaller objects, the distinction of both might become more important. While the approach is transferable to similar objects, more complex ones, with affordances like handles and other functional parts, could change the way the object is transferred and even further influence the social layer of implication of the motion.

11.2 OUTLOOK AND FUTURE WORK

In this chapter I give an outlook on future work in terms of improvements, applications, and implications of this work, addressing the before-mentioned limitations, as well as opening up new research opportunities.

VIRTUAL AND MIXED REALITY With the work described in Section 4.5 on MR and virtual reality (VR) we already explored virtualization technologies in the context of human-robot handover. In the future one could examine the influence of different head features and designs on handover without the need of changing the hardware. But not only overcoming limitations in regard to the appearance but also on velocities, accelerations, and joint limits offer further possibilities. One could imagine a system that completely decouples what the interactant sees with a head-mounted display (HMD) from what happens in the real world. A precise and fast (industrial-grade) robot could then be used to physically interact with the human while simulating a different visual impression. Such an environment would allow to research the influence of the shown motions individually and give even more precise answers to topics addressed in this thesis. As already proposed earlier, MR can be used as a tool for visualizing internal robot data like sensor measurements, predicted OTPs, or joint limits. These techniques can be used to adjust the aforementioned parameters or to train naive users to become quicker and to profit from the shortcuts experts are able to take like the addition of early in motion handover. One could also address the discussed WEIRD aspect more easily, as a virtual robot is not as hard to bring in different places for evaluation in diverse contexts and cultures.

RANGE OF APPLICATION It would be worthwhile to verify the proposed methods in a broader range of application with different human postures like sitting or lying, as scenarios such as assisting at work or care often require such handovers. With the current design making use of a database for the first part of the reaching motion and a generic approach for the fine adaption, trajectories that reach for a sitting or lying person could be easily added. In a preliminary study it could be shown that such positions are also reachable. As lying and sitting people are immobile, it would have switched the HPhase 1: Approach from the human to the robot. This would have made the study design and especially its recording more complicated and was consequently not examined. It would also be interesting to test the presented concepts and approaches with additional groups of persons like elderly people or children. This way limitations of the stated approaches could be revealed and the boundaries be pushed even further. Bringing Floka to care homes or kindergartens could give new insights in additional groups of interactants. Especially for an even wider view on the topic, the problem of parameter selection becomes vital. While machine learning offers an increasingly growing number of tools for such problems, most of them require large amounts of data for good results. Especially in HRI this data is hard to obtain. Thus, learning algorithms requiring few samples could be integrated in handover scenarios.

MOTION GENERATION Right now, the database of base trajectories (BTs) requires manual crafting of the first part of the reaching motion for different target positions. While the adaptive trajectory (AT) showed to continue the BT well, thus it does not require many entries for a good workspace coverage, it would need to be done for each type of robot. Nevertheless, a method that takes over such a task would give the approach a wider applicability. Filling the database with trajectories from the joint motion model (JMM) could be a promising approach. While this method is currently not fast enough for live application, filling the BT-databases of different robots can be validated. On the other hand, the AT can be improved by testing various approaches for dynamically selecting the ideal damped least squares (DLS) damping parameter [Bus09] for even faster and smoother convergence. It could also benefit from exploring dynamic collision avoidance methods to better keep distance between elbow and the robot's chest while converging to the OTP. Here, considering the robot as a whole with Whole-Body Control (WBC) might give fast and reliable solutions [Sen07]. Experiments, with that approach, on a Meka M1 Mobile Manipulator (Meka M1) have proved to be promising in various tasks, whereby HRI still needs to be addressed [PSK11; FS16]. It is still a challenge to transfer such implementations to different robots and to model tasks like handover in such frameworks.

EVALUATION The presented evaluation methods can be integrated in future HRI scenarios. During the two human-robot handover studies carried out in my work over eight hundred recordings were generated. These can be further analyzed to get an even deeper insight in the motions and forces during the process. The data of the final evaluation could be checked for timings to discover slowdowns in the process. While designed modules and integrated systems showed a good performance, an evaluation on different robots has not been accomplished yet. This remains a desideratum. Other humanoids could already benefit from the integrated NVC and the software stack offers opportunities for directly comparing different platforms in the same HRI scenario. Also, it would be interesting to see, to which degree industrial style robots could benefit from the NVC aspects.

12 CONCLUSION

In this thesis, I presented research in the field of human-robot object handover combined with nonverbal communication like gaze and gestures. I discussed, implemented, and examined the whole process on an autonomous humanoid robot. In contrast to many other approaches to the topic, I did not take shortcuts like using an external tracking system or artificial markers. All the sensing and processing was integrated into the single coherent robot called Floka.

In research question (RQ) 1: Handover Interaction Pattern, I investigated the general process of object handover by extracting information from previous work in this field of research, combining it to a coherent model and applying it to a humanoid. For RQ 2: Impact of Nonverbal Communication I presented concepts for integration of different types of NVC. While previous work had already experimented with NVC in HRI, the influence and effect depends highly on the overall interaction and behavior shown by the robot. Thus, I created an integrated behavior and approached the topic as a whole, from adding gestures with a second arm to making the HPhase 2 motion of object handover a gesture itself. I also complemented it by contributing an integrated structure and concept for robotic gaze. In regard to RQ 3: Influence of Expert Knowledge I addressed an often underrated topic in HRI. Already my first experiment revealed that naive and experienced users have different expectations and requirements in terms of human-robot object handover. Contributing to more useful and accepted robots, I presented a model that aims to meet the requirements of different experience levels by allowing shortcuts like in motion handover and adding recovery strategies for unexpected behavior. I evaluated it in a variety of scenarios to show the validity of the approach. RQ 4: Perception Requirements addresses technical aspects of the system. Therefore, I introduced an ample list of system requirements for smooth object handovers and contributed a core component for hand tracking and prediction of OTPs. This resulted in a system that is ready to use for human-robot object handover by a variety of interactants without any further explanation.

Overall, this work contributes to a better understanding of human-robot object handover and the involved nonverbal communication cues by presenting solutions and interaction studies with users having different levels of experience on a fully autonomous robot with its own inherent limitations and properties. Approaching the subject without shortcuts gives a holistic view on the topic of advancing human-robot interaction.

Part IV

APPENDIX

The Appendix starts with additional material from experiments and studys. It contains information on used abbreviations and important terms, followed by the bibliography starting with my own publications that originated during my phd, my RoboCup ToBi involvement, other literature that inspired this work, and the software packages used in this work.

A GESTURE STUDY MATERIAL

A.1 RECORDED ROS-TOPICS

- /hand_over/(.*)
- /meka_roscontrol/(.*)/follow_joint_trajectory/(.*)
- /joint_states
- (.*)/camera_info
- /rosout
- /tf
- /tf_static
- /usb_cam/image_raw/compressed
- /xtion/rgb/image_raw/compressed
- /meka_ros_pub/m3loadx6_ma29_l0/wrench
- /meka_ros_pub/m3loadx6_ma30_l0/wrench

A.2 HANDOVER QUESTIONNAIRE 2017

Floka-Studie 2017

Diese Umfrage enthält 12 Fragen.

Probandennummer

[]Vom Versuchsleiter vergebene ID *

Bitte geben Sie Ihre Antwort hier ein:

Figure A.1: Questionnaire Question Block #1

Persönliche Informationen

[]Haben sie einen VPN-Code aus dem Psychologie Forschungsportal?

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ Ja
- ○ Nein

[]Bitte VPN-Code angeben *

Bitte geben Sie Ihre Antwort hier ein:

[]Bitte wählen Sie Ihr Geschlecht:

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ weiblich
- ○ männlich

[]Bitte geben Sie Ihr Alter an: *

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ 18-24
- ○ 25-34
- ○ 35-44
- ○ 45-54
- ○ 65-74
- ○ >75

[]Bitte geben Sie an, wieviel Erfahrung Sie haben mit: *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1 (keine)	2	3	4	5	6	7 (sehr viel)
Nutzung von Computern	○	○	○	○	○	○	○
Programmierung von Computern	○	○	○	○	○	○	○
Nutzung von Robotersystemen	○	○	○	○	○	○	○
Dem Roboter Floka oder seiner Simulation	○	○	○	○	○	○	○

Figure A.2: Questionnaire Question Block #2

Einstellung gegenüber Robotern (NARS)

[] *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1 (Ich stimme gar nicht zu)	2	3	4	5	6	7 (Ich stimme voll zu)
Ich würde mich unwohl fühlen, wenn ich auf der Arbeit mit Robotern zu tun hätte.	○	○	○	○	○	○	○
Es könnte etwas Schlimmes passieren, wenn sich Roboter zu Lebewesen entwickeln würden.	○	○	○	○	○	○	○
Ich wäre entspannt, wenn ich mit einem Roboter spräche.	○	○	○	○	○	○	○
Wenn Roboter Emotionen besäßen, könnte ich mich mit ihnen anfreunden.	○	○	○	○	○	○	○
Ich würde mich unwohl fühlen, wenn Roboter tatsächlich Emotionen besäßen.	○	○	○	○	○	○	○
Das Wort "Roboter" hat keine Bedeutung für mich.	○	○	○	○	○	○	○
Ich wäre nervös, wenn ich vor anderen Leuten einen Roboter bedienen müsste.	○	○	○	○	○	○	○
Ich verabscheue die Vorstellung, dass Roboter oder künstliche Intelligenzen sich Urteile über Dinge bilden können.	○	○	○	○	○	○	○
Vor einem Roboter zu stehen, würde mich nervös machen.	○	○	○	○	○	○	○
Ich glaube, dass etwas Schlimmes passiert, wenn ich zu sehr von Robotern abhängig wäre.	○	○	○	○	○	○	○
Ich würde mich paranoid fühlen, wenn ich mit einem Roboter spräche.	○	○	○	○	○	○	○
Ich mache mir Sorgen, dass Roboter einen schlechten Einfluss auf Kinder haben könnten.	○	○	○	○	○	○	○
Ich glaube, dass die Gesellschaft in Zukunft von Robotern dominiert wird.	○	○	○	○	○	○	○

Figure A.3: Questionnaire Question Block #3

Interaktion

[]Die Interaktion mit dem System *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1 (Ich stimme gar nicht zu)	2	3	4	5	6	7 (Ich stimme voll zu)
fiel mir leicht	○	○	○	○	○	○	○
war effizient	○	○	○	○	○	○	○
war selbsterklärend	○	○	○	○	○	○	○
war angenehm	○	○	○	○	○	○	○
war leicht zu verstehen	○	○	○	○	○	○	○
hat Spaß gemacht	○	○	○	○	○	○	○
war flüssig	○	○	○	○	○	○	○
war vorhersehbar	○	○	○	○	○	○	○
war eindeutig	○	○	○	○	○	○	○
war zügig	○	○	○	○	○	○	○

Figure A.4: Questionnaire Question Block #4

Bewertungen

[]

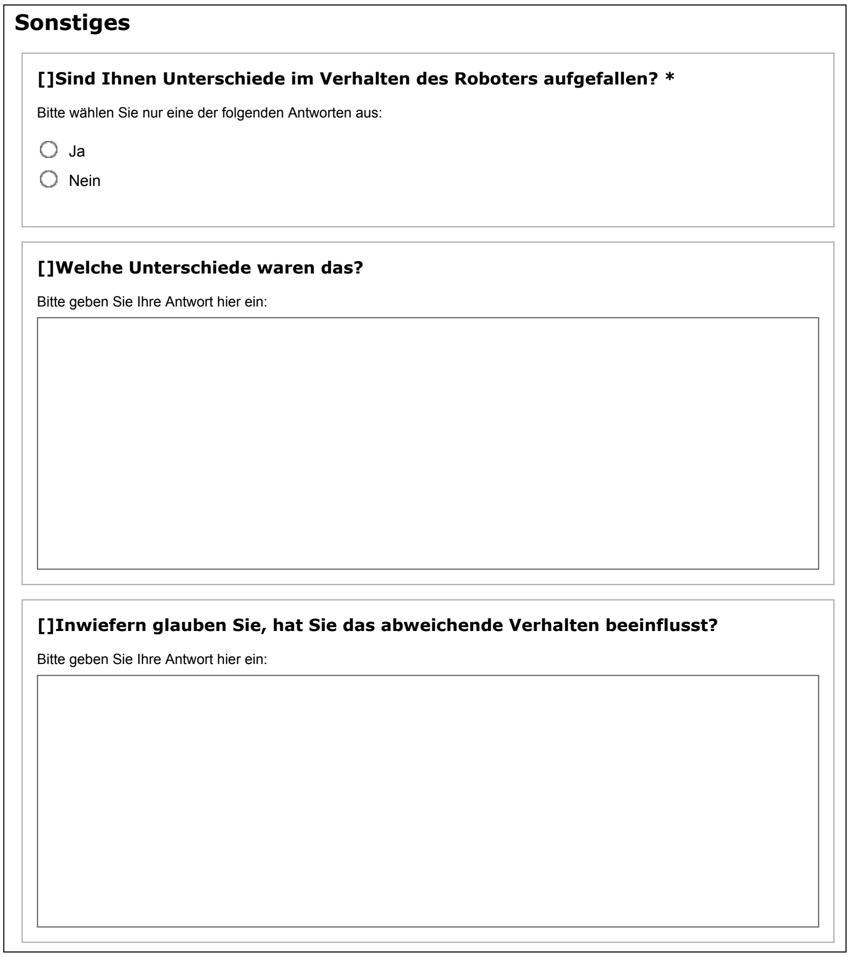
Bitte bewerten Sie das System auf der folgenden Skala:

*

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1	2	3	4	5	
maschinenhaft	○	○	○	○	○	menschenähnlich
hat kein Bewusstsein	○	○	○	○	○	hat ein Bewusstsein
künstlich	○	○	○	○	○	realistisch
bewegt sich steif	○	○	○	○	○	bewegt sich flüssig
unecht	○	○	○	○	○	natürlich
tot	○	○	○	○	○	lebendig
unbewegt	○	○	○	○	○	lebhaft
mechanisch	○	○	○	○	○	organisch
träge	○	○	○	○	○	interaktiv
teilnahmslos	○	○	○	○	○	ansprechbar
unsympathisch	○	○	○	○	○	sympathisch
unfreundlich	○	○	○	○	○	freundlich
unhöflich	○	○	○	○	○	höflich
unangenehm	○	○	○	○	○	angenehm
furchtbar	○	○	○	○	○	nett
inkompetent	○	○	○	○	○	kompetent
ungebildet	○	○	○	○	○	gebildet
verantwortungslos	○	○	○	○	○	verantwortungsbewusst
unintelligent	○	○	○	○	○	intelligent
unvernünftig	○	○	○	○	○	vernünftig
ängstlich	○	○	○	○	○	entspannt
aufgewühlt	○	○	○	○	○	ruhig
still	○	○	○	○	○	überrascht
demotivierend	○	○	○	○	○	motivierend

Figure A.5: Questionnaire Question Block #5



Sonstiges

[]Sind Ihnen Unterschiede im Verhalten des Roboters aufgefallen? *

Bitte wählen Sie nur eine der folgenden Antworten aus:

- Ja
- Nein

[]Welche Unterschiede waren das?

Bitte geben Sie Ihre Antwort hier ein:

[]Inwiefern glauben Sie, hat Sie das abweichende Verhalten beeinflusst?

Bitte geben Sie Ihre Antwort hier ein:

Figure A.6: Questionnaire Question Block #6

A.3 INSTRUCTIONS

A.3.1 *Versuchsleiter Anweisungen*

Deine Aufgabe heute ist es, mit Floka [zeige auf Floka] Objekte zu lernen, beziehungsweise ihn dabei zu unterstützen. Es geht dabei um diese drei Objtekte [zeige auf Objekte]. Du kannst dir auch gerne schon eins davon nehmen [zeige auf Objekte und warte, bis eins genommen wurde]. Der Ablauf ist dabei folgender: Der Roboter wird zuerst ein Objekt von dir erwarten. Dieses wird er dann lernen und dann zurückgeben. Nimm es ihm dann bitte ab und stell es wieder auf den Tisch und nimm das nächste Objekt. Das ganze wiederholst du dann für jedes Objekt dreimal, also insgesamt neunmal. Hast du noch fragen? [VL setzt sich neben Kamera, Experiment wird über "Notaus" gestartet]

A.3.2 *Examiner Instructions [translated]*

It is your task to learn objects with Floka [point at Floka], respectively help the robot at this task. These three objects [point at objects] have to be learned. You can take one already. [wait until participant took one] The procedure is as follows: The robot expects to receive an object from you. It will learn the given object. After that it will give back the object. Please take it from the robot and put it back on the table to take the next object. Repeat that procedure three times for each object, resulting in a total of nine. Do you have any questions? [sit down next to the camera, start the experiment with the remote]

B GESTURE STUDY TIMING PLOTS

B.1 DURATION BY EXPERIENCE

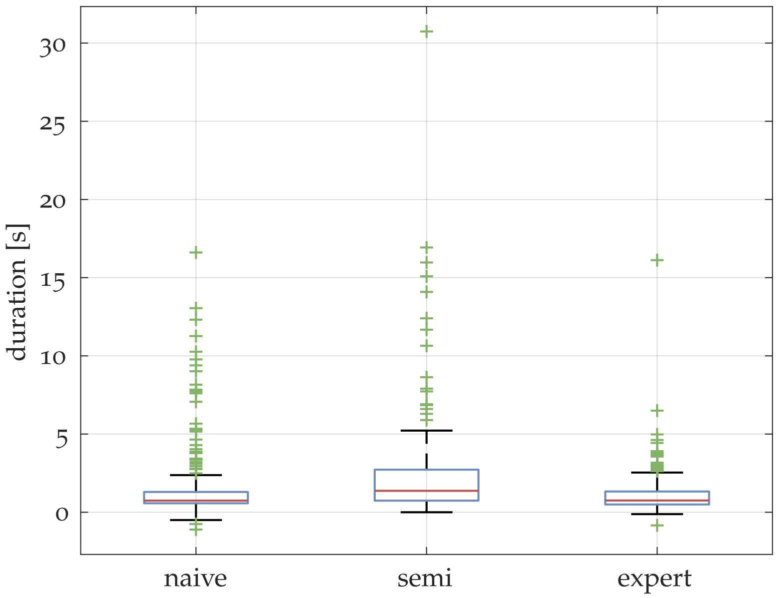


Figure B.1: The transfer times of my first object handover (handover) study grouped by the three experience groups. The extracted times are visualized as a boxplot. The central red line shows the median. Top and bottom edges of the blue box mark the 25th and 75th percentiles. The whiskers reach to minimum and maximum, not considering outliers, which are plotted individually using a green + symbol.

B.2 DURATION BY CONDITION

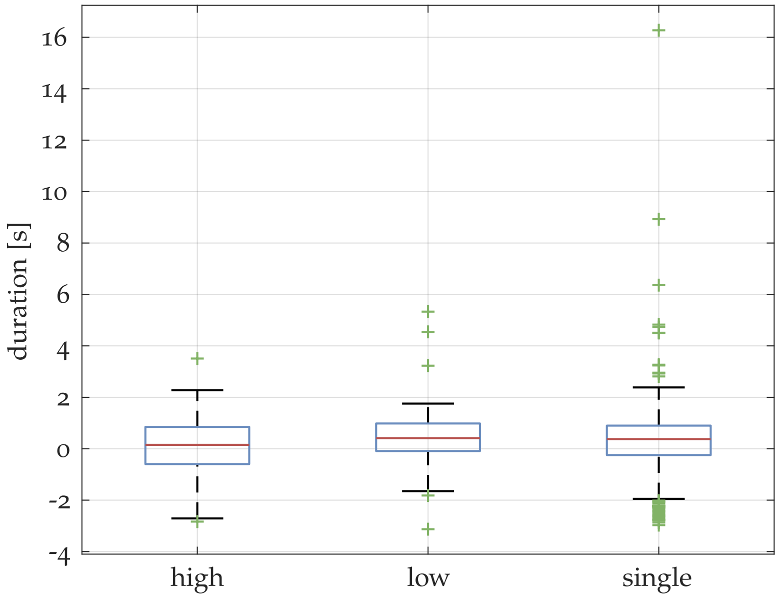


Figure B.2: The reaction times of my first handover study grouped by the three gesture conditions. The extracted times are visualized as a boxplot.

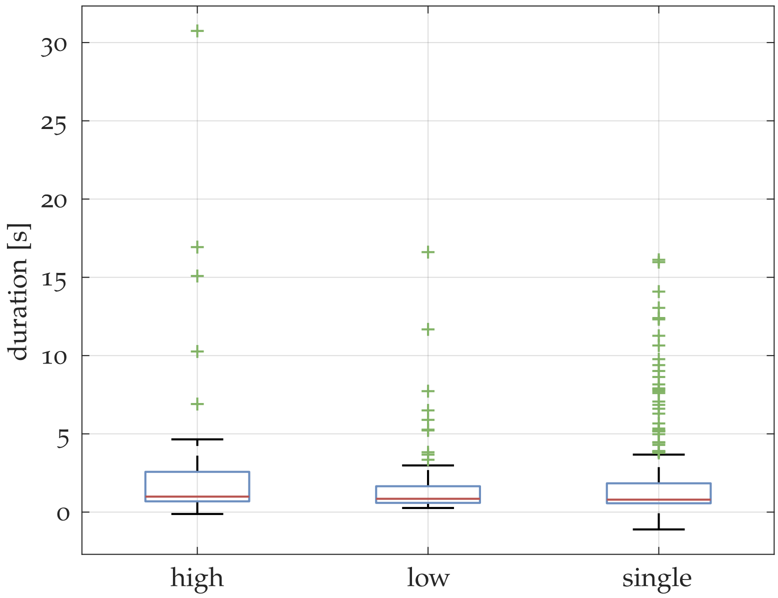


Figure B.3: The transfer times of my first handover study grouped by the three gesture conditions. The extracted times are visualized as a boxplot.

C

ADAPTION PLOTS

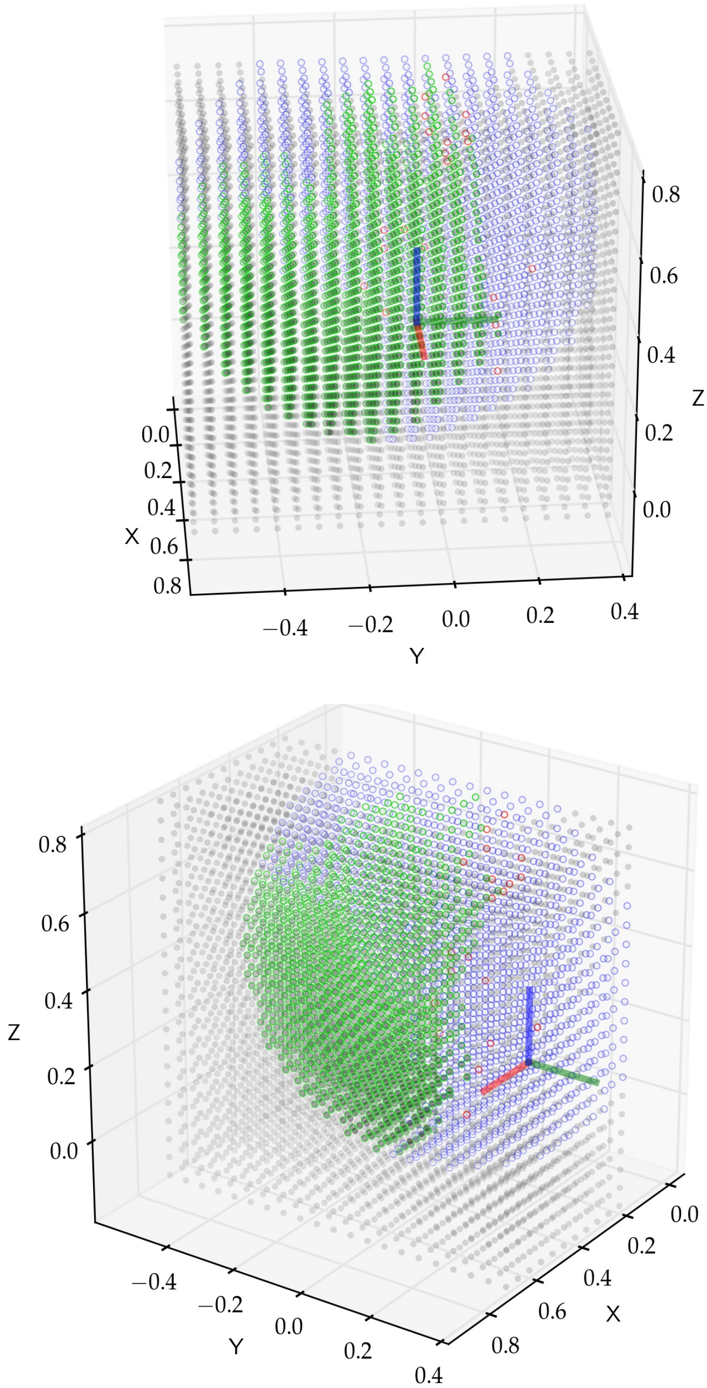
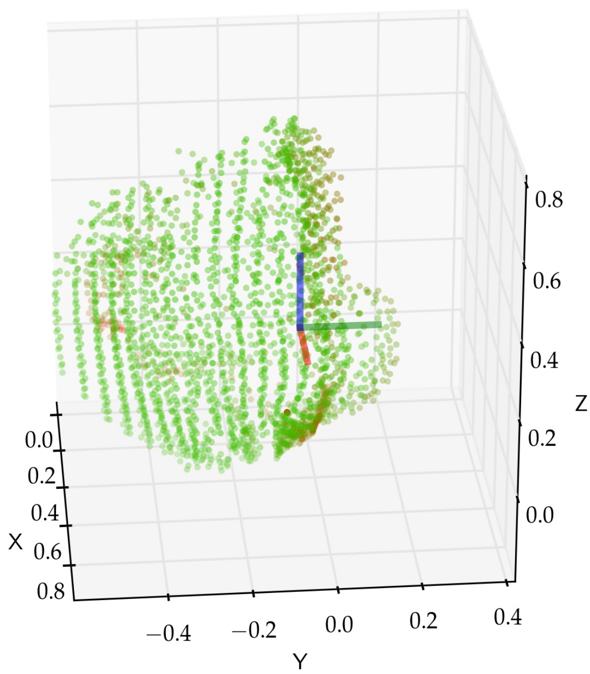


Figure C.1: Different perspectives on the adaption workspace (Fig. 7.4).



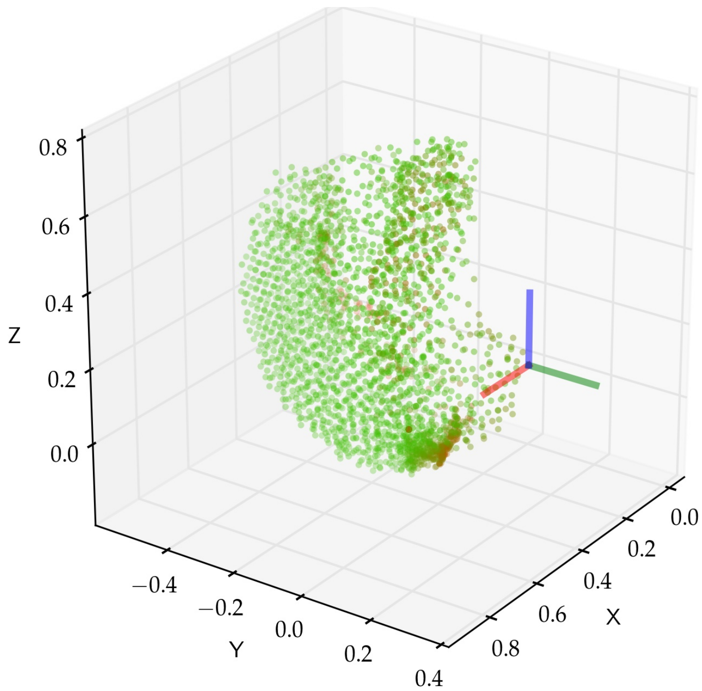


Figure C.2: Different perspectives on the adaptive trajectory Reachability (Fig. 7.5).

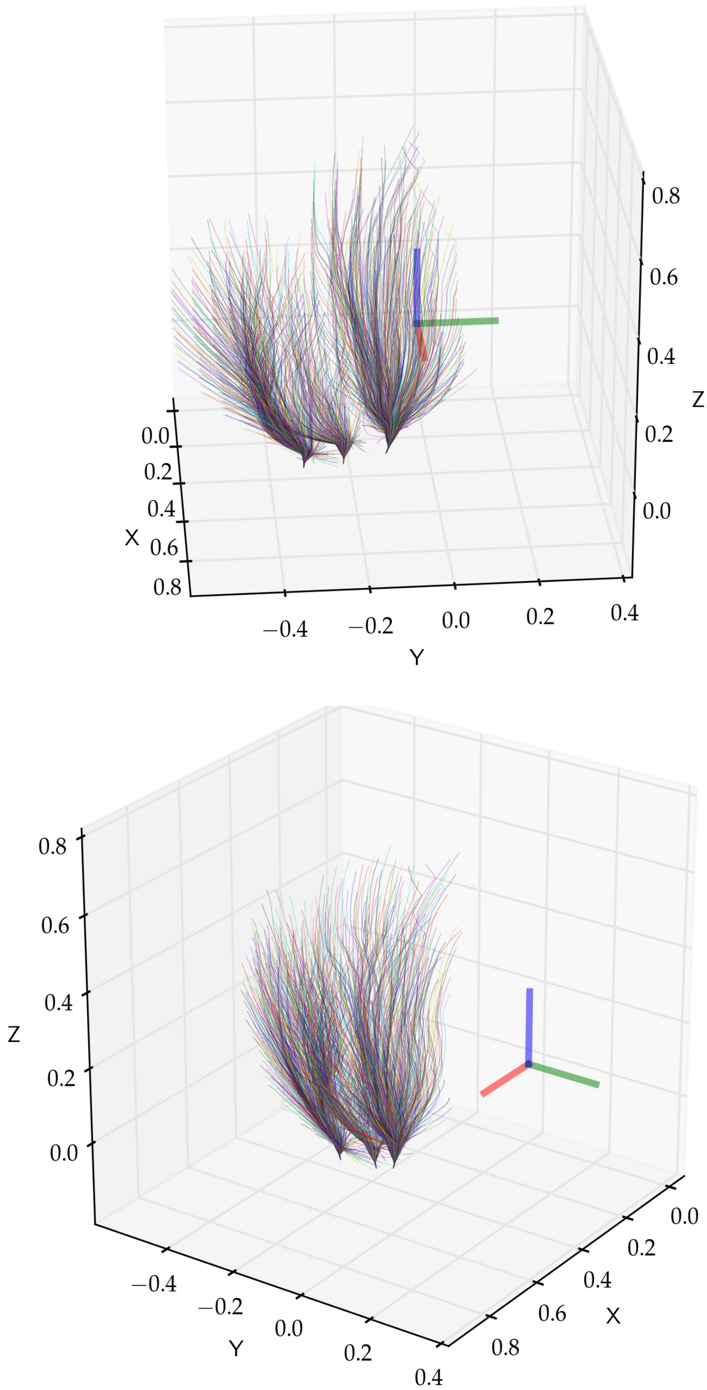


Figure C.3: Different perspectives on the adaptive trajectories (Fig. 7.6).

D FINAL SYSTEM-EVALUATION MATERIAL

D.1 RECORDED ROS-TOPICS

- /usb_cam/image_raw/compressed
- /usb_cam_top/image_raw/compressed
- /realsense_face/color/image_raw/compressed
- /tf
- /tf_static
- /meka_roscontrol/(.*)/follow_joint_trajectory/(.*)
- /do_adaption/(.*)
- (.*)/camera_info
- /otpprediction/(.*)
- /people_tracker(.*)
- /flexbe/(.*)
- /rosout
- /floka/eye_right/image_color/compressed
- /openface2/faces
- /force_helper
- /force_helper/arm
- /force_helper/raw
- /force_helper/result
- /gaze_relay/target
- /gaze_relay/target_point
- /meka_ros_pub/m3loadx6_ma29_l0/wrench
- /meka_roscontrol/right_arm_
  position_trajectory_controller/command
- /xsc3/joint_states
- /joint_states

D.2 BEHAVIOR STATEMACHINE

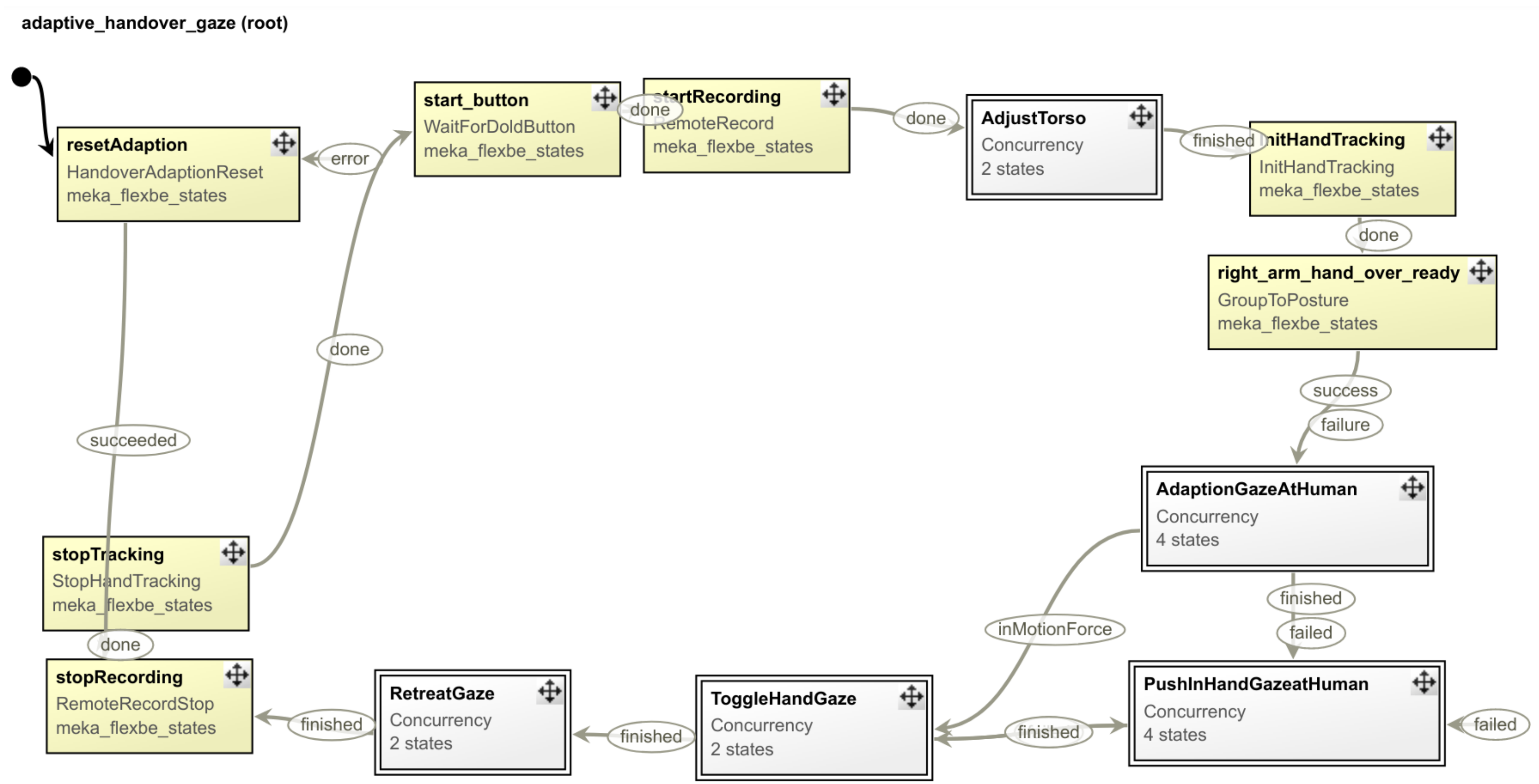


Figure D.1: The *FlexBE - The flexible behavior engine* [FlexBE] behavior created for the evaluation. Besides the important handover states, recording, resetting and stopping states were added.

D.3 HANDOVER QUESTIONNAIRE 2019

Floka-Studie 2019

Diese Umfrage enthält 10 Fragen.

Probandennummer

[]Vom Versuchsleiter vergebene ID *

Hier dürfen nur ganze Zahlen (integer) eingegeben werden.

Bitte geben Sie Ihre Antwort hier ein:

Figure D.2: Questionnaire Question Block #1

Persönliche Informationen

[]Bitte geben Sie Ihr Geschlecht an:

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ weiblich
- ○ männlich
- ○ anderes

[]Bitte geben Sie Ihr Alter an:

Jede Antwort muss zwischen 18 und 99 sein
Hier dürfen nur ganze Zahlen (integer) eingegeben werden.

Bitte geben Sie Ihre Antwort hier ein:

[]Bitte geben Sie an, wieviel Erfahrung Sie haben mit: *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1 (keine)	2	3	4	5	6	7 (sehr viel)
Nutzung von Computern	○	○	○	○	○	○	○
Programmierung von Computern	○	○	○	○	○	○	○
Nutzung von Robotersystemen	○	○	○	○	○	○	○
Dem Roboter Floka oder seiner Simulation	○	○	○	○	○	○	○

[]Haben oder hatten Sie ein Haustier?

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ Ich habe atkuell ein Haustier
- ○ Ich hatte schon mal ein Haustier
- ○ Ich hatte bisher kein Haustier

Figure D.3: Questionnaire Question Block #2

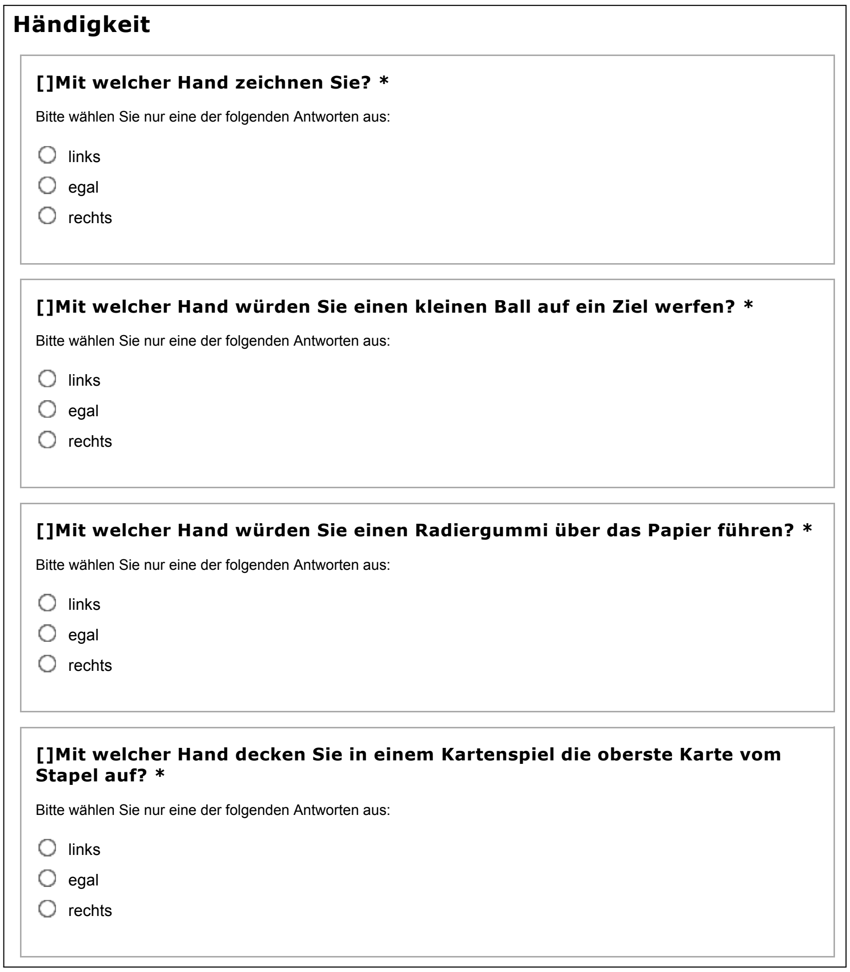
Einstellung gegenüber Robotern (NARS)

[] *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1 (Ich stimme gar nicht zu)	2	3	4	5	6	7 (Ich stimme voll zu)
Ich würde mich unwohl fühlen, wenn ich auf der Arbeit mit Robotern zu tun hätte.	○	○	○	○	○	○	○
Es könnte etwas Schlimmes passieren, wenn sich Roboter zu Lebewesen entwickeln würden.	○	○	○	○	○	○	○
Ich wäre entspannt, wenn ich mit einem Roboter spräche.	○	○	○	○	○	○	○
Wenn Roboter Emotionen besäßen, könnte ich mich mit ihnen anfreunden.	○	○	○	○	○	○	○
Ich würde mich unwohl fühlen, wenn Roboter tatsächlich Emotionen besäßen.	○	○	○	○	○	○	○
Das Wort "Roboter" hat keine Bedeutung für mich.	○	○	○	○	○	○	○
Ich wäre nervös, wenn ich vor anderen Leuten einen Roboter bedienen müsste.	○	○	○	○	○	○	○
Ich verabscheue die Vorstellung, dass Roboter oder künstliche Intelligenzen sich Urteile über Dinge bilden können.	○	○	○	○	○	○	○
Vor einem Roboter zu stehen, würde mich nervös machen.	○	○	○	○	○	○	○
Ich glaube, dass etwas Schlimmes passiert, wenn ich zu sehr von Robotern abhängig wäre.	○	○	○	○	○	○	○
Ich würde mich paranoid fühlen, wenn ich mit einem Roboter spräche.	○	○	○	○	○	○	○
Ich mache mir Sorgen, dass Roboter einen schlechten Einfluss auf Kinder haben könnten.	○	○	○	○	○	○	○
Ich glaube, dass die Gesellschaft in Zukunft von Robotern dominiert wird.	○	○	○	○	○	○	○

Figure D.4: Questionnaire Question Block #3



Händigkeit

[]Mit welcher Hand zeichnen Sie? *

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ links
- ○ egal
- ○ rechts

[]Mit welcher Hand würden Sie einen kleinen Ball auf ein Ziel werfen? *

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ links
- ○ egal
- ○ rechts

[]Mit welcher Hand würden Sie einen Radiergummi über das Papier führen? *

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ links
- ○ egal
- ○ rechts

[]Mit welcher Hand decken Sie in einem Kartenspiel die oberste Karte vom Stapel auf? *

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ○ links
- ○ egal
- ○ rechts

Figure D.5: Questionnaire Question Block #4

D.4 HANDOVER INSTRUCTIONS 2019

D.4.1 *Interaktionsaufgaben*

Deine Aufgabe ist es, mit Floka [zeige auf Floka] Objekte auszutauschen. Lies eine Aufgabe [zeige auf Aufgabenstapel] und führe sie aus, kehre danach zu den Aufgaben am Tisch zurück um die nächste Aufgabe zu lesen. Insgesamt gibt es 10 Aufgaben. Bitte führe alle Aufgaben mit der rechten Hand aus.

1. Gib Floka das Objekt.
2. Nimm das Objekt von Floka.
3. Gib das Objekt, indem du es hinhältst.
4. Nimm das Objekt von Floka, ohne daran zu ziehen.
5. Gib das Objekt, indem du es in die Hand drückst.
6. Nimm das Objekt, indem du es aus der Hand ziehst.
7. Gib das Objekt so früh wie möglich.
8. Nimm das Objekt so früh wie möglich.
9. Gib Floka das Objekt, aber entziehe es ihm wieder, bevor seine Hand geschlossen ist.
   Gib es ihm dann an einer anderen Position.
10. Nimm das Objekt von Floka.

D.4.2 *Scripted Interaction Tasks [translated]*

Your task is to exchange objects with Floka [point at Floka]. Read your task [point at the stack of cards] and execute it, after come back to the tasks at the table and start the next task. There are 10 in total. Please execute all tasks with the right hand.

For a list of translated tasks refer to 122.

D.5 HANDOVER INTERVIEW 2019

D.5.1 *Interview Fragen*

1. Wie hat Floka auf dich gewirkt, als du auf ihn zugegangen bist?
2. Wie hat der Roboter sich dabei verhalten?
3. Wie hast du empfunden,
   als du ihm deine Hand entgegengestreckt hast?
4. Wusstest du, wohin du das Objekt übergeben solltest?
5. Wie hat der Roboter dies signalisiert?
6. Wie hat Floka seinen Arm dabei bewegt?
7. Wohin hat der Roboter dabei geguckt?
8. Wie fühlte sich der Austausch des Objektes an?
9. Wusstest du, wann du das Objekt wieder loslassen konntest?
10. Wie hat der Roboter dies signalisiert?
11. Wie hat sich Floka dabei verhalten?
12. Wie hat der Roboter dabei gegeguckt?
13. Gab es Unterschiede ob du das Objekt genommen oder gegeben hast?
14. Welche der Aufgaben hat besonders gut funktioniert?
15. Hat sich die Interaktion über die Aufgaben hinweg verändert?
16. Bei welcher Aufgabe gab es besondere Probleme?
17. Hat dich etwas bei der Interaktion irritiert?
18. Ist dir insgesamt etwas besonders positiv aufgefallen?
19. Könntest du dir vorstellen, mit Floka zu arbeiten?
20. Wie sicher schätzt du die Übergabe ein?

D.5.2 *Interview Questions [translated]*

1. How did Floka appear to you, as you approached?
2. How did the robot behave in the course of this?
3. How did you feel, while holding out your hand?
4. Did you know where to hand over the object?
5. How did the robot signal it to you?
6. How did Floka move the arm in the process?
7. Where did the robot look at while doing so?
8. How did the object exchange feel?
9. Did you know when to release the object from your grasp?
10. How did the robot signal you to do so?
11. How did Floka behave while exchanging the object?
12. Where did the robot look in the course of this?
13. Did you notice differences between giving and taking the object?
14. Which of the tasks worked really good?
15. Did the interaction change during the tasks?
16. In which tasks did you encounter problems?
17. Did you find something irritating during the interaction?
18. Was there something positive you would like to mention?
19. Could you imagine working together with Floka?
20. How safe would you rate the handover?

E

EVALUATION DURATION AS BOXPLOT

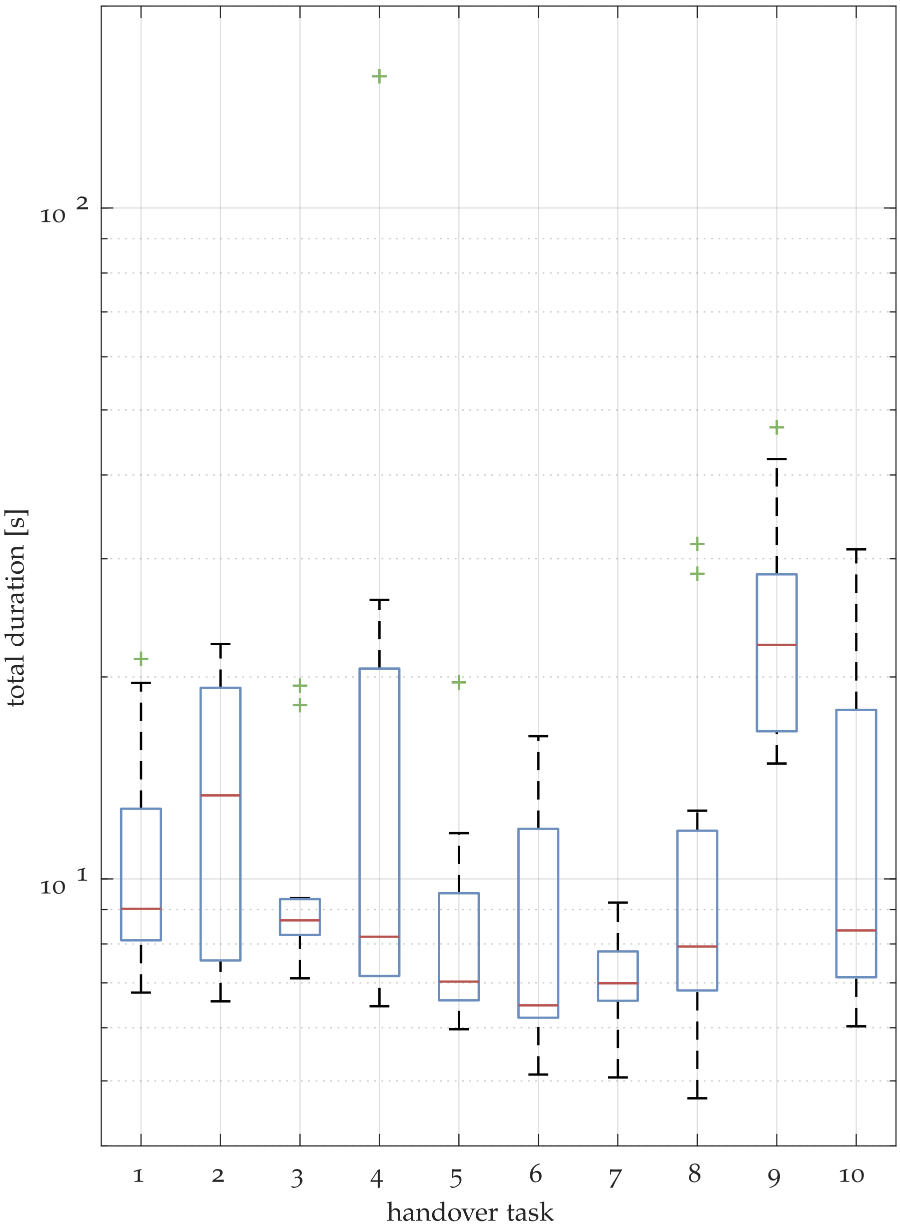


Figure E.1: The duration of the whole handover in my final evaluation grouped by the ten handover tasks. The extracted times are visualized as a boxplot using a logarithmic y-axis to compress the outlier.

ACRONYMS AND ABBREVIATIONS

Numbers

3D

three-dimensional. *used on: pp. 31, 38, 44, 58, 59, 90, 93, 95, 96*

A

ANOVA

ANalysis Of VAriance. *used on: pp. 70–72*

AR

augmented reality. *used on: pp. 56, 57*

AT

adaptive trajectory. *used on: pp. 82, 84–86, 89–91, 112, 116, 132, 139, 154, 155, 173*

B

BT

base trajectory. *used on: pp. 82, 84–86, 88–91, 112, 114, 116, 126, 132, 136, 139, 173*

C

CAVE

CAVE Automatic Virtual Environment. *used on: pp. xiii, 59, 60, 167*

CITEC

Cluster of Excellence Cognitive Interaction Technology. *used on: p. 55*

CITK

Cognitive Interaction Toolkit. *used on: p. 54*

CSRA

Cognitive Service Robotics Apartment. *used on: pp. vii, xiii, 55, 65, 120, 173*

D

DFD

data-flow diagram. *used on: pp. 86, 96*

DLS

damped least squares. *used on: pp. 83–86, 116, 139, 173*

DMPs

dynamic movement primitives. *used on: pp. 20, 42, 43, 173*

DOF

degree of freedom. *used on: pp. 40, 41, 43, 45, 51, 52, 82*

E

EEF
: end-effector. *used on: pp. 4, 5, 7, 11, 12, 16, 20, 25, 29, 31, 39–43, 45–47, 52, 53, 62, 64, 73, 77, 78, 80–84, 88, 101, 108, 112, 114–116, 125–127, 129, 130, 174, 175, 177*

EtherCAT
: Ethernet for Control Automation Technology. *used on: pp. 51, 53*

F

FK
: forward kinematics. *used on: p. 82*

Floka
: Floka humanoid. *used on: pp. v, 55, 57, 58, 75, 76, 80, 97, 107, 111, 115, 119–121, 123–131, 136, 139, 141, 174*

FoV
: field of view. *used on: pp. 56–58, 60, 90, 97, 101, 102, 111, 114*

FPS
: frames per second. *used on: pp. 94, 97, 99*

FSM
: finite-state machine. *used on: pp. 9, 79, 108, 112*

FTS
: force/torque-sensor. *used on: pp. 45–47, 52, 68, 112, 114, 115, 117, 130, 131, 137*

G

GPU
: graphics processing unit. *used on: pp. 51, 99*

H

H
: hypothesis. *used on: pp. 7, 9–12, 14, 29, 61, 64, 67, 71, 73, 74, 108, 135*

handover
: object handover. *used on: pp. v, 1, 4–12, 14–41, 43–48, 51–53, 56, 58, 60–62, 64, 67, 68, 70–75, 77–82, 84–86, 88, 93–95, 97, 99–102, 105, 107, 108, 110–112, 114, 116, 117, 119–122, 124–126, 128–133, 135–141, 151, 152, 158, 165, 174–176*

HMD
: head-mounted display. *used on: pp. 60, 138*

HPhase
: handover phase. *used on: pp. 29–33, 35, 36, 39–41, 44, 45, 47, 48, 62, 64, 65, 69, 73, 77, 81, 84, 86, 100–103, 105, 108, 111–116, 122, 125, 126, 130–132, 135, 136, 139, 141, 173, 174*

HRI
human-robot interaction. *used on: pp. v, 3, 4, 11–13, 16–18, 23, 36, 39, 43, 46, 48, 53, 56, 60, 61, 64, 69, 83, 124, 125, 135, 136, 139–141, 175*

HTask
handover task. *used on: pp. 119–123, 125–128, 130–132, 165*

HType
handover type. *used on: pp. 25, 26, 62, 107, 175*

humanoid
humanoid robot. *used on: pp. v, 3, 9, 16–18, 20, 22, 24, 33, 38, 42, 44, 46, 47, 51, 52, 54, 57, 61, 75, 81, 82, 107, 114, 117, 135–137, 140, 141, 168, 173–176*

I

IIK
inverse instantaneous kinematics. *used on: pp. 81–84, 90, 136, 173, 175*

IK
inverse kinematics. *used on: pp. 20, 40–44, 81–85*

IMU
inertial measurement unit. *used on: p. 47*

J

JMM
joint motion model. *used on: pp. 44, 82, 139*

JTC
joint_trajectory_controller. *used on: pp. 53, 84–87*

L

LiDAR
light detection and ranging. *used on: pp. 51, 111*

M

Meka M1
Meka M1 Mobile Manipulator. *used on: pp. v, xiii, 3, 4, 49, 51–55, 59, 61–71, 73, 75, 76, 80, 82, 84, 87, 88, 90, 93, 97, 101, 107, 108, 139, 174, 176*

MR
mixed reality. *used on: pp. 56, 57, 59, 60, 137, 138*

mshs-robot
mobile social humanoid service robot. *used on: pp. 17, 48, 49, 51, 75, 107, 174, 176*

N

NARS

Negative Attitude toward Robots Scale. *used on: pp. 68, 119, 128, 131*

NVC

nonverbal communication. *used on: pp. v, 1, 4–9, 11–15, 17–19, 22, 24, 25, 27, 30, 33, 35, 43, 48, 60, 62, 75, 131, 133, 135, 140, 141, 176*

O

OTP

object transfer point. *used on: pp. 4, 7, 29, 30, 38–43, 49, 58, 62, 77, 78, 81, 86, 88, 94–96, 99, 102, 103, 105, 106, 112, 113, 136, 138, 139, 141, 173, 174, 176, 178*

$\mathrm{OTP}_{dynamic}$

dynamic object transfer point. *used on: pp. 39, 95, 96, 102–104, 125, 173, 175*

OTP_{fixed}

fixed object transfer point. *used on: pp. 39, 40, 62, 81, 174*

$\mathrm{OTP}_{integrated}$

integrated object transfer point. *used on: pp. 39, 95, 96, 103, 104, 108, 112, 175*

OTP_{static}

static object transfer point. *used on: pp. 39, 81, 95, 96, 102–104, 112, 125, 175, 178*

P

PC

personal computer. *used on: pp. 51, 53, 121*

R

RANSAC

RANdom SAmple Consensus. *used on: p. 38*

RDTK

[Research & Robotics] Development Toolkit. *used on: pp. vii, 54, 209*

RGB

red, green, blue color model. *used on: pp. 93, 95, 96, 99*

ROS

Robot Operating System. *used on: pp. 53, 57, 76, 120, 125, 177*

RPY

roll pitch yaw. *used on: p. 84*

RQ

research question. *used on: pp. 7–10, 12, 14, 141*

S

SEA
: series elastic actuator. *used on: pp. 40, 51–53, 90, 114, 115*

SLAM
: simultaneous localization and mapping. *used on: p. 52*

SOM
: self-organizing map. *used on: p. 94*

SR
: system requirement. *used on: pp. 7, 10–15, 19, 24, 29, 35, 39, 43, 45, 47, 49, 51, 54, 62, 64, 73–75, 77, 79–84, 91, 93–95, 101, 102, 105–108, 111, 112, 114, 116, 117, 122, 125, 128–132, 135–137, 141*

U

UML
: Unified Modeling Language. *used on: pp. 111, 113*

V

VR
: virtual reality. *used on: pp. 56, 58, 60, 138*

W

WBC
: Whole-Body Control. *used on: p. 139*

WEIRD
: Western, Educated, Industrialized, Rich, and Democratic. *used on: pp. 74, 137, 138*

GLOSSARY

A

adaptive trajectory (AT)
: Second part of the motion trajectory during the handover phase (HPhase) 2: Reach which is adapting to the current object transfer point (OTP) (cf. *mode2 motion* by Kajikawa and Ishikawa [KI00]). *defined on p. 82. used on: pp. 82, 84–86, 89–91, 112, 116, 132, 139, 154, 155, 167, 173*

B

base trajectory (BT)
: First part of the motion during the HPhase 2: Reach. It is similar to the *mode1 motion* as described by Kajikawa and Ishikawa [KI00]. *defined on p. 82. used on: pp. 82, 84–86, 88–91, 112, 114, 116, 126, 132, 136, 139, 167, 173*

C

Cognitive Service Robotics Apartment (CSRA)
: A smart environment for research of long-term human-technology interaction [Wre+17]. *defined on p. 55. used on: pp. vii, 55, 65, 120, 167, 173*

D

damped least squares (DLS)
: A possible solution to the inverse instantaneous kinematics (IIK) problem. *defined on p. 83. used on: pp. 83–86, 116, 139, 167, 173*

deictic gesture
: Pointing gesture to reference something. Also called *indexical* gesture. *defined on p. 19. used on: pp. 19, 131*

dynamic movement primitives (DMPs)
: A framework for motor control in humans and humanoid robots (humanoids) [Sch06]. *defined on p. 20. used on: pp. 20, 42, 43, 167, 173*

dynamic object transfer point ($\mathrm{OTP}_{dynamic}$)
: A current prediction of the OTP based on the latest data. *defined on p. 39. used on: pp. 39, 95, 96, 102–104, 125, 170, 173, 175*

E

end-effector (EEF)

A tool at the end of a robotic manipulator that is used to interact with the environment. Here, impactive gripper or hand of a robot. *defined on p. 16. used on: pp. 4, 5, 7, 11, 12, 16, 20, 25, 29, 31, 39–43, 45–47, 52, 53, 62, 64, 73, 77, 78, 80–84, 88, 101, 108, 112, 114–116, 125–127, 129, 130, 168, 174, 175, 177*

F

f-Formation

Systems of Arrangements in interaction [Ken76]. *defined on p. 32. used on: pp. 32, 33*

fixed object transfer point (OTP_{fixed})

An OTP that is determined before the handover and not updated during the whole interaction. *defined on p. 39. used on: pp. 39, 40, 62, 81, 170, 174*

Floka Head

A humanoid robot head which is capable of expressing gaze cues, seeing through his eyes, and emitting social signals. *defined on p. 75. used on: pp. v, 59, 75, 76, 79, 80, 107, 108, 174*

Floka humanoid (Floka)

A combination of the Floka Head and the Meka M1 Mobile Manipulator (Meka M1) which has all features of a mobile social humanoid service robot (mshs-robot). *defined on p. 75. used on: pp. v, 55, 57, 58, 75, 76, 80, 97, 107, 111, 115, 119–121, 123–131, 136, 139, 141, 168, 174*

G

gaze

An intended look at something by fixing the eyes on it. *defined on p. 21. used on: pp. 4, 6–8, 11, 12, 18, 21–23, 58, 60, 64, 75–80, 107–113, 115–117, 119, 125, 128, 131, 132, 136, 141, 174, 176*

gesture

A motion of the body or limbs that nonverbally conveys information. *defined on p. 19. used on: pp. v, 6–9, 11, 18–21, 24, 40, 61, 62, 64, 70, 71, 73, 81, 85, 91, 93, 94, 99, 112, 116, 131, 135–137, 141, 152, 173, 176, 178*

H

handover phase (HPhase)

An isolated part or sub-process during a handover interaction. *defined on p. 29. used on: pp. 29–33, 35, 36, 39–41, 44, 45, 47, 48, 62, 64, 65, 69, 73, 77, 81, 84, 86, 100–103, 105, 108, 111–116, 122, 125, 126, 130–132, 135, 136, 139, 141, 168, 173, 174*

handover type (HType)
: Classification of handovers into four distinct types as described in Table 3.1. *defined on p. 26. used on: pp. 25, 26, 62, 107, 169, 175*

humanoid robot (humanoid)
: A robot with human properties, like having two arms with human proportions, the ability to walk upright, or having a human-like face. *defined on p. 16. used on: pp. v, 3, 9, 16–18, 20, 22, 24, 33, 38, 42, 44, 46, 47, 51, 52, 54, 57, 61, 75, 81, 82, 107, 114, 117, 135–137, 140, 141, 168, 169, 173–176*

human-robot interaction (HRI)
: A multidisciplinary research field on human factors, robotics, cognitive psychology, and design [GS07]. *defined on p. 17. used on: pp. v, 3, 4, 11–13, 16–18, 23, 36, 39, 43, 46, 48, 53, 56, 60, 61, 64, 69, 83, 124, 125, 135, 136, 139–141, 169, 175*

I

in motion handover
: A handover that happens while at least one of the interactants is moving. In the literature two variants can be found. One includes locomotion like walking or running and the other describes it as gls handover where the object is transferred before the end-effector (EEF) stops moving. The latter is the definition used in this work. *defined on p. 25. used on: pp. 25, 28, 41, 46, 73, 78, 113–115, 117, 122, 129–132, 136, 138, 141*

integrated object transfer point ($\mathrm{OTP}_{\mathit{integrated}}$)
: Fusion of the static object transfer point ($\mathrm{OTP}_{\mathrm{static}}$) and the dynamic object transfer point ($\mathrm{OTP}_{\mathrm{dynamic}}$) for an early as well as precise prediction [NDL19]. *defined on p. 39. used on: pp. 39, 95, 96, 103, 104, 108, 112, 170, 175*

interaction space
: Overlap of two peripersonal spaces [Ken90]. *defined on p. 32. used on: pp. 32, 34, 90, 111*

inverse instantaneous kinematics (IIK)
: Calculating the changes in joint angles for a given EEF position and target [SK08, p. 30]. *defined on p. 82. used on: pp. 81–84, 90, 136, 169, 173, 175*

J

joint attention
: The more intense form of shared gaze, where two individuals focus on the same subject. Also called *shared attention*. The strongest form is triadic joint attention and a weaker form is called dyadic joint attention [OG04]. *defined on p. 21. used on: pp. 11, 13, 21, 22, 78, 111, 112*

L

locomotion

The action or power of a human, animal, cell, etc., of moving from one place or position to another unaided [OED:locom.]. *defined on p. 16. used on: pp. 16, 25, 30, 34, 35, 175, 176*

M

Meka M1 Mobile Manipulator (Meka M1)

The M1 Mobile Manipulator is a humanoid with a torso ([Meka:torso]), two arms ([Meka:arm]), two hands ([Meka:hand]), a mobile base ([Meka:base]) with the optional prismatic lift and computation backpack, and a movable sensor-head ([Meka:head]) [Meka:robot]. *defined on p. 51. used on: pp. v, 3, 4, 49, 51–55, 59, 61–71, 73, 75, 76, 80, 82, 84, 87, 88, 90, 93, 97, 101, 107, 108, 139, 169, 174, 176*

mobile robot

A robot being able to move around (see locomotion) using legs or wheels. *defined on p. 16. used on: pp. 14, 16, 17, 28, 34, 35, 38, 93, 137, 176*

mobile social humanoid service robot (mshs-robot)

A robot that has all the capabilities of mobile robots, social robots, service robots, and humanoids. *defined on p. 17. used on: pp. 17, 48, 49, 51, 75, 107, 169, 174, 176*

mutual gaze

Mutual gaze occurs when two people make eye contact or look into each other's eyes [Rog13]. *defined on p. 21. used on: pp. 21, 23, 24, 79, 108*

N

nonverbal communication (NVC)

Types of communication that do not use speech, like gaze and gestures. *defined on p. 18. used on: pp. v, 1, 4–9, 11–15, 17–19, 22, 24, 25, 27, 30, 33, 35, 43, 48, 60, 62, 75, 131, 133, 135, 140, 141, 170, 176*

O

object handover (handover)

Transfer of an object between two interaction partners. *defined on p. 24. used on: pp. v, 1, 4–12, 14–41, 43–48, 51–53, 56, 58, 60–62, 64, 67, 68, 70–75, 77–82, 84–86, 88, 93–95, 97, 99–102, 105, 107, 108, 110–112, 114, 116, 117, 119–122, 124–126, 128–133, 135–141, 151, 152, 158, 165, 168, 174–176*

object transfer point (OTP)

The point in space where control over an object is transferred. *defined on p. 38. used on: pp. 4, 7, 29, 30, 38–43, 49, 58, 62, 77,* *78, 81, 86, 88, 94–96, 99, 102, 103, 105, 106, 112, 113, 136, 138, 139, 141, 170, 173, 174, 176, 178*

P

peripersonal space
The space in reach with the limbs of an individual [Ken90]. *defined on p. 32. used on: pp. 32, 175*

power grasp
Grasping something by clamping it between partly flexed fingers and the palm instead of using only fingers or fingertips (cf. precision grasp). The result is also called *power grip* [Nap56]. *defined on p. 32. used on: pp. 32, 53, 62, 115*

precision grasp
Clamping an object between fingers and thumb. The result is also called *precision grip* [Nap56]. *defined on p. 32. used on: pp. 32, 53, 177*

proxemics
A concept of interpersonal distance by Hall et al. with four distinct categories as describe in Table 3.3 [Hal+68; Hal69]. *defined on p. 33. used on: pp. 18, 33, 34, 79, 178*

R

referential gaze
Gazing at a target to create a reference in communication. *defined on p. 21. used on: pp. 11, 21*

robot
A term originally coined by Čapek for artificial humans. Today, used for a machine that can act on the environment with multiple actuated joints which move an EEF to solve programmable tasks [OED:robot; OLD:robot]. *defined on p. 15. used on: pp. v, 1, 3–29, 31–48, 51–62, 64, 65, 67–82, 84, 86–90, 93–97, 100–102, 105, 107, 108, 111–117, 119–133, 135–141, 169, 173–178*

robot companion
A useful robot that also behaves socially [Dau07]. *defined on p. 16. used on: pp. 8, 16, 33, 51*

Robot Operating System (ROS)
A collection of libraries and tools for software developers to create applications for robots, including middleware for message passing, build tools and visualization [Qui+09]. *defined on p. 53. used on: pp. 53, 57, 76, 120, 125, 170, 177*

S

service robot
A robot that can sense and act on the environment. *defined on p. 16. used on: pp. v, 4, 5, 8, 16, 17, 74, 176*

shared gaze
Two interactants look at the same location or object. *defined on p. 21. used on: pp. 21, 23, 24, 175*

signaling
"Intentional modification of one's own action plan (e.g., a plan for reaching a glass of wine) to make it more predictable" [DDP18]. *defined on p. 18. used on: pp. 18, 64, 81, 112, 115, 116*

social robot
A robot that can emit social signals. *defined on p. 16. used on: pp. 16, 17, 34, 176*

social signal
A kind of communication which sends direct or indirect information on social interactions like emotions, attitudes or relationships [PD10]. *defined on p. 18. used on: pp. v, 8, 9, 11, 12, 16–18, 20, 174, 178*

static object transfer point (OTP_{static})
An initial prediction of an OTP. *defined on p. 39. used on: pp. 39, 81, 95, 96, 102–104, 112, 125, 170, 175, 178*

symbolic gesture
A replacement of words, depending on culture and context. Also called *emblematic*. *defined on p. 19. used on: p. 19*

T

turn-taking
Mostly used to describe the process of switching the speaker in conversations. In this context it is used to describe who has the turn during the handover process. *defined on p. 21. used on: pp. 21–23, 64*

V

vis-à-vis
Two persons standing face to face [CK80]. Also *H-formation* from proxemics theory. *defined on p. 33. used on: pp. 20, 33, 35, 62, 111*

BIBLIOGRAPHY

OWN PUBLICATIONS

[Sim+19] Janneke Simmering, Sebastian Meyer zu Borgsen, Sven Wachsmuth, and Ayoub Al-Hamadi. "Combining Static and Dynamic Predictions of Transfer Points for Human Initiated Handovers." In: *ICSR 2019: International Conference on Social Robotics*. Ed. by John-John Cabibihan, Emilia Barakova, and Alan Wagner. Madrid, Spain: Springer, Cham, 2019-11, pp. 676–686. DOI: 10.1007/978-3-030-35888-4_63. *used on: pp. 93, 102, 104, 105*

[SMW19] Simon Schulz, Sebastian Meyer zu Borgsen, and Sven Wachsmuth. "See and Be Seen – Rapid and Likeable High-Definition Camera-Eye for Anthropomorphic Robots." In: *IEEE International Conference on Robotics and Automation*. Montreal, Canada, 2019-06. DOI: 10.1109/ICRA.2019.8794319. *used on: pp. 75, 76*

[Neu+18] Nils Neumann, Sebastian Meyer zu Borgsen, Phillip Lücking, and Sven Wachsmuth. "An Automated Pick-and-Place Benchmarking System in Robotics." In: *2018 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. Exc 277. IEEE, 2018-04, pp. 243–249. DOI: 10.1109/ICARSC.2018.8374190.

[Mey+18] Sebastian Meyer zu Borgsen, Patrick Renner, Florian Lier, Thies Pfeiffer, and Sven Wachsmuth. "Improving Human-Robot Handover Research by Mixed Reality Techniques." In: *VAM-HRI 2018. The Inaugural International Workshop on Virtual, Augmented and Mixed Reality for Human-Robot Interaction*. Chicago, USA, 2018-03. DOI: 10.4119/unibi/2919957. *used on: pp. 56, 57, 59*

[MBW17] Sebastian Meyer zu Borgsen, Jasmin Bernotat, and Sven Wachsmuth. "Hand in Hand with Robots: Differences Between Experienced and Naive Users in Human-Robot Handover Scenarios." In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, Cham, 2017-11, pp. 587–596. DOI: 10.1007/978-3-319-70022-9_58. *used on: pp. 61, 63, 65, 66, 69, 70*

[Mey+17] Sebastian Meyer zu Borgsen, Timo Korthals, Florian Lier, and Sven Wachsmuth. "ToBI – Team of Bielefeld: Enhancing Robot Behaviors and the Role of Multi-Robotics in RoboCup@Home." In: *RoboCup 2016: Robot World Cup XX. Behnke S, Raymond S, Sariel S, Lee DD (Eds); Lecture Notes on Artificial Intelligence, 9776*. Ed. by Sven Behnke, Sheh Raymond, Sanem Sariel, and Daniel D. Lee. Vol. 9776 LNAI.

Springer, Cham, 2017-11, pp. 577–588. ISBN: 978-3-319-68791-9. DOI: 10.1007/978-3-319-68792-6_48. *used on: p. 5*

[Ber+16] Jasmin Bernotat, Birte Schiffhauer, Friederike Eyssel, Patrick Holthaus, Christian Leichsenring, Viktor Richter, Marian Pohling, Birte Carlmeyer, Norman Köster, Sebastian Meyer zu Borgsen, et al. "Welcome to the Future – How Naïve Users Intuitively Address an Intelligent Robotics Apartment." In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2016-10, pp. 982–992. DOI: 10.1007/978-3-319-47437-3_96. arXiv: arXiv:1011.1669v3. *used on: p. 55*

[Ric+16] Viktor Richter, Birte Carlmeyer, Florian Lier, Sebastian Meyer zu Borgsen, David Schlangen, Franz Kummert, Sven Wachsmuth, and Britta Wrede. "Are you talking to me? Improving the robustness of dialogue systems in a multi party HRI scenario by incorporating gaze direction and lip movement of attendees." In: *HAI 2016 - Proceedings of the 4th International Conference on Human Agent Interaction*. 2016-10. DOI: 10.1145/2974804.2974823. *used on: p. 22*

[WWM16] Johannes Wienke, Sebastian Wrede, and Sebastian Meyer zu Borgsen. "A Data Set for Fault Detection Research on Component-Based Robotic Systems." In: *17th Towards Autonomous Robotic Systems Conference (TAROS-16)*. Exc 277. 2016-06. DOI: 10.1007/978-3-319-40379-3_35.

[Hol+16] Patrick Holthaus, Christian Leichsenring, Jasmin Bernotat, Viktor Richter, Marian Pohling, Birte Carlmeyer, Norman Köster, Sebastian Meyer zu Borgsen, René René Zorn, Birte Schiffhauer, et al. "How to Address Smart Homes with a Social Robot? A Multi-modal Corpus of User Interactions with an Intelligent Environment." In: *10th Language Resources and Evaluation Conference*. European Language Resources Association (ELRA), 2016-05, pp. 3440–3446. *used on: p. 55*

[Mey14] Sebastian Meyer zu Borgsen. *Incorporating Articulated Scene Models in Movement Strategies for a Robot Companion*. M.Sc. Thesis. Bielefeld, Germany, 2014-08.

[Mey+14] Sebastian Meyer zu Borgsen, Matthias Schöpfer, Leon Ziegler, and Sven Wachsmuth. "Automated Door Detection with a 3D-Sensor." In: *Proceedings - Conference on Computer and Robot Vision, CRV 2014*. 2014-05, pp. 276–282. DOI: 10.1109/CRV.2014.44.

ROBOCUP TOBI INVOLVEMENT

[WLM18] Sven Wachsmuth, Florian Lier, and Sebastian Meyer zu Borgsen. *ToBI - Team of Bielefeld: The Human-Robot Interaction System for RoboCup@Home 2018*. Tech. rep. Montreal, Canada: Bielefeld University, 2018. *used on: pp. 5, 61*

[Wac+17] Sven Wachsmuth, Florian Lier, Sebastian Meyer zu Borgsen, Johannes Kummert, Luca Lach, and Dominik Sixt. *ToBI - Team of Bielefeld: The Human-Robot Interaction System for RoboCup@Home 2017*. Tech. rep. Nagoya, Japan: Bielefeld University, 2017. *used on: pp. 5, 61*

[MKW16] Sebastian Meyer zu Borgsen, Timo Korthals, and Sven Wachsmuth. *ToBI - Team of Bielefeld: The Human-Robot Interaction System for RoboCup@Home 2016*. Tech. rep. Leipzig, Germany: Bielefeld University, 2016. *used on: pp. 5, 61*

[Zie+15] Leon Ziegler, Jens Wittrowski, Sebastian Meyer zu Borgsen, Sven Wachsmuth, Timo Korthals, Leon Ziegler, and Sven Wachsmuth. *ToBI - Team of Bielefeld: The Human-Robot Interaction System for RoboCup@Home 2015*. Tech. rep. Hefei, China: Bielefeld University, 2015. *used on: pp. 5, 61*

[Zie+14] Leon Ziegler, Jens Wittrowski, Sebastian Meyer zu Borgsen, and Sven Wachsmuth. *ToBI - Team of Bielefeld: The Human-Robot Interaction System for RoboCup@Home 2014*. Tech. rep. João Pessoa, Brazil: Bielefeld University, 2014. *used on: pp. 5, 61*

[Zie+13] Leon Ziegler, Jens Wittrowski, Matthias Schöpfer, Frederic Siepmann, and Sven Wachsmuth. *ToBI - Team of Bielefeld: The Human-Robot Interaction System for RoboCup@Home 2013*. Tech. rep. Eindhoven, Netherlands, 2013. *used on: pp. 5, 61*

[Wac+12] Sven Wachsmuth, Frederic Siepmann, Leon Ziegler, and Florian Lier. *ToBI - Team of Bielefeld: The Human-Robot Interaction System for RoboCup@Home 2012*. Tech. rep. Mexico City, Mexico: Bielefeld University, 2012. *used on: p. 61*

GENERAL REFERENCES

[Adm+14] Henny Admoni, Anca Dragan, Siddhartha S Srinivasa, and Brian Scassellati. "Deliberate Delays During Robot-to-Human Handovers Improve Compliance With Gaze Communication Henny." In: *International Conference on Human-Robot Interaction (HRI)*. 2014-03. DOI: `10.1145/2559636.2559682`. *used on: pp. 23, 78*

[AT97] A. Agah and K. Tanie. "Human Interaction with a Service Robot: Mobile-Manipulator Handing Over an Object to a Human." In: *Proceedings of the 1997 IEEE International Conference on Robotics and Automation* 1.April (1997), pp. 575–580. ISSN: 10504729. DOI: `10.1109/ROBOT.1997.620098`. *used on: pp. 25, 42*

[AMC14] Jacopo Aleotti, Vincenzo Micelli, and Stefano Caselli. "An Affordance Sensitive System for Robot to Human Object Handover." In: *International Journal of Social Robotics* 6.4 (2014-11), pp. 653–666. ISSN: 18754805. DOI: `10.1007/s12369-014-0241-3`. *used on: p. 31*

[AMC12] Jacopo Aleotti, Vincenzo Micelli, and Stefano Caselli. "Comfortable Robot to Human Object Hand-Over." In: *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. Paris, France, 2012-11, pp. 771–776. DOI: 10.1109/ROMAN.2012.6343845. *used on: pp. 31, 36*

[AI72] Michael Argyle and Roger Ingham. "Gaze, Mutual Gaze, and Proximity." In: *Semiotica* 6.1 (1972). *used on: p. 21*

[BDS84] A. Balestrino, G. De Maria, and L. Sciavicco. "Robust Control of Robotic Manipulators." In: *IFAC Proceedings Volumes* 17.2 (1984). DOI: 10.1016/S1474-6670(17)61347-8. *used on: p. 83*

[Bal+18] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis Philippe Morency. "OpenFace 2.0: Facial behavior analysis toolkit." In: *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018* (2018), pp. 59–66. DOI: 10.1109/FG.2018.00019. *used on: p. 209*

[Bar11] Christoph Bartneck. "The end of the beginning : a reflection on the first five years of the HRI conference." In: *Scientometrics* (2011), pp. 487–504. DOI: 10.1007/s11192-010-0281-x. *used on: p. 17*

[Bar+07] Christoph Bartneck, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. "Is the uncanny valley an uncanny cliff?" In: *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication* May 2014 (2007), pp. 368–373. ISSN: 1944-9445. DOI: 10.1109/ROMAN.2007.4415111. *used on: p. 5*

[Bar+09] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. *Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots*. 2009. DOI: 10.1007/s12369-008-0001-3. URL: http://link.springer.com/10.1007/s12369-008-0001-3. *used on: p. 67*

[Bas+09] Patrizia Basili, Markus Huber, Thomas Brandt, Sandra Hirche, and Stefan Glasauer. "Investigating Human-Human Approach and Hand-Over." In: *Human Centered Robot Systems: Cognition, Interaction, Technology*. Ed. by Helge Ritter, Gerhard Sagerer, Rüdiger Dillmann, and Martin Buss. Vol. 6. Berlin, Heidelberg: Springer, Berlin, Heidelberg, 2009, pp. 151–160. ISBN: 978-3-642-10403-9. DOI: 10.1007/978-3-642-10403-9_16. *used on: pp. 28, 34, 36, 38, 39, 102, 103*

[BSW13] Mohamad Bdiwi, Jozef Suchy, and Alexander Winkler. "Handing-over model-free objects to human hand with the help of vision/force robot control." In: *10th International Multi-Conferences on Systems, Signals & Devices 2013 (SSD13)*. IEEE, 2013, pp. 1–6. DOI: 10.1109/SSD.2013.6564138. *used on: pp. 25, 45*

[BMR19] Holger H. Bekemeier, Jonathan W. Maycock, and Helge J. Ritter. "What Does a Hand-Over Tell?—Individuality of Short Motion Sequences." In: *Biomimetics* 4.3 (2019), p. 55. ISSN: 2313-7673. DOI: 10.3390/biomimetics4030055. *used on: p. 39*

[BE17] Jasmin Bernotat and Friederike Eyssel. "An Evaluation Study of Robot Designs for Smart Environments." In: *HRI '17 Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. March 2017. 2017, pp. 87–88. DOI: 10.1145/3029798.3038429. *used on: pp. 58, 75*

[BK06] Rainer Bischoff and Johannes Kurth. "Concepts, Tools and Devices for Facilitating Human-Robot Interaction with Industrial Robots through Augmented Reality." In: *ISMAR Workshop on Industrial Augmented Reality*. Santa Barbara, CA, 2006, pp. 1–35. *used on: p. 57*

[Bou+12] Jean David Boucher, Ugo Pattacini, Amelie Lelong, Gerard Bailly, Frederic Elisei, Sascha Fagel, Peter Ford Dominey, and Jocelyne Ventre-Dominey. "I reach faster when I see you look: Gaze effects in human-human and human-robot face-to-face cooperation." In: *Frontiers in Neurorobotics* 6 (2012). ISSN: 16625218. DOI: 10.3389/fnbot.2012.00003. *used on: p. 22*

[BKV04] Matthieu Bray, Esther Koller-Meier, and Luc Van Gool. "Smart particle filtering for 3D hand tracking." In: *Proceedings - Sixth IEEE International Conference on Automatic Face and Gesture Recognition*. 2004, pp. 675–680. DOI: 10.1109/AFGR.2004.1301612. *used on: p. 94*

[Bri04] Robert Bringhurst. *The Elements of Typgraphic Style*. 3rd. Vol. 3. Hartley & Marks Point Roberts, WA, 2004. ISBN: 978-0-88179-205-8. *used on: p. 213*

[BG88a] Daniel Bullock and Stephen Grossberg. "Neural Dynamics of Planned Arm Movements: Emergent Invariants and Speed-Accuracy Properties During Trajectory Formation." In: *Psychological Review* 95.1 (1988), pp. 49–90. ISSN: 0033295X. DOI: 10.1037/0033-295X.95.1.49. *used on: p. 44*

[BG88b] Daniel Bullock and Stephen Grossberg. "The VITE Model: A Neural Command Circuit for Generating Arm and Articulator Trajectories." In: *Dynamic patterns in complex systems* (1988). ISSN: 0007-1250. DOI: 10.1192/bjp.112.483.211-a. *used on: p. 44*

[Bus09] Samuel R. Buss. *Introduction to Inverse Kinematics with Jacobian Transpose, Pseudoinverse and Damped Least Squares methods*. Tech. rep. San Diego, CA, USA: University of California, 2009. *used on: pp. 83, 139*

[But+30] Stephen Butterworth et al. "On the Theory of Filter Amplifiers." In: *Experimental Wireless and the Wireless Engineer* 7.6 (1930), pp. 536–541. *used on: p. 115*

[Cak+11a] Maya Cakmak, Siddhartha S. Srinivasa, Min Kyung Lee, Jodi Forlizzi, and Sara Kiesler. "Human Preferences for Robot-Human Hand-over Configuration." In: *IEEE International Conference on Intelligent Robots and Systems*. San Francisco, CA, USA, 2011, pp. 1986–1993. DOI: 10.1109/IROS.2011.6048340. *used on: pp. 27, 31, 36, 41*

[Cak+11b] Maya Cakmak, Siddhartha S Srinivasa, Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. "Using Spatial and Temporal Contrast for Fluent Robot-Human Hand-overs." In: *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Lausanne, Switzerland, 2011. DOI: 10.1145/1957656.1957823. *used on: p. 41*

[Cao+17] Zhe Cao, Tomas Simon, Shih En Wei, and Yaser Sheikh. "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields." In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* XXX (2017), pp. 1302–1310. DOI: 10.1109/CVPR.2017.143. arXiv: arXiv:1812.08008v2. *used on: p. 209*

[Čap20] Karel Čapek. *R. U. R. - Rossum's Universal Robots*. Ed. by Paul Selver and Nigel Playfair. 1920. *used on: pp. 3, 15, 177*

[Car+19] Alessandro Carfì, Francesco Foglino, Barbara Bruno, and Fulvio Mastrogiovanni. "A multi-sensor dataset of human-human handover." In: *Data in Brief* 22 (2019), pp. 109–117. ISSN: 23523409. DOI: 10.1016/j.dib.2018.11.110. *used on: p. 26*

[Cha+14] Wesley P. Chan, Iori Kumagai, Shunichi Nozawa, Yohei Kakiuchi, Kei Okada, and Masayuki Inaba. "Implementation of a Robot-Human Object Handover Controller on a Compliant Underactuated Hand Using Joint Position Error Measurements for Grip Force and Load Force Estimations." In: *Proceedings - IEEE International Conference on Robotics and Automation*. Hong Kong, China: ieee, 2014, pp. 1190–1195. DOI: 10.1109/ICRA.2014.6907004. *used on: pp. 37, 46*

[Cha+15] Wesley P. Chan, Matthew K X J Pan, Elizabeth A. Croft, and Masayuki Inaba. "Characterization of Handover Orientations used by Humans for Efficient Robot to Human Handovers." In: *IEEE International Conference on Intelligent Robots and Systems*. 2015, pp. 1–6. DOI: 10.1109/IROS.2015.7353106. *used on: p. 36*

[Cha+19] Wesley P. Chan, Matthew K.X.J. Pan, Elizabeth A. Croft, and Masayuki Inaba. "An Affordance and Distance Minimization Based Method for Computing Object Orientations for Robot Human Handovers." In: *International Journal of Social Robotics* (2019). ISSN: 18754805. DOI: 10.1007/s12369-019-00546-7. *used on: p. 38*

[Cha+12] Wesley P. Chan, Chris A.C. Parker, H.F. Machiel Van der Loos, and Elizabeth A. Croft. "Grip Forces and Load Forces in Handovers: Implications for Designing Human-Robot Handover Controllers." In: *7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Boston, Massachusetts, USA, 2012-03. doi: 10.1145/2157689.2157692\. *used on: pp. 45, 46*

[Cha+13] Wesley P Chan, Chris A.C. C Parker, Hf F Machiel Van Der Loos, and Elizabeth a Croft. "A human-inspired object handover controller." In: *International Journal of Robotics Research* 32.8 (2013), pp. 971–983. doi: 10.1177/0278364913488806. *used on: pp. 4, 45–47*

[Che+11] Chia Ping Chen, Yu Ting Chen, Ping Han Lee, Yu Pao Tsai, and Shawmin Lei. "Real-time Hand Tracking on Depth Images." In: *2011 IEEE Visual Communications and Image Processing, VCIP 2011*. 1. 2011. doi: 10.1109/VCIP.2011.6115983. *used on: pp. 99, 100*

[Chi+17] Sachin Chitta, Eitan Marder-Eppstein, Wim Meeussen, Vijay Pradeep, Adolfo Rodríguez Tsouroukdissian, Jonathan Bohren, David Coleman, Bence Magyar, Gennaro Raiola, Mathias Lüdtke, and Enrique Fernandez Perdomo. "ros_-control: A generic and simple control framework for ROS." In: *The Journal of Open Source Software* 2.20 (2017), p. 456. doi: 10.21105/joss.00456. *used on: p. 208*

[Cin+19] F. Cini, V. Ortenzi, P. Corke, and M. Controzzi. "On the choice of grasp type and location when handing over an object." In: *Science Robotics* 4.27 (2019). issn: 2470-9476. doi: 10.1126/scirobotics.aau9757. *used on: p. 32*

[CK80] T. Matthew Ciolek and Adam Kendon. "Environment and the Spatial Arrangement of Conversational Encounters." In: *Sociological Inquiry* 50.3-4 (1980), pp. 237–271. issn: 1475682X. doi: 10.1111/j.1475-682X.1980.tb00022.x. *used on: pp. 32, 33, 178*

[Cla96] Herbert H. Clark. *Using language*. Cambridge : Cambridge University Press, 1996. *used on: p. 29*

[Coa09] Gordon Coates. "Notes on Communication: A few thoughts about the way we interact with the people we meet." In: (2009), p. 180. *used on: p. 18*

[CR00] Dorin Comaniciu and Visvanathan Ramesh. "Real-Time Tracking of Non-Rigid Objects using Mean Shift." In: *Computer Vision and Pattern Recognition, IEEE Conference on. Vol 2*. 7. 2000, pp. 142–149. *used on: p. 93*

[Con+18] Marco Controzzi, Harmeet Singh, Francesca Cini, Torquato Cecchini, Alan Wing, and Christian Cipriani. "Humans adjust their grip force when passing an object according to the observed speed of the partner's reaching out movement." In: *Experimental Brain Research* 236.12 (2018), pp. 3363–3377. issn: 14321106. doi: 10.1007/s00221-018-5381-5. *used on: p. 46*

[Coo77] Mark Cook. "Gaze and Mutual Gaze in Social Encounters: How long—and when—we look others "in the eye"." In: *American Scientist* 65.3 (1977), pp. 328–333. *used on: p. 21*

[Dar72] Charles Darwin. "The expression of emotions in Man and Animals." In: (1872), p. 399. ISSN: 0002-9629. DOI: 10.1037/10001-000. *used on: p. 4*

[Dau07] Kerstin Dautenhahn. "Socially intelligent robots: dimensions of human-robot interaction." In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 362.1480 (2007), pp. 679–704. DOI: 10.1098/rstb.2006.2004. *used on: pp. 16, 177*

[Deh+11] Frédéric Dehais, Emrah Akin Sisbot, Rachid Alami, and Mickaël Causse. "Physiological and subjective evaluation of a human-robot object hand-over task." In: *Applied Ergonomics* 42.6 (2011), pp. 785–791. DOI: 10.1016/j.apergo.2010.12.005. *used on: p. 41*

[Don+15] Christian Dondrup, Nicola Bellotto, Ferdian Jovan, and Marc Hanheide. "Real-time multisensor people tracking for human-robot spatial interaction." In: *Workshop on Machine Learning for Social Robotics at International Conference on Robotics and Automation (ICRA)* (2015). *used on: p. 209*

[DDP18] Francesco Donnarumma, Haris Dindo, and Giovanni Pezzulo. "Sensorimotor Communication for Humans and Robots: Improving Interactive Skills by Sending Coordination Signals." In: *IEEE Transactions on Cognitive and Developmental Systems* 10.4 (2018), pp. 903–917. ISSN: 23798939. DOI: 10.1109/TCDS.2017.2756107. *used on: pp. 18, 178*

[DB09] Michael Donoser and Horst Bischof. "Real Time Appearance Based Hand Tracking." In: *19th International Conference on Pattern Recognition*. 2009, pp. 1–4. DOI: 10.1109/icpr.2008.4761485. *used on: p. 93*

[DM03] Lorenzo Dozio and Paolo Mantegazza. "Linux Real Time Application Interface (RTAI) in low cost high performance motion control." In: *Motion Control*. National Italian Association for Automation, 2003, pp. 27–28. *used on: p. 209*

[Dra+15] Anca D Dragan, Shira Bauman, Jodi Forlizzi, and Siddhartha S Srinivasa. "Effects of Robot Motion on Human-Robot Collaboration." In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*. 2015, pp. 51–58. DOI: 10.1145/2696454.2696473. *used on: p. 40*

[DHO06] Mauro Dragone, Thomas Holz, and Gregory M.P. O'Hare. "Mixing robotic realities." In: *Proceedings of the 2006 International Conference on Intelligent User Interfaces*. 2006. Sydney, Australia: ACM Press, 2006, p. 261. DOI: 10.1145/1111449.1111504. *used on: p. 57*

[E D12] E. DOLD & SÖHNE KG. *SAFEMASTER W Wireless Safety System Remote Control for E-Stop RE 5910*. Tech. rep. Furtwangen, 2012. *used on: p. 53*

[EK07] Aaron Edsinger and Charles C. Kemp. "Human-Robot Interaction for Cooperative Manipulation: Handing Objects to One Another." In: *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. 2007, pp. 1167–1172. DOI: 10.1109/ROMAN.2007.4415256. *used on: pp. 37, 40, 41*

[EB16] Kristian Ehlers and Konstantin Brama. "A human-robot interaction interface for mobile and stationary robots based on real-time 3D human body and hand-finger pose estimation." In: *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*. 2016. DOI: 10.1109/ETFA.2016.7733719. *used on: p. 94*

[Ekm97] Paul Ekman. "Should we call it expression or communication?" In: *Innovation: The European Journal of Social Science Research* 10.4 (1997). DOI: 10.1080/13511610.1997.9968538. *used on: p. 18*

[EM16] Hanna Engstrand and Janet Mattsson. "The non-verbal communication in handover situations are the spice between the lines, to understand the severity of the patient's condition." In: *Journal of Nursing Education and Practice* 7.5 (2016). ISSN: 1925-4040. DOI: 10.5430/jnep.v7n5p1. *used on: p. 24*

[Far+02] Teresa Farroni, Gergely Csibra, Francesca Simion, and Mark H. Johnson. "Eye contact detection in humans from birth." In: *Proceedings of the National Academy of Sciences of the United States of America* 99.14 (2002), pp. 9602–9605. ISSN: 00278424. DOI: 10.1073/pnas.152159999. *used on: p. 21*

[Fei+16] Thomas Feix, Javier Romero, Heinz Bodo Schmiedmayer, Aaron M. Dollar, and Danica Kragic. "The GRASP Taxonomy of Human Grasp Types." In: *IEEE Transactions on Human-Machine Systems* 46.1 (2016), pp. 66–77. ISSN: 21682291. DOI: 10.1109/THMS.2015.2470657. *used on: p. 32*

[FH85] Tamar Flash and Neville Hogan. "The Coordination of Arm Movements: An Experimentally Confirmed Mathematical Model." In: *The Journal of Neuroscience* 5.7 (1985), pp. 1688–1703. DOI: 10.1523/JNEUROSCI.05-07-01688.1985. *used on: pp. 95, 102*

[FS16] Chien-Liang Fok and Luis Sentis. "Integration and Usage of a ROS-based Whole Body Control Software Framework." In: *Studies in Computational Intelligence* 625 (2016), pp. 535–563. ISSN: 1860949X. DOI: 10.1007/978-3-319-26054-9_21. *used on: p. 139*

[Fon+06] Terrence Fong, Moffett Field, Clayton Kunz, Laura M Hiatt, and Magda Bugajska. "The Human-Robot Interaction Operating System." In: *Human-Robot Interaction*. Salt Lake City, UT USA, 2006. *used on: p. 17*

[Foo13] Tully Foote. "Tf: The transform library." In: *IEEE Conference on Technologies for Practical Robot Applications, TePRA*. 2013. DOI: 10.1109/TePRA.2013.6556373. *used on: p. 209*

[Fra+12] Richard M. Frankel, Mindy Flanagan, Patricia Ebright, Alicia Bergman, Colleen M. O'Brien, Zamal Franks, Andrew Allen, Angela Harris, and Jason J. Saleem. "Context, cultureand (non-verbal) communication affect handover quality." In: *BMJ Quality and Safety* 21.SUPPL. 1 (2012), pp. 121–128. ISSN: 20445415. DOI: 10.1136/bmjqs-2012-001482. *used on: p. 24*

[Gha+13] Mamoun Gharbi, Séverin Severin Lemaignan, Jim Mainprice, and Rachid Alami. "Natural Interaction for Object Hand-Over." In: *ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, 2013-03, p. 401. DOI: 10.1109/HRI.2013.6483617. *used on: p. 35*

[Gha+15] Mamoun Gharbi, Pierre Vincent Paubel, Aurelie Clodic, Ophelie Carreras, Rachid Alami, and Jean Marie Cellier. "Toward a better understanding of the communication cues involved in a human-robot object transfer." In: *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. 2015, pp. 319–324. DOI: 10.1109/ROMAN.2015.7333626. *used on: pp. 23, 24*

[Gla+10] S. Glasauer, M. Huber, P. Basili, A. Knoll, and T. Brandt. "Interacting in time and space: Investigating human-human and human-robot joint action." In: *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. 2010, pp. 252–257. DOI: 10.1109/ROMAN.2010.5598638. *used on: pp. 28, 38*

[GS07] Michael A Goodrich and Alan C Schultz. "Human – Robot Interaction : A Survey. Foundations and Trends in Human-Computer Interaction." In: 1.3 (2007). DOI: 10.1561/1100000005. *used on: pp. 3, 17, 18, 175*

[Gri+13] Elena Corina Grigore, Kerstin Eder, Anthony G. Pipe, Chris Melhuish, and Ute Leonards. "Joint action understanding improves robot-to-human object handover." In: *IEEE International Conference on Intelligent Robots and Systems*. 2013, pp. 4622–4629. DOI: 10.1109/IROS.2013.6697021. *used on: pp. 4, 22*

[HMB11] Georg Hackenberg, Rod McCall, and Wolfgang Broll. "Lightweight Palm and Finger Tracking for Real-Time 3D Gesture Control." In: *Proceedings - IEEE Virtual Reality*. March. 2011, pp. 19–26. DOI: 10.1109/VR.2011.5759431. *used on: p. 94*

[Hal+68] Edward T. Hall, Ray L. Birdwhistell, Bernhard Bock, Paul Bohannan, A. Richard Diebold, Marshall Durbin, Munro S. Edmonson, J. L. Fischer, Dell Hymes, Solon T. Kimball, et al. "Proxemics [and Comments and Replies]." In: *Current Anthropology* 9.2/3 (1968-04), pp. 83–108. ISSN: 0011-3204. DOI: 10.1086/200975. *used on: pp. 33, 177*

[Hal69] Edward Twitchell Hall. *The Hidden Dimension: Man's Use of Space in Public and Private*. Bodley Head, 1969. ISBN: 978-0-370-01308-4. *used on: pp. 33, 177*

[HY18] Zhao Han and Holly Yanco. "Investigating Human-Robot Handover Release Behaviors." In: *New England Manipulation Symposium*. New England, 2018. *used on: p. 46*

[HY19] Zhao Han and Holly Yanco. "The Effects of Proactive Release Behaviors during Human-Robot Handovers." In: *ACM/IEEE International Conference on Human-Robot Interaction* 2019-March (2019), pp. 440–448. ISSN: 21672148. DOI: 10.1109/HRI.2019.8673085. *used on: p. 46*

[Han+17] Clint Hansen, Paula Arambel, Khalil Ben Mansour, Véronique Perdereau, and Frédéric Marin. "Human–Human Handover Tasks and How Distance and Object Mass Matter." In: *Perceptual and Motor Skills* 124.1 (2017), pp. 182–199. ISSN: 1558688X. DOI: 10.1177/0031512516682668. *used on: pp. 36, 94, 95*

[Har+15] Justin W Hart, Sara Sheikholeslami, Elizabeth A Croft, Wesley Patrick Chan, and Matthew K X J Pan. "Predictions of Human Task Performance and Handover Trajectories for Human-Robot Interaction." In: *HRI Workshop on Timing in Human-Robot Teaming*. 2015, p. 4. *used on: pp. 94, 95*

[HS15] Wuwei He and Daniel Sidobre. "Improving Human-Robot Object Exchange by Online Force Classification." In: *Journal of Human-Robot Interaction* 4.1 (2015), p. 75. ISSN: 2163-0364. DOI: 10.5898/JHRI.4.1.He. *used on: pp. 25, 45*

[Hen+14] Norman Hendrich, Hannes Bistry, Johannes Liebrecht, and Jianwei Zhang. "Natural Robot-Human Handover Combining Force and Tactile Sensors." In: *Workshop on Assistance and Service Robotics in a Human Environment, IEEE/RSJ IROS*. Chicago, USA, 2014, pp. 1–7. *used on: pp. 25, 27, 28, 42, 45, 46*

[Hen+16] Norman Hendrich, Johannes Liebrecht, Hannes Bistry, and Jianwei Zhang. *Optimal Interaction Forces for Natural Robot-Human Handover*. Tech. rep. 2016, pp. 4–5. *used on: pp. 25, 28, 45, 46*

[HHN10] Joseph Henrich, Steven J. Heine, and Ara Norenzayan. "The weirdest people in the world?" In: *Behavioral and Brain Sciences* 33.2-3 (2010-06), pp. 61–83. ISSN: 0140-525X. DOI: 10.1017/S0140525X0999152X. *used on: pp. 74, 137*

[HB08] Micha Hersch and Aude G. Billard. "Reaching with multi-referential dynamical systems." In: *Autonomous Robots* 25.1-2 (2008), pp. 71–83. ISSN: 09295593. DOI: 10.1007/s10514-007-9070-7. *used on: p. 44*

[Hic+07] Halim Hicheur, Quang Cuong Pham, Gustavo Arechavaleta, Jean Paul Laumond, and Alain Berthoz. "The formation of trajectories during goal-oriented locomotion in humans. I. A stereotyped behaviour." In: *European Journal of Neuroscience* 26.8 (2007), pp. 2376–2390. ISSN: 0953816X. DOI: 10.1111/j.1460-9568.2007.05836.x. *used on: p. 34*

[Hof13] Guy Hoffman. "Evaluating Fluency in Human-Robot Collaboration." In: *Robotics: Science and Systems, Workshop on Human Robot Collaboration*. 2013, pp. 1–8. DOI: 10.1021/ja035318g. *used on: p. 18*

[Hol14] Patrick Holthaus. "Approaching Human-Like Spatial Awareness in Social Robotics." Doctoral dissertation. Bielefeld University, 2014. *used on: pp. 32, 111*

[HS12] Dirk Holz and Jesus Savage. *RoboCup @ Home Rulebook 2012*. Tech. rep. 2012. *used on: p. 5*

[HCM15] Chien-Ming Huang, Maya Cakmak, and Bilge Mutlu. "Adaptive Coordination Strategies for Human-Robot Handovers." In: *Robotics: Science and Systems*. 2015. *used on: p. 42*

[Hub+09a] Markus Huber, Alois Knoll, Thomas Brandt, and Stefan Glasauer. "Handing over a cube: Spatial features of physical joint-action." In: *Annals of the New York Academy of Sciences*. 2009, pp. 380–382. DOI: 10.1111/j.1749-6632.2008.03743.x. *used on: pp. 35, 36, 39*

[Hub+08a] Markus Huber, Claus Lenz, Markus Rickert, and Alois Knoll. "Human preferences in industrial human-robot interactions." In: *Proceedings of the International Workshop on Cognition for Technical Systems* (2008). *used on: pp. 27, 44*

[Hub+09b] Markus Huber, Helmuth Radrich, Cornelia Wendt, Markus Rickert, Alois Knoll, Thomas Brandt, and Stefan Glasauer. "Evaluation of a novel biologically inspired trajectory generator in human-robot interaction." In: *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication* (2009), pp. 639–644. DOI: 10.1109/ROMAN.2009.5326233. *used on: pp. 38, 94*

[Hub+08b] Markus Huber, Markus Rickert, Alois Knoll, Thomas Brandt, and Stefan Glasauer. "Human-Robot Interaction in Handing-Over Tasks Markus." In: *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*. IEEE, 2008, pp. 107–112. DOI: 10.1109/ROMAN.2008.4600651. *used on: pp. 27, 38, 43, 44*

[Ima+02] Michita Imai, Takayuki Kanda, Testuo Ono, Hiroshi Ishiguro, and Kenji Mase. "Robot Mediated Round Table : Analysis of the Effect of Robot ' s Gaze." In: *Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communication*. 2002. *used on: p. 22*

[IOI03] Michita Imai, Tetsuo Ono, and Hiroshi Ishiguro. "Physical Relation and Expression : Joint Attention for Human–Robot Interaction." In: *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS* 50.4 (2003). *used on: p. 22*

[Ino+05] Kenji Inoue, Seri Nonaka, Yoshihiro Ujiie, Tomohito Takubo, and Tatsuo Arai. "Comparison of Human Psychology for Real and Virtual Mobile Manipulators." In: *RO-MAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005*. 13. Nashville, TN, USA, USA, 2005, pp. 73–78. *used on: p. 56*

[Int17] Intel Corporation. *Intel® RealSense™ Depth Camera D400-Series Datasheet*. Tech. rep. September. 2017. *used on: pp. 90, 97, 99, 107*

[IRS10] Luca Iocchi, Javier Ruiz-del-solar, and Jesus Savage. *RoboCup @ Home Rulebook 2010*. Tech. rep. 2010. *used on: p. 5*

[Ish06] Hiroshi Ishiguro. "Interactive Humanoids and Androids as Ideal Interfaces for Humans." In: *International Conference on Intelligent User Interfaces, Proceedings IUI* (2006), pp. 2–9. DOI: 10.1145/1111449.1111451. *used on: p. 16*

[JB04] D. Jansen and H. Buttner. "Real-time ethernet the EtherCAT solution." In: *Computing & Control Engineering Journal* 15.1 (2004). DOI: 10.1049/cce:20040104. *used on: p. 51*

[Jin+03] Mitsuru Jindai, Satoru Shibata, Tomonori Yamamoto, and Akira Shimizu. "A Study on Robot-Human System with Consideration of Individual Preferences (1st Report, Construction of a Planar-Two-Degrees-of-Freedom Robot System for Handing over Motion Based on Image Processing)." In: *JSME International Journal, Series C: Mechanical Systems, Machine Elements and Manufacturing* 46.3 (2003), pp. 1075–1083. DOI: 10.1299/jsmec.46.1075. *used on: p. 40*

[Jin+07] Mitsuru Jindai, Satoru Shibata, Tomonori Yamamoto, and Tomio Watanabe. "A Study on Robot-Human System with Consideration of Individual Preferences (2nd Report, Multimodal Human-Machine Interface for Object-Handing Robot System)." In: *JSME International Journal, Series C: Mechanical Systems, Machine Elements and Manufacturing* 49.4 (2007), pp. 1033–1039. ISSN: 13447653. DOI: 10.1299/jsmec.49.1033. *used on: p. 40*

[KI00] S. Kajikawa and E. Ishikawa. "Trajectory Planning for Hand-Over between Human and Robot." In: *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. 2000, pp. 281–287. DOI: 10.1109/ROMAN.2000.892509. *used on: pp. 41, 81, 173*

[Kaj+95] S Kajikawa, T Okino, K Ohba, and H Inooka. "Motion Planning for Hand-Over between Human and Robot." In: *Proceedings 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots*. IEEE, 1995. DOI: 10.1109/IROS.1995.525796. *used on: pp. 25, 40*

[KSO02] Shinya Kajikawa, Naoki Saito, and Hideharu Okano. "Receiver robot's motion for handing-over with a human." In: *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication* (2002), pp. 494–499. DOI: 10.1109/ROMAN.2002.1045670. *used on: p. 41*

[Kam+11] Hiroko Kamide, Mika Yasumoto, Yasushi Mae, Tomohito Takubo, Kenichi Ohara, and Tatsuo Arai. "Comparative Evaluation of Virtual and Real Humanoid with Robot-Oriented Psychology Scale." In: *2011 IEEE International Conference on Robotics and Automation*. Shanghai, China, 2011. *used on: pp. 56, 57*

[Kat+19] Saki Kato, Natsuki Yamanobe, Gentiane Venture, Eiichi Yoshida, and Gowrishankar Ganesh. "The where of handovers by humans: Effect of partner characteristics, distance and visual feedback." In: *PLoS ONE* 14.6 (2019), pp. 1–14. DOI: 10.1371/journal.pone.0217129. *used on: p. 40*

[KC69] A. Kendon and M. Cook. "The consistency of gaze patterns in social interaction." In: *British journal of psychology (London, England : 1953)* 60.4 (1969), pp. 481–494. ISSN: 00071269. DOI: 10.1111/j.2044-8295.1969.tb01222.x. *used on: p. 21*

[Ken90] Adam Kendon. *Conducting interaction: Patterns of behavior in focused encounters*. 7th ed. CUP Archive, 1990. *used on: pp. 175, 177*

[Ken04] Adam Kendon. *Gesture: Visible Action as Utterance*. Cambridge University Press, 2004, p. 400. ISBN: 978-0-521-54293-7. *used on: p. 19*

[Ken76] Adam Kendon. "The f-formation system: Spatial-orientational relations in face to face interaction." In: *Man Environment Systems* 6 (1976). *used on: pp. 32, 174*

[KI92] I. Kim and H. Inooka. "Hand-over of an object between human and robot." In: *Proceedings IEEE International Workshop on Robot and Human Communication*. Tokyo, Japan, Japan, 1992, pp. 199–203. DOI: 10.1109/roman.1992.253888. *used on: p. 45*

[Kim+04] Jinsul Kim, Jihwan Park, Yong K Hwang, and Manjai Lee. "Advanced Grasp Planning for Handover Operation Between Human and Robot: Three Handover Methods in Esteem Etiquettes Using Dual Arms and Hands of Home-Service Robot." In: c (2004), pp. 34–39. *used on: p. 31*

[KAD11] Nathan Kirchner, Alen Alempijevic, and Gamini Dissanayake. "Nonverbal Robot-Group Interaction Using an Imitated Gaze Cue." In: *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 2011. DOI: 10.1145/1957656.1957824. *used on: p. 22*

[Klu+14] A. Klußmann, J. B. Sattler, B. Arnold-Schulz-gahmen, A. Vasterling, H. Wagner, and B. Hartmann. "Leitlinie Händigkeit - Bedeutung und Untersuchung." In: *Die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V.* (2014). ISSN: 23634669. *used on: p. 47*

[Koa+07] KL L Koay, EA A Sisbot, DS S Syrdal, and M L Walters. "Exploratory Study of a Robot Approaching a Person in the Context of Handing Over an Object." In: *Spring Symp. on Multidisciplinary Collaboration for Socially Assistive Robotics*. 2007, pp. 18–24. *used on: pp. 34, 36, 47, 48*

[Koe+14] Ansgar Koene, Anthony Remazeilles, Miguel Prada, Ainara Garzo, Mildred Puerto, Satoshi Endo, and Alan M Wing. "Relative importance of spatial and temporal precision for user satisfaction in human-robot object handover interactions." In: *Third International Symposium on New Frontiers in Human-Robot Interaction* 14 (2014). *used on: p. 42*

[KT04] Mathias Kölsch and Matthew Turk. “Robust Hand Detection.” In: *FGR' 04 Proceedings of the Sixth IEEE international conference on Automatic face and gesture recognition*. 7. Korea, 2004, pp. 614–619. *used on: p. 93*

[Kru+13] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. “Human-aware robot navigation: A survey.” In: *Robotics and Autonomous Systems* 61.12 (2013), pp. 1726–1743. ISSN: 09218890. DOI: 10.1016/j.robot.2013.05.007. *used on: p. 35*

[KKH19] Alap Kshirsagar, Hadas Kress-Gazit, and Guy Hoffman. “Specifying and Synthesizing Human-Robot Handovers.” In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Macau, China: IEEE, 2019-11, pp. 5930–5936. *used on: p. 43*

[KHL16] Andras Kupcsik, David Hsu, and Wee Sun Lee. “Learning Dynamic Robot-to-Human Object Handover from Human Feedback.” In: *International Symposium on Robotics Research (ISRR)*. 2016. DOI: 10.1007/978-3-319-51532-8_10. arXiv: arXiv:1603.06390v1. *used on: pp. 43, 46*

[Lan62] J. M. F. Landsmeer. “Power grip and precision handling.” In: *Annals of the rheumatic diseases* 21.164 (1962). DOI: 10.1136/ard.21.2.164. *used on: p. 32*

[Lee] Min Kyung Lee. *Designing a Workplace Robotic Service*. Tech. rep. *used on: p. 38*

[Lee+11] Min Kyung Lee, Jodi Forlizzi, Sara Kiesler, Maya Cakmak, and Siddhartha Srinivasa. “Predictability or adaptivity?: designing robot handoffs modeled from trained dogs and people.” In: *Human-robot interaction* (2011), pp. 179–180. DOI: 10.1145/1957656.1957720. *used on: pp. 18, 24, 28, 112*

[LH15] Zhi Li and Kris Hauser. “Predicting Object Transfer Position and Timing in Human-Robot Handover Tasks.” In: *Science and Systems*. 2015. *used on: pp. 38, 95*

[Lie+17] Florian Lier, L Phillip, Joshua De Leeuw, Sven Wachsmuth, and Selma Sabanovi. “Can we Reproduce it ? Toward the Implementation of good Experimental Methodology in Interdisciplinary Robotics Research.” In: *CRA 2017 Workshop on Reproducible Research in Robotics: Current Status and Road Ahead*. Ed. by Bonsignorio FP (Ed). Proceedings of IEEE International Conference on Robotics and Automation (ICRA), IEEE Xplore. IEEE, 2017, pp. 3–5. *used on: p. 54*

[LSW14] Florian Lier, Simon Schulz, and Sven Wachsmuth. *Reality check! - A physical robot versus its simulation*. Tech. rep. 2014, p. 331. DOI: 10.1145/2559636.2559787. *used on: p. 75*

[Lie+14] Florian Lier, Johannes Wienke, Arne Nordmann, Sven Wachsmuth, and Sebastian Wrede. “The Cognitive Interaction Toolkit – Improving Reproducibility of Robotic Systems Experiments.” In: *SIMPAR 2014, LNAI*. Ed. by Davide Brugali. Springer International Publishing Switzerland, 2014, pp. 400–411. *used on: p. 54*

[LW10] Christine E. Looser and Thalia Wheatley. "The Tipping Point of Animacy: How, When, and Where We Perceive Life in a Face." In: *Psychological Science* 21.12 (2010), pp. 1854–1862. ISSN: 09567976. DOI: 10.1177/0956797610388044. *used on: p. 21*

[Lop+06] Efrain Lopez-Damian, Daniel Sidobre, Stephane De Tour, Rachid Alami, and De Narbonne. "Grasp Planning for Interactive Object Manipulation." In: *Proceedings of the International Symposium on Robotics and Automation*. 2006. *used on: p. 31*

[LMH13] Tamara Lorenz, Alexander Mörtl, and Sandra Hirche. *Movement synchronization fails during non-adaptive human-robot interaction*. Tech. rep. 2013, pp. 189–190. DOI: 10.1109/HRI.2013.6483565. *used on: p. 39*

[Lor+11] Tamara Lorenz, Alexander Mortl, Bjorn Vlaskamp, Anna Schubo, and Sandra Hirche. "Synchronization in a goal-directed task: Human movement coordination with each other and robotic partners." In: *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. 2011, pp. 198–203. DOI: 10.1109/ROMAN.2011.6005253. *used on: p. 39*

[Lue+10] Ingo Luetkebohle, Frank Hegel, Simon Schulz, Matthias Hackel, Britta Wrede, Sven Wachsmuth, and Gerhard Sagerer. "The Bielefeld Anthropomorphic Robot Head "Flobi"." In: *International Conference on Robotics and Automation*. IEEE, 2010-05, pp. 3384–3391. DOI: 10.1109/ROBOT.2010.5509173. *used on: p. 75*

[Mae+17a] Guilherme J. Maeda, Gerhard Neumann, Marco Ewerton, Rudolf Lioutikov, Oliver Kroemer, and Jan Peters. "Probabilistic movement primitives for coordination of multiple human–robot collaborative tasks." In: *Autonomous Robots* 41.3 (2017-03), pp. 593–612. ISSN: 15737527. DOI: 10.1007/s10514-016-9556-2. *used on: pp. 44, 45*

[Mae+17b] Guilherme Maeda, Marco Ewerton, Gerhard Neumann, Rudolf Lioutikov, and Jan Peters. "Phase estimation for fast action recognition and trajectory generation in human–robot collaboration." In: *International Journal of Robotics Research* 36.13-14 (2017), pp. 1579–1594. ISSN: 17413176. DOI: 10.1177/0278364917693927. *used on: pp. 38, 94*

[Mai+12] Jim Mainprice, Mamoun Gharbi, Thierry Simeon, and Rachid Alami. "Sharing effort in planning human-robot handover tasks." In: *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. IEEE, 2012-09, pp. 764–770. DOI: 10.1109/ROMAN.2012.6343844. *used on: p. 35*

[MHB15] Jim Mainprice, Rafi Hayne, and Dmitry Berenson. "Using Inverse Optimal Control To Predict Human Reaching Motion in Collaborative Tasks." In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015. *used on: p. 94*

[MES18] Jameel Malik, Ahmed Elhayek, and Didier Stricker. "Structure-Aware 3D hand pose regression from a single depth image." In: *International Conference on Virtual Reality and Augmented Reality*. 2018, pp. 3–17. DOI: 10.1007/978-3-030-01790-3_1. *used on: p. 94*

[Man+08] C. Manders, F. Farbiz, J. H. Chong, K. Y. Tang, G. G. Chua, M. H. Loke, and M. L. Yuan. "Robust hand tracking using a skin tone and depth joint probability model." In: *2008 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008*. September. 2008. DOI: 10.1109/AFGR.2008.4813459. *used on: p. 93*

[MBR17] Julieta Martinez, Michael J. Black, and Javier Romero. "On human motion prediction using recurrent neural networks." In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. 2017, pp. 4674–4683. DOI: 10.1109/CVPR.2017.497. *used on: p. 94*

[MM05] Andrea H Mason and Æ Christine L Mackenzie. "Grip forces when passing an object to a partner." In: (2005), pp. 173–187. DOI: 10.1007/s00221-004-2157-x. *used on: pp. 25, 45, 130*

[Mat14] Rob Matheson. *'Sociable' humanoid robots could help advance human-robot interaction*. Tech. rep. February. 2014, pp. 1–4. *used on: p. 206*

[Med+16] Jose R. Medina, Felix Duvallet, Murali Karnam, and Aude Billard. "A human-inspired controller for fluid human-robot handovers." In: *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2016-11, pp. 324–331. DOI: 10.1109/HUMANOIDS.2016.7803296. *used on: pp. 28, 42, 47*

[Mek10] Meka Robotics. *B1 Omni Base Product Sheet*. Tech. rep. San Francisco, CA, USA, 2010, pp. 1–4. *used on: p. 206*

[Mek09] Meka Robotics. *H2 Compliant Hand Datasheet*. Tech. rep. San Francisco, CA, USA, 2009, pp. 2–4. *used on: p. 206*

[Meu+07] Ruud G.J. Meulenbroek, Jurjen Bosga, Majken Hulstijn, and Stephan Miedl. "Joint-action coordination in transferring objects." In: *Experimental Brain Research* 180.2 (2007), pp. 333–343. ISSN: 00144819. DOI: 10.1007/s00221-007-0861-z. *used on: pp. 31, 32*

[MSS11] Vincenzo Micelli, Kyle Strabala, and Siddhartha Srinivasa. "Perception and Control Challenges for Effective Human-Robot Handoffs." In: *Research Showcase @ CMU* (2011). *used on: p. 24*

[Mit+08] Noriaki Mitsunaga, Christian Smith, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. "Adapting Robot Behavior for Human-Robot Interaction." In: *IEEE Transactions on Robotics* 24.4 (2008), pp. 911–916. ISSN: 15523098. DOI: 10.1109/TRO.2008.926867. *used on: p. 18*

[Moo+14] AJung Moon, Daniel M. Troniak, Brian Gleeson, Matthew K.X.J. X. J. Pan, Minhua Zeng, Benjamin A. Blumer, Karon MacLean, and Elizabeth A. Croft. "Meet Me where I'm Gazing: How Shared Attention Gaze Affects Human-Robot Handover Timing." In: *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14*. New York, New York, USA: ACM Press, 2014, pp. 334–341. DOI: 10.1145/2559636.2559656. *used on: pp. 4, 23, 25, 28*

[Mor70] Masahiro Mori. "The Uncanny Valley." In: *Energy* 7.4 (1970), pp. 33–35. *used on: p. 5*

[MM11] Jonathan Mumm and Bilge Mutlu. "Human-robot proxemics: Physical and psychological distancing in human-robot interaction." In: *Proceedings of the 6th International Conference on Human-Robot Interaction, HRI 2011*. May 2014. Lausanne, Switzerland, 2011, p. 331. DOI: 10.1145/1957656.1957786. *used on: p. 22*

[MBM12] Matteo Munaro, Filippo Basso, and Emanuele Menegatti. "Tracking people within groups with RGB-D data." In: *IEEE International Conference on Intelligent Robots and Systems*. October. 2012, pp. 2101–2107. DOI: 10.1109/IROS.2012.6385772. *used on: p. 94*

[Mut09] Bilge Mutlu. "Designing Gaze Behavior for Humanlike Robots." Doctoral dissertation. Carnegie Mellon University, 2009. *used on: pp. 22, 119*

[Mut+09a] Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. "Footing In Human-Robot Conversations : How Robots Might Shape Participant Roles Using Gaze Cues." In: *Proceedings of the 4th ACM/IEEE international conference on Human-Robot Interaction*. 1. 2009, pp. 61–68. *used on: pp. 22, 79*

[Mut+09b] Bilge Mutlu, Fumitaka Yamaoka, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. "Nonverbal leakage in robots: Communication of Intentions through Seemingly Unintentional Behavior." In: *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* 2.1 (2009), p. 69. DOI: 10.1145/1514095.1514110. *used on: p. 18*

[Nag+98] K. Nagata, Y. Oosaki, M. Kakikura, and H. Tsukune. "Delivery by Hand between Human and Robot Based on Fingertip Force-Torque Information." In: *Proceedings. 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems. Innovations in Theory, Practice and Applications (Cat. No.98CH36190)*. Victoria, B.C., Canada: IEEE/RSJ, 1998-10, pp. 750–757. DOI: 10.1109/IROS.1998.727283. *used on: pp. 25, 45*

[NH86] Yoshihiko Nakamura and Hideo Hanafusa. "Inverse kinematic solutions with singularity robustness for robot manipulator control." In: *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME* 108.3 (1986), pp. 163–171. ISSN: 15289028. DOI: 10.1115/1.3143764. *used on: p. 83*

[Nap56] J. R. Napier. "The prehensile movements of the human hand." In: *The Journal of Bone and Joint Surgery* 38-B.4 (1956). DOI: 10.1302/0301-620X.38B4.902. *used on: pp. 32, 177*

[NDL19] Heramb Nemlekar, Dharini Dutia, and Zhi Li. "Object Transfer Point Estimation for Fluent Human-Robot Handovers." In: *International Conference on Robotics and Automation (ICRA)*. Montreal, Canada, 2019. *used on: pp. 38, 39, 95, 102, 103, 175*

[NDL18] Heramb Nemlekar, Dharini Dutia, and Zhi Li. "Prompt Human to Robot Handovers by Estimation of Object Transfer Point based on Human Partner ' s Motion." In: *Workshop on Human-Robot Cooperation and Collaboration in Manipulation: Advancements and Challenges*. Madrid, Spain, 2018. *used on: pp. 38, 93, 95*

[Ngu+18a] Dong Hai Phuong Nguyen, Matej Hoffmann, Alessandro Roncone, Ugo Pattacini, and Giorgio Metta. "Compact Real-time Avoidance on a Humanoid Robot for Human-robot Interaction." In: *ACM/IEEE International Conference on Human-Robot Interaction*. 2018, pp. 416–424. DOI: 10.1145/3171221.3171245. arXiv: arXiv:1801.05671v1. *used on: pp. 44, 93*

[Ngu+18b] Phuong D.H. Nguyen, Fabrizio Bottarel, Ugo Pattacini, Matej Hoffmann, Lorenzo Natale, and Giorgio Metta. "Merging Physical and Social Interaction for Effective Human-Robot Collaboration." In: *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*. October. Beijing, China, 2018-11, pp. 710–717. DOI: 10.1109/HUMANOIDS.2018.8625030. *used on: pp. 38, 44*

[Nis+08] Koichi Nishiwaki, Kazuhiko Kobayashi, Shinji Uchiyama, Hiroyuki Yamamoto, and Satoshi Kagami. "Mixed Reality Environment for Autonomous Robot Development." In: *2008 IEEE International Conference on Robotics and Automation*. Pasadena, CA, USA, 2008, pp. 2211–2212. *used on: p. 56*

[Nom+06] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. "Measurement of negative attitudes toward robots." In: *Interaction StudiesInteraction Studies. Social Behaviour and Communication in Biological and Artificial Systems* 7.3 (2006-11), pp. 437–454. ISSN: 1572-0373. DOI: 10.1075/is.7.3.14nom. *used on: p. 68*

[OG04] John Ed Oates and Andrew Ed Grayson. *Cognitive and language development in children*. Oxford, UK: Blackwell Publishing, 2004. ISBN: 978-1-4051-1045-7. *used on: pp. 21, 175*

[Pan+19] Matthew K X J Pan, Espen Knoop, Moritz Bächer, and Günter Niemeyer. "Fast Handovers with a Robot Character : Small Sensorimotor Delays Improve Perceived Qualities." In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Macau, China: IEEE/RSJ, 2019-11, pp. 6735–6741. *used on: p. 45*

[Pan+17] Matthew KXJ Pan, Vidar Skjervøy, Wesley P Chan, Masayuki Inaba, and Elizabeth A Croft. "Automated detection of handovers using kinematic features." In: *International Journal of Robotics Research* 36.5-7 (2017-06), pp. 721–738. ISSN: 17413176. DOI: 10.1177/0278364917692865. *used on: pp. 94, 95*

[Pan+11a] Amit Kumar Pandey, Muhammad Ali, Matthieu Warnier, and Rachid Alami. "Towards multi-state visuo-spatial reasoning based proactive human-robot interaction." In: *IEEE 15th International Conference on Advanced Robotics: New Boundaries for Robotics, ICAR 2011*. 2011, pp. 143–149. DOI: 10.1109/ICAR.2011.6088642. *used on: p. 36*

[Pan+11b] Maja Pantic, Roderick Cowie, Francesca D'Errico, Dirk Heylen, Marc Mehu, Catherine Pelachaud, Isabella Poggi, Marc Schroeder, and Alessandro Vinciarelli. "Social Signal Processing: The Research Agenda." In: *Visual Analysis of Humans* (2011), pp. 511–538. DOI: 10.1007/978-0-85729-997-0_26. *used on: p. 18*

[Par18] Mohammad Sina Parastegari. "Modelling and Control of Object Handover , A Study in Human-Robot Interaction." Doctoral dissertation. University of Illinois at Chicago, 2018. *used on: pp. 29, 37*

[PAN17] Sina Parastegari, Bahareh Abbasi, and Ehsan Noohi. "Modeling Human Reaching Phase in Human-Human Object Handover with Application in Robot-Human Handover." In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2017), pp. 3597–3602. DOI: 10.1109/IROS.2017.8206205. *used on: pp. 25, 29, 38, 94*

[Par+16] Sina Parastegari, Ehsan Noohi, Bahareh Abbasi, and Miloš Žefran. "A fail-safe object handover controller." In: *Proceedings - IEEE International Conference on Robotics and Automation* (2016), pp. 2003–2008. ISSN: 10504729. DOI: 10.1109/ICRA.2016.7487346. *used on: p. 45*

[PPM19] Jae Sung Park, Chonhyon Park, and Dinesh Manocha. "I-Planner: Intention-aware motion planning using learning-based human motion prediction." In: *International Journal of Robotics Research* 38.1 (2019), pp. 23–39. ISSN: 17413176. DOI: 10.1177/0278364918812981. *used on: p. 43*

[Pat+10] Ugo Pattacini, Francesco Nori, Lorenzo Natale, Giorgio Metta, and Giulio Sandini. "An experimental evaluation of a novel minimum-jerk Cartesian controller for humanoid robots." In: *IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings* (2010), pp. 1668–1674. DOI: 10.1109/IROS.2010.5650851. *used on: p. 44*

[PP04] Allan Pease and Barbara Pease. "The Definitive Book of Body Language: How to read others thoughts by their gestures." In: *Library* (2004). ISSN: 03630277. DOI: 10.1162/asep.2010.9.1.54. arXiv: 1007.5476. *used on: p. 4*

[PE94] Ira B. Perelle and Lee Ehrman. "An International Study of Human Handedness: The Data." In: *Behavior Genetics* 24.3 (1994). *used on: p. 48*

[PS15] Claudia Perez-D'Arpino and Julie A. Shah. "Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification." In: *Proceedings - IEEE International Conference on Robotics and Automation* 2015-June.June (2015), pp. 6175–6182. ISSN: 10504729. DOI: 10.1109/ICRA.2015.7140066. *used on: p. 94*

[PA15] Sammy Pfeiffer and Cecilio Angulo. "Gesture learning and execution in a humanoid robot via dynamic movement primitives." In: *Pattern Recognition Letters* 000 (2015-12), pp. 1–8. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2015.07.042. *used on: p. 20*

[Pha+07] Quang Cuong Pham, Halim Hicheur, Gustavo Arechavaleta, Jean Paul Laumond, and Alain Berthoz. "The formation of trajectories during goal-oriented locomotion in humans. II. A maximum smoothness model." In: *European Journal of Neuroscience* 26.8 (2007), pp. 2391–2403. DOI: 10.1111/j.1460-9568.2007.05835.x. *used on: p. 34*

[PSK11] Roland Philippsen, Luis Sentis, and Oussama Khatib. "An open source extensible software package to create whole-body compliant skills in personal mobile manipulators." In: *IEEE International Conference on Intelligent Robots and Systems*. 2011, pp. 1036–1041. DOI: 10.1109/IROS.2011.6048762. *used on: p. 139*

[PD10] Isabella Poggi and Francesca D'Errico. "Cognitive modelling of human social signals." In: *SSPW'10 - Proceedings of the 2010 ACM Social Signal Processing Workshop, Co-located with ACM Multimedia 2010* June (2010), pp. 21–26. DOI: 10.1145/1878116.1878124. *used on: pp. 18, 178*

[Poh18] Jan Pohlmeyer. "Human-Inspired Gaze Control Strategies for an Anthropomorphic Robot Head during Human-Robot Handover." B.Sc. Thesis. Bielefeld University, 2018. *used on: pp. 79, 80*

[Pra+13] Miguel Prada, Anthony Remazeilles, Ansgar Koene, and Satoshi Endo. "Dynamic Movement Primitives for Human-Robot interaction: Comparison with human behavioral observation." In: *IEEE International Conference on Intelligent Robots and Systems* (2013), pp. 1168–1175. ISSN: 21530858. DOI: 10.1109/IROS.2013.6696498. *used on: p. 42*

[Pra+14] Miguel Prada, Anthony Remazeilles, Ansgar Koene, and Satoshi Endo. "Implementation and experimental validation of Dynamic Movement Primitives for object handover." In: *IEEE International Conference on Intelligent Robots and Systems* (2014), pp. 2146–2153. ISSN: 21530866. DOI: 10.1109/IROS.2014.6942851. *used on: pp. 25, 42*

[PKM02] Jerry Pratt, Ben Krupp, and Chris Morse. “Series elastic actuators for high fidelity force control.” In: *Industrial Robot* 29.3 (2002), pp. 234–241. DOI: 10.1108/01439910210425522. *used on: p. 51*

[Pri12] Primesense Ltd. *The PrimeSense 3D Awareness Sensor*. Tech. rep. 2012, pp. 1–4. *used on: p. 52*

[Qia+14] Chen Qian, Xian Sun, Yichen Wei, Xiaoou Tang, and Jian Sun. “Realtime and Robust Hand Tracking from Depth.” In: *IEEE Conference on Computer Vision and Pattern Recognition*. 10. 2014, pp. 1133–1144. DOI: 10.1109/CVPR.2014.145. *used on: p. 94*

[Qui+09] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. “ROS: an open-source Robot Operating System.” In: *ICRA Workshop on Open Source Software*. 2009. DOI: 10.1109/IECON.2015.7392843. arXiv: 1106.4561. *used on: pp. 53, 177*

[QBS15] Ana Huaman Quispe, Heni Ben Amor, and Mike Stilman. “Handover planning for every occasion.” In: *IEEE-RAS International Conference on Humanoid Robots*. 2015, pp. 431–436. DOI: 10.1109/HUMANOIDS.2014.7041396. *used on: p. 41*

[RWK18] Robin Rasch, Sven Wachsmuth, and Matthias König. “A Joint Motion Model for Human-Like Robot-Human Handover.” In: *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*. Beijing, China: IEEE, 2018. DOI: 10.1109/HUMANOIDS.2018.8624967. arXiv: 1808.09280. *used on: pp. 44, 82*

[RWK17] Robin Rasch, Sven Wachsmuth, and Matthias König. “Understanding Movements of Hand-Over Between two Persons to Improve Humanoid Robot Systems.” In: *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*. Birmingham, UK, 2017, pp. 856–861. DOI: 10.1109/HUMANOIDS.2017.8246972. *used on: p. 44*

[Ren+18] Patrick Renner, Florian Lier, Felix Friese, Thies Pfeiffer, and Sven Wachsmuth. “Facilitating HRI by Mixed Reality Techniques.” In: *HRI '18 Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. Chicago, IL, USA, 2018, pp. 215–216. *used on: p. 56*

[Rie+10] Laurel D Riek, Tal-chen Rabinowitch, Paul Bremner, Anthony G Pipe, Mike Fraser, and Peter Robinson. “Cooperative Gestures : Effective Signaling for Humanoid Robots.” In: *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*. IEEE Press, 2010, pp. 61–68. *used on: p. 20*

[Rog13] Sally J. Rogers. “Mutual Gaze.” In: *Encyclopedia of Autism Spectrum Disorders*. Ed. by Fred R. Volkmar. New York, NY: Springer, 2013, pp. 1966–1967. ISBN: 978-1-4419-1698-3. DOI: 10.1007/978-1-4419-1698-3_628. *used on: p. 176*

[Sat+12] Satoru Satake, Takayuki Kanda, Dylan F. Glas, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. "How to Approach Humans?-Strategies for Social Robots to Initiate Interaction-." In: *Journal of the Robotics Society of Japan* 28.3 (2012), pp. 327–337. ISSN: 0289-1824. DOI: 10.7210/jrsj.28.327. *used on: p. 35*

[SG64] Abraham Savitzky and Marcel J.E. Golay. "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." In: *Analytical Chemistry* 36.8 (1964), pp. 1627–1639. ISSN: 15206882. DOI: 10.1021/ac60214a047. *used on: p. 115*

[Sch06] Stefan Schaal. "Dynamic Movement Primitives - A Framework for Motor Control in Humans and Humanoid Robotics." In: *Adaptive Motion of Animals and Machines* 1 (2006), pp. 261–280. DOI: 10.1007/4-431-31381-8_23. *used on: pp. 20, 173*

[Sch+16] Simon Schulz, Florian Lier, Andreas Kipp, and Sven Wachsmuth. "Humotion - A human inspired gaze control framework for anthropomorphic robot heads." In: *HAI 2016 - Proceedings of the 4th International Conference on Human Agent Interaction*. 2016, pp. 207–214. DOI: 10.1145/2974804.2974827. *used on: pp. 75, 77, 209*

[Sen07] Luis Sentis. "Synthesis and Control of Whole-Body Behaviors in Humanoid Systems." Doctoral dissertation. Stanford University, 2007, pp. 1–208. ISBN: 978-0-549-24604-6. *used on: p. 139*

[STS95] S. Shibata, K. Tanaka, and A. Shimizu. "Experimental analysis of handing over." In: *Robot and Human Communication, 1995. RO-MAN'95 TOKYO, Proceedings., 4th IEEE International Workshop on* (1995), pp. 53–58. DOI: 10.1109/ROMAN.1995.531934. *used on: pp. 35, 40*

[Shi+97] Satoru Shibata, Benlamine Mohamed Sahbi, Kanya Tanaka, and Akira Shimizu. "An Analysis of The Process of Handing Over An Object and Its Application to Robot Motions." In: 04 (1997). *used on: pp. 40, 94*

[SK08] Bruno Siciliano and Oussama Khatib. *Handbook of Robotics*. Vol. 109. 7. Springer, 2008. ISBN: 978-3-540-23957-4. *used on: pp. 83, 175*

[SPD18] Antonis Sidiropoulos, Efi Psomopoulou, and Zoe Doulgeri. "A human inspired handover policy using Gaussian Mixture Models and haptic cues." In: *Autonomous Robots* (2018-02), pp. 1–16. ISSN: 15737527. DOI: 10.1007/s10514-018-9705-x. *used on: pp. 43, 47*

[Sim19] Janneke Simmering. "Object Transfer Point Prediction on Humanoid Robots for Human Initiated Handovers." B.Sc. Thesis. Bielefeld University, 2019. *used on: pp. 97, 138*

[Sim+17] Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. "Hand Keypoint Detection in Single Images using Multiview Bootstrapping." In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. 2017-04, pp. 4645–4653. DOI: 10.1109/CVPR.2017.494. arXiv: 1704.07809. *used on: pp. 69, 99, 209*

[Sin+18] Harmeet Singh, Marco Controzzi, Torquato Cecchini, and Christian Cipriani. "Online Prediction of Robot to Human Handover Events Using Vibrations." In: *International Symposium on Robot and Human Interactive Communication*. 2018. *used on: p. 47*

[SA12] E. A. Sisbot and R. Alami. "A Human-Aware Manipulation Planner." In: *IEEE Transactions on Robotics* 28.5 (2012), pp. 1045–1057. ISSN: 1552-3098. DOI: 10.1109/TRO.2012.2196303. *used on: pp. 22, 41*

[Sis08] Emrah Akin Sisbot. "Towards human-aware robot motions." Doctoral dissertation. Université Paul Sabatier - Toulouse III, 2008. *used on: p. 38*

[Sis+07] Emrah Akin Sisbot, Luis F. Marin-Urias, R. Alami, and T. Simeon. "A Human Aware Mobile Robot Motion Planner." In: *IEEE Transactions on Robotics* 23.5 (2007), pp. 874–883. ISSN: 1552-3098. DOI: 10.1109/TRO.2007.904911. *used on: pp. 34, 41*

[Six18] Dominik Sixt. "Responsive Strategy for Smooth Handover with a Humanoid Robot." M.Sc. Thesis. Bielefeld University, 2018. *used on: pp. 84, 86, 90, 91*

[SE17] Roy Someshwar and Yael Edan. "Givers & Receivers perceive handover tasks differently: Implications for Human-Robot collaborative system design." In: *arXiv* (2017). DOI: arXiv:1708.06207v2. arXiv: 1708.06207. *used on: pp. 25, 36, 39*

[SOT13] Srinath Sridhar, Antti Oulasvirta, and Christian Theobalt. "Interactive Markerless Articulated Hand Motion Tracking Using RGB and Depth Data." In: *Proceedings of the IEEE International Conference on Computer Vision* (2013), pp. 2456–2463. DOI: 10.1109/ICCV.2013.305. *used on: p. 93*

[Sti+05] Mike Stilman, Philipp Michel, Joel Chestnutt, Koichi Nishiwaki, Satoshi Kagami, and James J Kuffner. *Augmented Reality for Robot Development and Experimentation*. Tech. rep. Pittsburgh, PA, 2005, pp. 1–11. *used on: p. 56*

[Str+13] Kyle Strabala, Min Kyung Lee, Anca Dragan, Jodi Forlizzi, Siddhartha S. Srinavasa, Maya Cakmak, and Vincenzo Micelli. "Towards Seamless Human-Robot Handovers." In: *Journal of Human-Robot Interaction (2013)* 1.1 (2013), pp. 112–132. ISSN: 2163-0364. DOI: 10.5898/jhri.v2i1.114. *used on: pp. 4, 20, 23, 29, 35*

[Str+12] Kyle Strabala, Min Kyung Lee, Anca Dragan, Jodi Forlizzi, and Siddhartha S. Srinivasa. "Learning the communication of intent prior to physical collaboration." In: *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication* (2012), pp. 968–973. ISSN: 1944-9445. DOI: 10.1109/ROMAN.2012.6343875. *used on: pp. 4, 22–24, 26, 28, 94*

[SS15] Halit Bener Suay and Emrah Akin Sisbot. "A Position Generation Algorithm Utilizing A Biomechanical Model For Robot-Human Object Handover." In: *Proceedings - IEEE International Conference on Robotics and Automation*. IEEE, 2015, pp. 3776–3781. DOI: 10.1109/ICRA.2015.7139724. *used on: pp. 4, 28, 35*

[TGB00] Deepak Tolani, Ambarish Goswami, and Norman I. Badler. "Real-time inverse kinematics techniques for anthropomorphic limbs." In: *Graphical Models* 62.5 (2000), pp. 353–388. ISSN: 15240703. DOI: 10.1006/gmod.2000.0528. *used on: p. 41*

[UO99] Akira Utsumi and Jun Ohya. "Multiple-Hand-Gesture Tracking using Multiple Cameras." In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1999, pp. 473–478. DOI: 10.1109/cvpr.1999.786980. *used on: p. 93*

[Vah+16] Nikolaus Vahrenkamp, Harry Arnst, Mirko Wächter, David Schiebener, Panagiotis Sotiropoulos, Michal Kowalik, and Tamim Asfour. "Workspace analysis for planning human-robot interaction tasks." In: *IEEE-RAS International Conference on Humanoid Robots*. September 2017. 2016, pp. 1298–1303. DOI: 10.1109/HUMANOIDS.2016.7803437. *used on: pp. 36, 37*

[Vog+18] David Vogt, Simon Stepputtis, Bernhard Jung, and Heni Ben Amor. "One-shot learning of human–robot handovers with triadic interaction meshes." In: *Autonomous Robots* 42.5 (2018), pp. 1053–1065. ISSN: 15737527. DOI: 10.1007/s10514-018-9699-4. *used on: pp. 29, 42, 43*

[VE90] V. Volterra and C.J. Erting. *From Gesture to Language in Hearing and Deaf Children*. 1990. ISBN: 978-3-642-67760-1. DOI: 10.1007/978-3-642-74859-2. *used on: p. 19*

[WGA15] Jules Waldhart, Mamoun Gharbi, and Rachid Alami. "Planning handovers involving humans and robots in constrained environment." In: *IEEE International Conference on Intelligent Robots and Systems*. IEEE, 2015-09, pp. 6473–6478. DOI: 10.1109/IROS.2015.7354302. *used on: p. 35*

[Wal+05a] Michael L. Walters, Kerstin Dautenhahn, Kheng Lee Koay, Christina Kaouri, René Te Boekhorst, Chrystopher Nehaniv, Iain Werry, and David Lee. "Close Encounters: Spatial Distances between People and a Robot of Mechanistic Appearance." In: *Proceedings of 2005 5th IEEE-RAS International Conference on Humanoid Robots*. 2005, pp. 450–455. DOI: 10.1109/ICHR.2005.1573608. *used on: p. 33*

[Wal+07] Michael L. Walters, Kerstin Dautenhahn, Sarah N. Woods, and Kheng Lee Koay. "Robotic etiquette: Results from user studies involving fetch and carry task." In: *Proceeding of the ACM/IEEE international conference on Human-robot interaction - HRI '07*. 2007. 2007. DOI: 10.1145/1228716.1228759. *used on: p. 34*

[Wal+05b] Michael L Walters, Kerstin Dautenhahn, René te Boekhorst, Kheng Lee Koay, Christina Kaouri, Sarah N Woods, Chrystopher L Nehaniv, David Lee, and Iain Werry. "The Influence of Subjects' Personality Traits on Predicting Comfortable Human-Robot Approach Distances." In: *Proceedings of Cog Sci 2005 Workshop: Toward Social Mechanisms of Android Science*. Stresa, Italy, 2005-07. *used on: p. 33*

[Wal+06] Michael L Walters, Kerstin Dautenhahn, Sarah N Woods, Kheng Lee Koay, R Te Boekhorst, and David Lee. "Exploratory Studies on Social Spaces between Humans and a Mechanical-looking Robot." In: *Connection Science* 18.Android Science Special Issue (2006), pp. 429–439. DOI: 10.1080/09540090600879513. *used on: p. 34*

[Wam86] Charles W. Wampler. "Manipulator Inverse Kinematic Solutions Based on Vector Formulations and Damped Least-Squares Methods." In: *IEEE Transactions on Systems, Man and Cybernetics*. 1. 1986, pp. 93–101. DOI: 10.1109/TSMC.1986.289285. *used on: p. 83*

[Wan+19] Weitian Wang, Rui Li, Zachary Max Diekel, Yi Chen, Zhujun Zhang, and Yunyi Jia. "Controlling object hand-over in human-robot collaboration via natural wearable sensing." In: *IEEE Transactions on Human-Machine Systems* 49.1 (2019), pp. 59–71. ISSN: 21682291. DOI: 10.1109/THMS.2018.2883176. *used on: p. 20*

[WBJ67] P. Watzlawick, J. Beavin-Bavelas, and D. Jackson. "Some Tentative Axioms of Communication." In: *In Pragmatics of Human Communication - A Study of Interactional Patterns, Pathologies and Paradox.* Ed. by s. W. W. Norton. New York, 1967. *used on: p. 18*

[Web+16] Matt Webster, David Western, Dejanira Araiza-Illan, Clare Dixon, Kerstin Eder, Michael Fisher, and Anthony G. Pipe. "An Assurance-based Approach to Verification and Validation of Human–Robot Teams." In: August (2016). arXiv: 1608.07403. *used on: p. 37*

[Wei+16] Shih En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. "Convolutional Pose Machines." In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2016-01, pp. 4724–4732. DOI: 10.1109/CVPR.2016.511. arXiv: 1602.00134. *used on: p. 209*

[WE84] W. A. Wolovich and H Elliott. "A computational technique for inverse kinematics." In: *The 23rd IEEE Conference on Decision and Control.* December. Las Vegas, Nevada, USA: IEEE, 1984. *used on: p. 83*

[Wre+17] Sebastian Wrede, Christian Leichsenring, Patrick Holthaus, Thomas Hermann, and Sven Wachsmuth. "The Cognitive Service Robotics Apartment." In: *KI - Künstliche Intelligenz* 31.3 (2017), pp. 299–304. ISSN: 0933-1875. DOI: 10.1007/s13218-017-0492-x. *used on: pp. 55, 173*

[Xu+09] Jiang Xu, J. Patrick Gannon, Karen Emmorey, Jason F. Smith, and Allen R. Braun. "Symbolic gestures and spoken language are processed by a common neural system." In: *Proceedings of the National Academy of Sciences* 106.49 (2009). DOI: 10.1073/pnas.0909197106. *used on: p. 19*

[YRA13] Katsu Yamane, Marcel Revfi, and Tamim Asfour. "Synthesizing object receiving motions of humanoid robots with human motion database." In: *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 1629–1636. DOI: 10.1109/ICRA.2013.6630788. *used on: pp. 25, 42*

[YD02] Ha Holly A Yanco and Jill L Jl Drury. "A Taxonomy for Human-Robot Interaction." In: *AAAI Fall Symposium on Human-Robot Interaction*. 2002-11, pp. 111–119. DOI: 10.1109/ICSMC.2004.1400783. *used on: p. 18*

[YD04] Holly A. Yanco and Jill Drury. "Classifying human-robot interaction: An updated taxonomy." In: *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics* 3.May (2004), pp. 2841–2846. ISSN: 1062922X. DOI: 10.1109/ICSMC.2004.1400763. *used on: p. 18*

[ZW06] Tijn van der Zant and Thomas Wisspeintner. "RoboCup X: A Proposal for a New League Where RoboCup Goes Real World." In: 2006, pp. 166–172. DOI: 10.1007/11780519_15. *used on: pp. 5, 61*

[ZW07] Tijn van der Zant and Thomas Wisspeintner. "Robocup@home: Creating and benchmarking tomorrows service robot applications." In: *Robotic Soccer*. IntechOpen, 2007, p. 64. DOI: 10.5772/32009. arXiv: 9809069v1 [arXiv:gr-qc]. *used on: pp. 5, 61*

[ZM12] Minhua Zheng and Max Q.H. Meng. "Designing Gestures with Semantic Meanings for Humanoid Robot." In: *2012 IEEE International Conference on Robotics and Biomimetics, ROBIO 2012 - Conference Digest*. IEEE, 2012, pp. 287–292. DOI: 10.1109/ROBIO.2012.6490981. *used on: p. 20*

[Zhe+14] Minhua Zheng, A. Jung Moon, Brian Gleeson, Daniel M. Troniak, Matthew K.X.J. Pan, Benjamin A. Blumer, Max Q.H. Meng, and Elizabeth A. Croft. "Human Behavioural Responses to Robot Head Gaze during Robot-to-Human Handovers." In: *Proceedings of the 2014 IEEE International Conference on Robotics and Biomimetics, IEEE ROBIO 2014*. Bali, Indonesia: IEEE, 2014, pp. 362–367. DOI: 10.1109/ROBIO.2014.7090357. *used on: pp. 4, 23, 37*

[Zhe+15] Minhua Zheng, AJung J. Moon, Elizabeth A. Croft, and Max Q.H. Meng. “Impacts of Robot Head Gaze on Robot-to-Human Handovers.” In: *International Journal of Social Robotics* 7.5 (2015), pp. 783–798. ISSN: 18754805. DOI: 10.1007/s12369-015-0305-z. *used on: p. 23*

ONLINE RESOURCES

[ATI:Mini40] *6-Axis Force and Torque Sensor (Mini40 Series)*. ATI Industrial Automation. URL: https://www.ati-ia.com/products/ft/ft_models.aspx?id=Mini40 (visited on 2019-08-29). *used on: pp. 52, 62*

[Meka:arm] *A2 Compliant Arm*. Meka Robotics. URL: http://mekabot.com/products/compliant-arm/ (visited on 2013-02-08). *used on: pp. 52, 176*

[Meka:base] *B1 Omni Base*. [Mek10]. Meka Robotics. URL: http://mekabot.com/products/omni-base/ (visited on 2013-10-02). *used on: pp. 51, 176*

[CK:m-n] Sebastian Meyer zu Borgsen. *Distribution Details: Meka nightly*. Bielefeld University. URL: https://citkat-citec.bob.ci.cit-ec.net/distribution/meka-nightly.xml (visited on 2019-08-22). *used on: p. 108*

[CITEC] CITEC. *Press Images*. Cluster of Excellence - Cognitive Interaction Technology. URL: https://www.cit-ec.de/en/press-images (visited on 2018-09-12). *used on: p. 3*

[Der17] Savitzky-Golay Derivative. *Least squares filtering for estimating the rate of change, including Savitzky-Golay*. 2017. URL: https://gregstanleyandassociates.com/whitepapers/FaultDiagnosis/Filtering/Derivative-Estimation/derivative-estimation.htm#least-squares (visited on 2019-08-14). *used on: p. 115*

[laterality] Walter Ehrenstein and Birgit Arnold-Schulz-Gahmen. *Seitenbevorzugung – Selbsttest*. IfADo – Leibniz-Institut für Arbeitsforschung an der TU Dortmund. 2000. URL: https://www.ifado.de/seitenbevorzugung-selbsttest/ (visited on 2019-09-29). *used on: p. 119*

[Meka:hand] *H2 Compliant Hand*. [Mek09]. Meka Robotics. URL: http://mekabot.com/products/compliant-hand/ (visited on 2013-10-02). *used on: pp. 52, 176*

[hri:conf] Human-Robot Interaction. *Human-Robot Interaction | A Research Portal for the HRI Community*. URL: http://humanrobotinteraction.org (visited on 2019-02-14). *used on: p. 17*

[Meka:robot] *M1 Mobile Manipulator. ’Sociable’ humanoid robots could help advance human-robot interaction* [Mat14]. Meka Robotics. URL: http://mekabot.com/products/m1-mobile-manipulator/ (visited on 2013-10-02). *used on: pp. 51, 176*

[MW:gesture] Merriam-Webster. *Gesture | Definition of Gesture by Merriam-Webster*. URL: https://www.merriam-webster.com/dictionary/gesture (visited on 2019-02-12). *used on: p. 19*

[MW:haov] Merriam-Webster. *Handover | Definition of Robot by Merriam-Webster*. URL: https://www.merriam-webster.com/dictionary/handover (visited on 2019-02-11). *used on: pp. 24, 25*

[MW:robot] Merriam-Webster. *Robot | Definition of Robot by Merriam-Webster*. URL: https://www.merriam-webster.com/dictionary/robot (visited on 2019-02-12). *used on: p. 15*

[Classicthesis] André Miede. *A Classic Thesis Style – An Homage to The Elements of Typographic Style*. URL: https://bitbucket.org/amiede/classicthesis (visited on 2019-02-05). *used on: p. 213*

[OED:hand] Oxford University Press. *hand, n. : Oxford English Dictionary*. The Oxford English Dictionary. URL: http://www.oed.com/view/Entry/83801?rskey=Melk8f&result=1 (visited on 2019-02-13). *used on: p. 24*

[OED:haov] Oxford University Press. *handover, n.2 : Oxford English Dictionary*. The Oxford English Dictionary. URL: http://www.oed.com/view/Entry/361463?rskey=iNoY5c&result=2#eid (visited on 2019-02-11). *used on: p. 24*

[OED:locom.] Oxford University Press. *locomotion, n. : Oxford English Dictionary*. The Oxford English Dictionary. URL: http://www.oed.com/view/Entry/109645?redirectedFrom=locomotion#eid (visited on 2019-05-07). *used on: p. 176*

[OED:robot] Oxford University Press. *robot, n.2 : Oxford English Dictionary*. The Oxford English Dictionary. URL: http://www.oed.com/view/Entry/166641?rskey=GRqkpy&result=2#eid (visited on 2019-02-12). *used on: pp. 15, 16, 177*

[Meka:head] *S3 Sensor Head*. Meka Robotics. URL: http://mekabot.com/products/s3-sensor-head/ (visited on 2012-04-24). *used on: pp. 52, 176*

[Meka:torso] *T2 Humanoid Torso*. Meka Robotics. URL: http://mekabot.com/products/humanoid-torso/ (visited on 2013-10-02). *used on: pp. 52, 176*

[Meka:tec] *Technology*. Meka Robotics. URL: http://mekabot.com/technology/ (visited on 2012-02-23). *used on: pp. 51, 53*

[OLD:haov] The Oxford Living Dictionaries. *handover | Definition of handover in English by Oxford Dictionaries*. URL: https://en.oxforddictionaries.com/definition/handover (visited on 2019-02-11). *used on: pp. 24, 25*

[OLD:robot] The Oxford Living Dictionaries. *robot | Definition of robot in English by Oxford Dictionaries*. URL: https://en.oxforddictionaries.com/definition/robot (visited on 2019-02-12). *used on: pp. 15, 177* [t.r.] *Understanding trajectory replacement*. Graphics with *CC BY 3.0* (https://creativecommons.org/licenses/by/3.0/). Open Source Robotics Foundation. URL: http://wiki.ros.org/joint_trajectory_controller/UnderstandingTrajectoryReplacement (visited on 2019-09-29). *used on: p. 87*

SOFTWARE PACKAGES

[BART] *Bielefeld Augmented Reality Tracker*. CITEC. URL: https://opensource.cit-ec.de/projects/bart/. *used on: pp. 68, 121*

[SMACH] Jonathan Bohren. *executive smach*. A procedural python-based task execution framework with ROS integration. ROS. URL: http://wiki.ros.org/smach (visited on 2019-08-14). *used on: p. 79*

[FlexBE] *FlexBE - The flexible behavior engine*. Philipp Schillinger. URL: http://philserver.bplaced.net/fbe/ (visited on 2019-08-14). *used on: pp. 79, 80, 108, 119, 126, 158*

[nodelet] Tully Foote and Radu Bogdan Rusu. *ROS nodelet*. designed to provide [..] zero copy transport between algorithms. Open Source Robotics Foundation. URL: http://wiki.ros.org/nodelet (visited on 2019-08-21). *used on: pp. 99, 121*

[Gazebo] *Gazebo - Robot simulation made easy*. Open Source Robotics Foundation. URL: https://github.com/RDTK (visited on 2019-08-14). *used on: pp. 53, 87*

[KDL] *Kinematics and Dynamics Library*. Orocos Kinematics and Dynamics. URL: https://orocos.org/kdl. *used on: p. 85*

[M3 Core] *M3 Core*. Meka Robotics LLC. URL: https://github.com/CentralLabFacilities/m3core (visited on 2019-03-20). *used on: p. 51*

[m3bie] *M3bie - Robot config*. CITEC - CLF. URL: https://github.com/CentralLabFacilities/m3bie. *used on: p. 53*

[self-filter] Eitan Marder-Eppstein. *Robot Self Filtering*. URL: http://wiki.ros.org/robot_self_filter. *used on: p. 101*

[ros_control] Wim Meeussen. *ROS Control*. "ros_control: A generic and simple control framework for ROS" [Chi+17]. Open Source Robotics Foundation. URL: http://wiki.ros.org/ros_control (visited on 2019-08-14). *used on: pp. 53, 84, 86, 87*

[MQTT] *Message Queuing Telemetry Transport*. Organization for the Advancement of Structured Information Standards. URL: http://mqtt.org/ (visited on 2018-05-10). *used on: p. 57*

[gaze_relay] Sebastian Meyer zu Borgsen and Luca Lach. *gaze forwarding with minimum duration*. URL: https://github.com/CentralLabFacilities/gaze_relay/. *used on: p. 108*

[MoveIt] *MoveIt Motion Planning Framework*. Ioan A. Sucan and Sachin Chitta. URL: https://moveit.ros.org/ (visited on 2019-04-10). *used on: pp. 42, 84, 86, 116, 124*

[Nuitrack] *Nuitrack Full Body Skeletal Tracking Software*. 3DiVi Inc. URL: https://nuitrack.com/. *used on: pp. 84, 90, 93*

[OpenFace] *OpenFace 2.2.0*. A facial behavior analysis toolkit [Bal+18]. CMU MultiComp Lab. URL: https://github.com/TadasBaltrusaitis/OpenFace (visited on 2019-06-25). *used on: pp. 108, 111*

[OpenPose] *OpenPose*. Multi-person keypoint detection library [Cao+17; Wei+16; Sim+17]. CMU - Perceptual Computing Laboratory. URL: https://github.com/CMU-Perceptual-Computing-Lab/openpose (visited on 2018-06-25). *used on: pp. 69, 79, 95, 98, 99, 105*

[RDTK] *[Research & Robotics] Development Toolkit (RDTK)*. CITEC. URL: https://github.com/RDTK (visited on 2019-08-14). *used on: p. 54*

[humotion] Simon Schulz. *[hu]man [motion] low-level robot control library [Sch+16]*. URL: https://github.com/CentralLabFacilities/humotion. *used on: pp. 75, 77, 80, 108*

[PPerception] *STRANDS people perception pipeline*. "Real-time multisensor people tracking for human-robot spatial interaction" [Don+15]. STRANDS. URL: https://github.com/strands-project/strands_perception_people (visited on 2019-08-14). *used on: p. 111*

[Jenkins] *the Jenkins project*. Kohsuke Kawaguchi. URL: https://jenkins.io (visited on 2019-08-14). *used on: p. 55*

[RTAI] *The RealTime Application Interface for Linux*. "Linux Real Time Application Interface (RTAI) in low cost high performance motion control" [DM03]. Dipartimento di scienze e tecnologie aerospaziali del Politecnico di Milano. URL: https://www.rtai.org/ (visited on 2019-10-15). *used on: pp. 51, 53*

[tf] *Transform Library*. "Tf: The transform library" [Foo13]. Open Source Robotics Foundation. URL: http://wiki.ros.org/tf (visited on 2019-08-14). *used on: pp. 53, 54, 68, 80, 84, 85, 121, 125*

[ROS URDF] *Unified Robot Description Format*. Ioan Sucan and Jackie Kay. URL: http://wiki.ros.org/urdf (visited on 2019-08-14). *used on: pp. 53, 86*

[Unity] *Unity 3D*. Unity Technologies. URL: https://unity3d.com/de/unity (visited on 2018-05-10). *used on: p. 57*

[VDemo] *VDemo*. cor-lab. URL: https://code.cor-lab.de/projects/vdemo/wiki (visited on 2019-08-14). *used on: p. 55*

DECLARATION OF AUTHORSHIP

According to the Bielefeld University's doctoral degree regulations §8(1)g: I hereby declare to acknowledge the current doctoral degree regulations of the Faculty of Technology at Bielefeld University. Furthermore, I certify that this thesis has been composed by me and is based on my own work, unless stated otherwise. Third parties have neither directly nor indirectly received any monetary advantages in relation to mediation advises or activities regarding the content of this thesis. Also, no other person's work has been used without due acknowledgment. All references and verbatim extracts have been quoted, and all sources of information, including graphs and data sets, have been specifically acknowledged. This thesis or parts of it have neither been submitted for any other degree at this university nor elsewhere.

Bielefeld, May 18, 2020

Sebastian Meyer zu Borgsen

COLOPHON

The style and layout of this thesis is based on a template created by Johannes Wienke and is typeset using the `classicthesis` typographical look-and-feel developed by André Miede [Classicthesis]. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*" [Bri04].