BEN-GURION UNIVERSITY OF THE NEGEV
THE FACULTY OF ENGINEERING SCIENCES
DEPARTMENT OF INDUSTRIAL ENGINEERING AND MANAGEMENT

HUMAN-ROBOT HANDOVERS: HUMAN PREFERENCES AND ROBOT LEARNING

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE M.Sc. DEGREE

By: Tair Faibish

Supervised by: Prof. Yael Edan and Dr. Armin Biess

January 2022

BEN-GURION UNIVERSITY OF THE NEGEV
THE FACULTY OF ENGINEERING SCIENCES
DEPARTMENT OF INDUSTRIAL ENGINEERING AND MANAGEMENT

HUMAN-ROBOT HANDOVERS: HUMAN PREFERENCES AND ROBOT LEARNING

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE M.Sc. DEGREE

By: Tair Faibish

Supervised by: Prof. Yael Edan and Dr. Armin Biess

Author:………………………………………… Date: 02/01/2022
Supervisor:………………………………….. Date: 02/01/2022
Supervisor:……………………… ………..… Date: 02/01/2022
Chairman of Graduate Studies Committee:………………… Date: 10/2/2022

January 2022

Acknowledgements

I would like to thank the following people without whom I would not have been able to complete this research.

Firstly, I am so grateful to have had the dedicated, caring, and open-minded Prof. Yael Edan as my supervisor. You have been a great mentor in mapping my journey, advising on research topics, connecting me with the resources I needed, and being available and responding to every single email and question. I am filled with sincere and heartfelt gratitude and appreciation for having had the privilege to work with you, and I consider you a role model for my future career and life.

Secondly, I would like to thank my second supervisor Dr. Armin Biess for all the guidance and assistance in the project. I cherish your wise counsel throughout the different phases of the project. You have been instrumental in my development as a scientist.

Thirdly, I would like to thank my colleague and non-formal mentor, Alap Kshirsagar. I am very thankful for all the support you have granted me throughout our mutual work on these two researches. I sincerely appreciate your continuous encouragement and guidance in the last two years. You pushed me to succeed and supported me at those inevitable times when I did not.

Fourthly, I would like to thank Moshe Bardea, Nissim Abuhazira, and Yossi Zahavi. I cannot thank you enough for the excellent technical support you provided me within the last couple of years.

Finally, profound thanks to my supportive husband, family and friends who escorted, heard, and motivated me throughout this research work.

Abstract

This thesis investigates the collaborative task of human to robot object handovers. Handovers are a vital capability for collaborative robots. We focused on two crucial issues for embedding human-like characteristics into robots. First, we examined the impact of robot's non-verbal communication on human's experience and fluency of human to robot handovers. Second, we developed and evaluated a robot controller based on reinforcement learning to perform a more natural sequential handover.

The first part in the research investigated human's preference of the robot's eye gaze during human-robot handovers. While there is some literature on robot gaze in robot-to-human handovers, there is a dearth of literature on robot gaze in human-to-robot handovers. Prior research that studied robot gaze behavior in human-to-robot handovers considered only the receiver's gaze patterns in the "reach" phase and used only one particular object in one configuration. Building upon this work, this research studied gaze patterns for all three phases of the handover process: reach, transfer, and retreat, both in video and in-person studies. This included investigation of whether the object's size and fragility or the giver's posture affect the human's preference of the robot gaze in terms of the perceived liking, anthropomorphism, and timing communication of the handover.

A public data-set of handovers videos were analyzed frame-by-frame to determine the most frequent gaze behaviors in human-human handovers. The most frequent gaze behaviors were found to be: gazing at the giver's hand and then at the giver's face (Hand-Face gaze), gazing initially at the giver's face and then at the giver's hand and then back to look at the giver's face (Face-Hand-Face gaze), and continuously look at the giver's hand (Hand gaze).

A Sawyer collaborative 7 DOF (degrees of freedom) robot was programmed to perform the handover task and exhibit these gaze behaviors. Different objects with different types of giver-receiver configurations were analyzed in two studies – a video study and an in-person study. In the video study, 72 participants watched and compared videos of human to robot handovers between an actor and a robot demonstrating the three gaze behaviors. In the in-person study, a different set of participants physically performed object handovers with the robot and evaluated their perception of the handovers for the robot's different gaze behaviors. Results revealed that for both studies when the robot initially gazes at the giver's face and then at the giver's hand and then back at the giver's face (Face-Hand-Face gaze), participants consider the handover to be more likable, anthropomorphic, and communicative of timing ($p < 0.005$). However, we did not find evidence of any effect of the object's size or fragility or the giver's posture on the gaze preference.

In the second part of the research, we assessed the potential of a model-based reinforcement learning (RL) method, the Guided Policy Search (GPS), to train a robot controller for human-robot object handovers. GPS is a data-efficient system that does not necessitate prior knowledge of the robot and environment dynamics, providing a promising approach for the handover task. Nevertheless, despite GPS demonstration on various navigation tasks and autonomous manipulation, testing GPS in a physical human-robot collaborative task has not been reported. In this study, the reach phase of a handover is formulated as an RL problem, with subsequent training of the Panda collaborative 7 DOF robot arm both in a simulation environment and directly on the physical robot.

Our results indicate that testing the policy learnt in the simulation environment on the real robot, is an infeasible solution for real world implementation. When estimating only static targets, we found that the performance of the global policies learnt by GPS generalize relatively well. However, the global policy performance got slightly improved by adding local controllers in regions with highest test errors. When evaluating the global policy trained with static targets on a moving target, the robot generated highly inefficient trajectories and reached areas outside of its cartesian position limits. Training on moving targets improved trajectories, but resulted with significantly larger worst-case errors. However, this issue can be addressed by adding local controllers to the training phase, improving the global policy’s performance.

Key Words: human-robot handovers, fluency, human-robot interaction, physical human-robot interaction, robot eye gaze, non-verbal communication, manipulation planning, reinforcement learning.

Publications

This research has yielded the following publications:

J1. **Faibish, T.**, Kshirsagar, A., Hoffman G., Edan, Y. 2022. Human Preferences for Robot Eye Gaze in Human-to-Robot Handovers. International Journal of Social Robotics, 1-18. https://doi.org/10.1007/s12369-021-00836-z

J2. **Faibish, T.**, Kshirsagar, A., Hoffman G., Biess A. Implementation and Evaluation of Guided Policy Search for Robot Reaching Towards Moving Targets. 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). And RAL. (Submitted, Appendix G).

Table of Contents

List of Appendices

List of Figures

Chapter 3

Chapter 4

Appendices

List of Tables

Chapter 3

Chapter 4

Chapter 1. Introduction

1.1 Overview

Until recent years, the traditional paradigm dominating human-robot interaction and collaborative robotics was keeping and operating robots in safety cages and separated from human operators. A recent review of the literature shows that this approach is increasingly being left behind, granting humans the capability to work alongside robots to complete various complex tasks (Magrini et al., 2020; Robotics, 2019; Tantawi et al., 2019).

One of the key challenges in these collaborative systems is coordination among the partners (Glasauer et al., 2010). Human-robot collaboration is often structured in a stop-and-go rigid regime of turn-taking operations inducing delays (Hoffman & Breazeal, 2009). For robots to become social or human-like in collaborative actions, robot-human interactions must reach a level of fluency, close to that of human-human interactions (Hoffman & Breazeal, 2007).

Object handover is one of the essential skills required for a collaborative or assistive robot hence, it is important for robots to carry out handovers autonomously. Tasks such as surgical assistance, housekeeping, rehabilitation assistance, and collaborative assembly require a robot to give objects to a human (robot-to-human handover) and take objects from a human (human-to-robot handover). This seemingly simple action involves coordination in both time and space of hand movements, grip forces, body postures, and other non-verbal cues like eye gaze.

There has been a constant growth of studies published regarding human-robot handovers over the year due to the importance and complexity of these tasks. This research focuses on two main gaps that remain understudied and unanswered. In the first part, with an agenda to excel robot's human-like characteristics, we implemented non-verbal communication in human-robot handovers. In the second part, we focused on motion planning of the reach phase of handovers. We implemented and evaluated a robot controller that uses Guided Policy Search (GPS) algorithm to perform object handovers with humans.

1.2 Background and problem description

1.2.1 Human-Robot interaction in industrial manufacturing

The industrial robotics field is expanding the usage of industrial robots, which used to be mostly in safety cages, by developing human-robot collaborative systems (Kuo, 2020; Hentout et al., 2019; Unhelkar et al., 2014; Duan et al., 2012). In order to achieve this goal, robots should be designed with attention to qualities such as flexibility and adaptability (Umbrico et al., 2020). Additionally, the robots should have the ability to share work and time space with humans (Fitzgerald, 2013). These types of robots can be used for a variety of tasks in different industries as multi-purpose robots which will work collaboratively with human workers. Examples of those industries are assisting in the assembly of complicated objects or in the packaging of products with various sizes, weights and shapes (Cherubini et al. 2016; Tsarouchi et al., 2016). Rethink Robotics Sawyer and Baxter (Fitzgerald, 2013), and Panda (Franka Emika GmbH, 2020) robots are considered to be among the new robotic developments promoting this approach.

According to recent HRI studies, humans tend to interact with computers in social ways (Reeves & Nass, 1996; Sproull et al., 1996). Humanlike stimuli are more likely to evoke social responses than machinelike stimuli, because people have a propensity to seek an embodiment for intelligence and a social locus of attention (Cassell, 2001). In support of this argument, Sproull et al. (Sproull et al., 1996) showed that explicit humanlike cues such as a humanlike face presented on a computer screen as compared with a text-based computer led people to make stronger attributions of personality to the computer, present themselves more positively to it, and feel more relaxed and assured by the computer. These results suggest that humanlike cues provide a sense of presence and disambiguate what communicative channels are open to people (e.g., speech, gaze, facial expressions, gesture), making communication more fluent and allowing people to have a more certain mental representation of the computer (Kiesler, 2005).

1.2.2 Handovers

A handover is a complicated collaboration. In order to transfer control of an object, actors are required to be coordinated in time and space. Handovers are an integral part of our day to day, for example: a caregiver bringing a patient a glass of water, a mechanic receiving a tool from his assistant, someone passing a bucket of water as part of a fire brigade, and a man on the sidewalk handing over a flyer to a busy passer-by. People rely on understanding the context of events and communication cues with those around them for successful handovers. The situation that surrounds an action is called its context, it provides the knowledge for people on the ways to interpret others' behaviors and to know what to expect from others. Through communication humans establish the what, when, and where/how of the handover. For example: the mechanic establishes the "what" by asking for a certain tool from his assistant and using context (the assistant is nearby) to expect a handover, when a caregiver and a patient exchange looks to establish the" when" they are ready to reach out and transfer the glass (Strabala et al., 2013).

1.2.3 Structure of the handover process

The handover process comprises a physical channel and a social cognitive channel (Strabala et al., 2013). The physical channel is subdivided into three distinct phases (Strabala et al., 2013). In the "approach" phase, both participants are heading towards each other. In the "reach" phase, both participants spread their arms to the estimated handover location, and in the "transfer" phase, the object is exchanged between the giver and the receiver, who then exit the joint activity. The social cognitive channel provides the context needed for a fluent handover to occur. For example, where and when the "transfer" phase will take place, and the establishment of the handover's object.

Basili et al. (Basili et al., 2009) showed that the three handover phases: approach, reach and transfer are dynamic actions that blend seamlessly into each other, rather than separate and consecutive actions. Nevertheless, researchers supporting a reductionist approach have analyzed the handover phases as three distinct phases to get a methodical understanding of the characteristics and actions of each phase.

1.2.4 Human-Human handovers

Handovers are complex interactions, yet humans are capable of performing handovers seamlessly and without conscious thought (Roy & Edan, 2017) . This suggests that people share a common procedure that guides the handover interaction. Experiments conducted to examine how people hand over objects to each other (Strabala et al., 2013) revealed a structure consisting of carrying (approaching with the object), signaling a readiness to do a handover, and transferring the object. In 89% of the cases, the exact time when an actor starts reaching can be predicted from communication cues that the actor uses right before the act, meaning the communication between humans is so rich that signaling a readiness to do a handover can happen before either actor starts reaching out. The experimenter reported that the cues came mainly from facial expressions, gestures, and eye movement.

1.2.5 Robot-to-Human handovers

In order to understand the best way in which a robot should approach a human to initiate a handover, many studies have been conducted (Basili et al., 2009; Koay et al., 2007; Mainprice et al., 2012; Sayfeld et al., 2017; Someshwar et al., 2012, Someshwar, 2017). Basili et al. (Basili et al., 2009) examined the way a human giver, with the purpose of handing over an object, carries the object and approaches a human receiver. They noted that their findings could be transferred to a robot giver. Koay et al. (Koay et al., 2007) investigated the interaction when a robot hands over a can to a human, and specifically, human preference of robot coordination during these handovers. Preferences such as the preferable distance from the human receiver in which the robot should stop, and the direction of approach. In this study, the human receiver approached the robot, which was at a fixed position. Still, advice concerning the positioning of the robot is offered in the above studies.

Others (e.g., Cakmak et al., 2011; Cakmak et al., 2011; Dehais et al., 2011; Edsinger et al., 2007; Huber et al., 2008) have shown that handover quality is affected by the route and the configuration or pose of a robot. Edsinger and Kemp (Edsinger et al., 2007) showed that subjects understood the robot's intention during a handover by the robot's approaching motion, even without vast knowledge in robotics or exact directions. Cakmak et al. (Cakmak et al., 2011) proved that handover intent also relies on handover poses. They showed that inadequately designed handover poses might fail to carry handover intent. Creating a distinct difference between the handing the object pose and the holding the object pose was their proposition to solve this issue. A different research (Cakmak et al.,2011) suggested a handover configuration that best conveys the handover intent. This configuration is composed along three Cartesian axes and includes an almost entirely extended arm with a persistently monotonic configuration of the distal tip of the object and the robot's elbow and wrist joint. Our work employed findings from the above studies in the design of our robot's handover trajectory and configuration.

1.2.6 Gaze cues in social interaction

During social interaction, people spend more time looking at others (an average of 61% of the interaction's time) than speaking (Argyle & Ingham, 1972). People study others' behavior by gazing at others and, particularly, by looking in their eyes region (Cook, 1977).

The function of eye gaze in human social interaction is versatile. One can both perceive information from other humans, and signal to others using his gaze (Argyle & Cook, 1976; Cañigueral & Hamilton, 2019; Gobel et al., 2015; Risko et al., 2016). Simmel already stated that "*the eye cannot take unless at the same time it gives*." (Simmel, 1921). This is contrary to auditory modality, where we use our ears to hear, but our mouth to speak. This makes our eyes a powerful tool for social interactions, with a "*uniquely sociological function*" (Simmel, 1921).

For any social interaction to be initiated and maintained, parties must establish eye contact. Through establishing eye contact, people form "an ecological eye-to-eye huddle" through which they signal each other that they agree to engage in social interaction (Goffman, 1963). Simmel (Simmel, 1921) describes this mutual behavior as "*a wholly new and unique union between two people [that] represents the most perfect reciprocity in the entire field of human relationship*".

People are extremely sensitive to being looked at (Gibson & Pick, 1963). The detection of direct eye contact is a crucial element for survival, as it can manifest predator's intentions for an attack. That may explain the evolving human's sensitivity to it (Emery, 2000). A designated 'eye direction detector' in human's brain is postulated to support that kind of mechanism, according to neurophysiological proofs (Baron-Cohen, 1995). Human's decision-making manners were found to be influenced not only by pictures of eyes (Bateson et al., 2006), but also by imitated "eyespots" on a computer screen (Haley & Fessler, 2005). Pedestrians who engage drivers, using their gaze, have better chances to get stopped for on the road (Mutlu, 2009; Snyder, Grather, & Keller, 1974). The tight coupling between gaze behavior and many other aspects of social interaction has made the study of gaze behavior central to social psychology (Mutlu, 2009). Argyle and Cook argue, *"Any account of social behavior which fails to deal with the phenomena of gaze is quite inadequate"* (Argyle & Cook, 1976).

1.2.7 Controllers for human-robot handovers

Handovers possess a substantial role in physical human-robot interactions. Following the realization of this concept, numerous studies regarding robot controllers for handovers have been published. These controllers utilize various sensor interfaces, e.g., wearable devices, visual sensors and physical sensors (Leal & Yihun, 2019). Several methods for controlling the robot in the different handover phases exist today. For the handover's reach phase, robot controllers can be subclassified as either offline or online. Offline controllers determine the motion plan of the robot prior to the initiation of the reach phase without further adjustments during the reach phase. In comparison, online controllers take into consideration the perceived behavior of the human while continuously updating the robot's motion plan during the reach phase.

1.2.8 Guided policy search

Reinforcement Learning (RL) is a subfield of machine learning. The RL methods let the agent use the rewards received in the interaction with the environment for learning the control policy (Du et al., 2021). In recent years, it has developed rapidly, achieving profitable results in sequential decision-making problems like robot learning (Kaelbling, 2020). Guided policy search (GPS) is one of the well-established RL methods developed over the years and is used in various robot's manipulation (Chebotar et al., 2017; Levine et al., 2016; Levine et al., 2015; Levine & Abbeel, 2014), and locomotion (Zhang et al., 2016; Levine & Abbeel, 2014;Levine & Koltun, 2013, Levine & Koltun, 2013b) tasks.

The GPS (Levine et al., 2014; Levine et al., 2015; Levine et al., 2016) method employs trajectory optimization methods to instruct the optimization of neural network policy parameters without encountering the local optimal dilemma. The sample efficiency is enhanced by the trajectory optimization methods with learned dynamics. Benefitting from the great framework, GPS can employ a more general neural network to parameterize the policy, increasing its ability to express and generalize without damaging the data's efficiency (Du et al., 2021).

Most of the commercial robots and custom-built robots' dynamics are unknown, partly because these parameters may be difficult to obtain. One method to deal with this challenge is the implementation of system identification techniques to develop dynamical models. However, this requires extensive training data, notably for formulating global models of complex systems (Ibarz et al., 2021). Hence, GPS is a data-efficient system that does not necessitate prior knowledge of the robot and environment dynamics, providing a promising approach for the handover task.

1.3 Objectives

This thesis investigates on two crucial issues in the collaborative task of human to robot object handovers. First, we examined the impact of robot's non-verbal communication on human's experience and fluency of human to robot handovers. Second, we develop and evaluate a robot controller based on reinforcement learning to perform a more natural sequential handover.

The main objective of the first study of the research is to investigate how different eye gaze behaviors of a robot receiving an object from a human influence the perceived liking, anthropomorphism, and timing of the handover. The specific research objectives in this first study are to investigate:

1. Human-Human joint-actions in handover tasks for developing H-R collaborative systems for handover tasks.
2. Parameters affecting Robot-to-Human handover actions:
    - The robot's eye gaze pattern for better H-R team coordination and improved system productivity in handover tasks.
    - Investigate if the object size / fragility affects the user's ratings of the robot's gaze in a human-to-robot handover.
    - Investigate if the human-robot configuration affects the user's ratings of the robot's gaze in a human-to-robot handover.

The second study in this thesis aims to implement and evaluate a robot controller that uses Guided Policy Search (GPS), a model-based reinforcement learning (RL) method to perform object handovers with humans. We investigate how does GPS perform with large variations in target locations, moving targets, with a physical robot and compare training in a simulation environment with training conducted directly on the physical robot.

1.4 Thesis overview

The overall research methodology is depicted in chapter 2. The research includes two separate parts corresponding to two gaps in the handover process: implementing non-verbal communication in human-robot handovers (study 1, chapter 3), and motion planning of the reach phase of handovers (study 2, chapter 4). Conclusions and future research are discussed in chapter 5.

Chapter 2. Methodology

The overall methodology is presented in this chapter. This includes the research questions regarding human-robot interaction during handover tasks. In the first study we examined what are the most frequent gaze behaviors in a human-human handover. Then, with the purpose of implementing these behaviors on a collaborative robot, we investigated whether and to what extent the user's preference of the robot's gaze, when it is receiving an object from the human, and is this dependent on the object size and type and on different human-robot configurations. In the second study we developed a robot controller that uses Guided Policy Search (GPS) to perform object handovers and evaluated the effect of different training scenarios (simulation and physical robot) on performance.

2.1 Study 1: Human Preferences for Robot Eye Gaze in Human-to-Robot Handovers

This study aims to investigate how the gaze behaviors of a robot, receiving an object from a human, affect the human's subjective experience of a handover. Details are provided in Chapter 3 and in publication J1. Previous research that studied robot gaze behavior in human-to-robot handovers has only considered the receiver's "head gaze" behaviors in the "reach" phase and used only one particular object in one configuration (i.e., they only used a bottle of water as the object, and only considered situations in which the person was standing, Kshirsagar et al., 2020).

In this study, gaze patterns for all three phases of the handover process: reach, transfer, and retreat were considered for different objects with a different type of giver-receiver configuration. First, to identify the most frequent gaze behaviors in a handover, a frame-by-frame video analysis of a public data-set of human-human handovers (Carfì et al., 2019) was performed. The database consists of over 1000 videos of object handovers with 18 volunteers, 10 objects, and several handover scenarios. The handover scenarios vary in terms of experiment type (volunteer-volunteer or volunteer-experimenter), role of the volunteer (giver or receiver), and starting position (with approach or without approach). For video analysis, we only considered the volunteer-volunteer handovers as these would be more natural. This yielded a total of 288 videos recorded at 8pfs with a resolution of 1280X720 pixels. These videos were analyzed for both the givers and receivers and included a total of 18 people. We used video analysis even though the dataset contained motion capture data for participants' heads, as we found that gaze is often enacted only with the participant's eyes, without noticeable head-movement. Analysis of videos of human-human handovers provided three candidate gaze patterns that were implemented on the robot:

1) **Hand-Face**: Initially look at the other person's face and then at the other person's hand. The duration of *Hand* gaze is 70% of the total duration of the handover.
2) **Face-Hand-Face**: Initially look at the other person's face and then at the other person's hand and then back to look at the other person's face.
   The total duration of the handover is divided as follows: the first *Face* gaze (15%), the *Hand* gaze (55%), and then back again to *Face* gaze (30%).
3) **Hand gaze**: Continuously look at the other person's hand.

We performed two types of user studies (video and in-person) with a collaborative robot that exhibited these gaze behaviors while receiving an object from a human. The robot arm was autonomous and programmed to reach a predefined position once the handover began. The robot grasped the object when the object was close enough. Finally, the robot retreated to its home position after the human released the object and started to retreat.

The system includes a robot arm receiving an object from a human, a distance sensor to detect the giver's movement, and an infrared proximity sensor placed on the robot arm to detect the object distance from the robot gripper. The sensors are controlled by an Arduino microcontroller. We used Rethink Robotics' Intera SDK to program the robot, and Robot Operating System (ROS) framework to connect all the components. Details of the Robotic system development are provided in Appendix A.

To investigate the effect of object's size, object's fragility or the human's posture on human's preferences for the robot gaze, objects of different sizes (a small box and a large box), different fragility (a plastic bottle and a glass bottle) and different giver's posture (standing and sitting) were used. In order to examine people's perception about the fragility of these objects, we conducted an online survey. Details are provided in Appendix B. Ten different objects were used in the human-to-human handovers videos, and three gaze patterns were received. Therefore, we chose to examine whether the type of objects affected the human preferences of robot gaze in human-to-robot handovers.

We preformed statistical analysis using the one-sample Wilcoxon signed-rank tests, the Bradley-Terry model, and also conducted Binary proportion difference tests and Equivalence tests[1].

A repetitive observation was attained in our open-ended responses regarding the preferred robot gaze (open-ended responses are presented in Appendix C). Participants favored the robot gaze perceived with the most human-like characteristics. This fact directed our search for additional ways to anthropomorphize our robot, generating a more intuitive human-robot interaction. After a thorough inspection of the literature, we found that other key components of HRI, which may influence humans, are the perceived naturalness and smoothness of the robot's movements. Therefore, we decided to pursue our second study regarding human-robot handover, implementing an online controller to produce reaching motion of the robot to further develop the acceptance and practical use of collaborative robots in the industry.

2.2 Study 2: Guided Policy Search for Human-Robot Handovers for Human-Robot Handovers in a Real-World Environment

This study aims to evaluate the potential of a model-based RL method, Guided Policy Search (GPS), to train a robot controller for human-robot object handovers both in a simulation environment, and directly on a physical robot. Details are provided in Chapter 4.

The controllers available nowadays for human-robot handovers necessitate precise robot kinematic/dynamic models. Moreover, tuning controller parameters which are non-intuitive, i.e., weights of movement primitives or velocity tracking gain, is required. To address these issues, we used a "Guided Policy Search (GPS)" (Levine et al., 2014; Levine et al., 2015; Levine et al., 2016) to generate an online handover controller which does not necessitate tuning non-intuitive controller parameters or the robot's dynamic/kinematic models. Also previous research evaluating GPS for human-robot handover was merely conducted in a simulation environment, without implementation on a real robot (Kshirsagar et al., 2021). The application in a real-world context is important.

[1] The equivalence test was added based on a request of a single reviewer of the IJSR paper; we are not convinced it should have been used in this type of rsearch.

We formulated the reach phase of a human-to-robot handover as a policy search problem. The system state representation consists of the robot's joint angles, joint velocities, object's positions and velocities, and the human's hand in relation to the robot end-effector, and load-share estimate (proportion of the weight of the object supported by the robot). The control input consists of torques applied to the robot's joints. We define a multi-modal cost function for the handovers task that rewards the robot's movement towards the human hand only if the human hand is moving towards the robot.

We evaluated the controller with a physical robot while the training was conducted both in a simulation environment and directly on the physical robot. The physical robot used to perform handover reaching motions in a real environment was the Panda robot. Panda Robot is a sensitive and agile 7 DOF arm with torque sensors at each joint, allowing adjustable stiffness/compliance and advanced torque control. We used a physics engine called MuJoCo (Multi-Joint dynamics with Contact) (Todorov et al., 2012) to train the robot to perform handover reaching motions in simulation and then tested the policy in real environment on a Panda robot. The performance of the global policy was measured in terms of the error between the human hand's position and the end-effector's position.

In the first experiment, we wanted to test the policy learnt in the simulation environment on the real robot. To do so, we tried to tune the MuJoCo model parameters to match the real robot parameters. It was proved to be an infeasible solution and did not achieve operational results. Thus, we decided to train the physical robot instead of a simulated robot, with a simulated target.

In the second experiment, we trained and tested the real Panda collaborative robot to perform handovers over repeated trials for two scenarios: large variations in target locations and moving targets. We used recorded human hand motions in all training iterations during the training/testing process. The region for training and testing was selected by trial and error to ensure that the robot does not run into joint position/velocity limits in the training/testing process. For each angle in 5deg increments, we tested it on a grid of 3×3 targets, resulting in 90 test locations.

The first research question examined in our study was how does the GPS perform for significant spatial differences between training and testing locations. We compared two scenarios of local controllers: one with 8 local controllers and another with 12 local controllers.

The second research question examined in our study was how does the GPS perform with moving targets. First, we used the global policy trained with static targets, but instead of a static tester, we used a recorded human reaching motion. In this case, the robot generated highly inefficient trajectories and reached areas outside of its cartesian position limits. To address this issue, we trained the robot with moving targets (recorded human reaching motions), and tested the policy on another set of recorded human's reaching motions.

Chapter 3. Human Preferences for Robot Eye Gaze in Human-to-Robot Handovers

Faibish, T., Kshirsagar, A., Hoffman G., Edan, Y. 2022. Human Preferences for Robot Eye Gaze in Human-to-Robot Handovers. *International Journal of Social Robotics*, 1-18.

International Journal of Social Robotics
https://doi.org/10.1007/s12369-021-00836-z

Human Preferences for Robot Eye Gaze in Human-to-Robot Handovers

Tair Faibish[1] · Alap Kshirsagar[1,2] · Guy Hoffman[2] · Yael Edan[1]

Accepted: 5 October 2021

Abstract
This paper investigates human's preferences for a robot's eye gaze behavior during human-to-robot handovers. We studied gaze patterns for all three phases of the handover process: reach, transfer, and retreat, as opposed to previous work which only focused on the reaching phase. Additionally, we investigated whether the object's size or fragility or the human's posture affect the human's preferences for the robot gaze. A public data-set of human-human handovers was analyzed to obtain the most frequent gaze behaviors that human receivers perform. These were then used to program the robot's receiver gaze behaviors. In two sets of user studies (video and in-person), a collaborative robot exhibited these gaze behaviors while receiving an object from a human. In the video studies, 72 participants watched and compared videos of handovers between a human actor and a robot demonstrating each of the three gaze behaviors. In the in-person studies, a different set of 72 participants physically performed object handovers with the robot and evaluated their perception of the handovers for the robot's different gaze behaviors. Results showed that, for both observers and participants in a handover, when the robot exhibited *Face-Hand-Face* gaze (gazing at the giver's face and then at the giver's hand during the reach phase and back at the giver's face during the retreat phase), participants considered the handover to be more likable, anthropomorphic, and communicative of timing ($p < 0.0001$). However, we did not find evidence of any effect of the object's size or fragility or the giver's posture on the gaze preference.

Keywords Human-robot handovers · Human-robot interaction · Robot eye gaze · Human-human-handovers · Non-verbal communication

1 Introduction

People frequently hand over objects to others or receive objects from others. Robots in domestic and industrial environments will be expected to perform such handovers with humans. For example, collaborative manufacturing (e.g., assembly), surgical assistance, household chores, shopping assistance, and elder care involve object handovers between the actors. In this work, we investigate where should a robot direct its gaze when it is receiving an object from a human.

A handover typically consists of three phases [1]: a reach phase in which both actors extend their arms towards the handover location, a transfer phase in which the object is transferred from the giver's hand to the receiver's hand, and a retreat phase in which the actors exit the interaction. These phases involve both physical and social interactions consisting of hand movements, grasp forces, body postures, verbal cues and eye gazes.

Most of the research on human-human and human-robot handovers has focused on arm movement and grasping in handovers, with only a few works that studied the social inter-

Tair Faibish and Alap Kshirsagar contributed equally to this work.

This work was part of Tair Faibish Engineering Final Project and MSc thesis.

✉ Tair Faibish
tairse@post.bgu.ac.il

Alap Kshirsagar
ak2458@cornell.edu

Guy Hoffman
hoffman@cornell.edu

Yael Edan
yael@bgu.ac.il

1 Department of Industrial Engineering and Management and the ABC Robotics Initiative, Ben-Gurion University of the Negev, Be'er Sheva, Israel

2 Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, USA

Published online: 21 January 2022

actions. Eye gaze is an important non-verbal communication mode in human-human and human-robot interactions, and it has been shown to affect the human's subjective experience of human-robot handovers [2–6]. However, except for our previous work [6], all of the prior studies of gaze behaviors in handovers considered only the robots-as-givers scenario i.e. robot-to-human handovers. Human-to-robot handovers are equally important with many applications in various domains. Some examples include a collaborative assembly task in which the robot receives parts from the human or an elder care robot that takes an empty tray from an older adult after giving him/her food.

In our previous work [6], we studied the effects of robot head gaze during the reach phase of human-to-robot handover. Results revealed that observers of a handover perceived a *Face-Hand transition* gaze, in which the robot initially looks at the giver's face and then at the giver's hand, as more anthropomorphic, likable and communicative of timing compared to continuously looking at the giver's face (*Face* gaze) or hand (*Hand* gaze). Participants in a handover perceived *Face* gaze or *Face-Hand transition* gaze as more anthropomorphic and likable compared to *Hand* gaze. However, these results were limited to a specific scenario where the giver stood in front of the robot and handed over a specific object (a plastic bottle) to the robot. Furthermore, the robot's gaze behaviors were studied only in the reach phase of the handover.

The goal of this paper is to expand and generalize the findings from our previous work. Here, we study the human's preference for robot gaze behaviors in human-to-robot handovers for all three phases of a handover for four different object types and two giver postures. Also, we use eye gaze instead of head gaze since it is more common. We also contribute to the literature on human-human handovers by identifying common gaze behaviors of humans in handovers.

2 Related Work

2.1 Human-to-Robot Handovers

Researchers have studied human-to-robot handovers to understand human preferences for robot behaviors in the approach, reach and transfer phases of handovers. In this work, we use the findings from these studies to design the robot's handover trajectory and configuration.

Investigation of the interaction of a robot handing over a can to a human [7] revealed that the preferred interpersonal distance between the human and the robot is within personal distance (0.6m - 1.25m), suggesting that people may treat robots similar to other humans. Previous research also showed that subjects understood the robot's intention during a handover by the robot's approaching motion, even without prior knowledge in robotics or exact directions [8]. Furthermore, Cakmak et al. [9] found that handover intent also relies on handover poses, and inadequately designed handover poses might fail to convey the handover intent. Their recommendation was to create the handover pose distinct from the object holding pose. They also suggested that the best handover intent is conveyed by an almost extended arm [10]. A study of effect of participant's previous encounters with robots on human-robot handovers showed that naive users, as opposed to experienced ones, expect the robot to monitor the handover visually, rather than merely use the force sensor [11]. A study of the impact of repeated handover experiments on the robot's social perception [12] showed that participants' emotional warmth towards the robot and comfort were improved by repeated experiments.

2.2 Gaze in Handovers

There is surprisingly little work on gaze behaviors in human-to-human handovers or object passing tasks [6]. Flanagan et al. [13] investigated gaze behavior in a block stacking task. Contrary to previous assumptions, they showed that human gazes were not reactive during the task i.e. people did not focus on the gripped object or the object in movement. Instead, human gazes were found to be predictive; their gazes focused on the object's final destinations. Investigation of the discriminative features that represent the intent to start a handover revealed that mutual gaze during the task, which is often considered crucial for communication, was not a critical discriminative feature [14]. Instead, givers' initiation of a handover was better predicted using asynchronous eye gaze exchange.

In a human-to-human handover study of a water bottle [2], it was found that the givers exhibited two types of gaze behaviors: shared attention gaze and turn-taking gaze. In shared attention gaze, the giver looked at the handover location, and in turn-taking gaze, the giver initially looked at the handover location and then at the receiver's face. In our prior work [6], we found that the most common gaze behavior for both the giver and the receiver was to continuously look at the other person's hand during the reach phase of a handover. Receivers exhibited this behavior almost twice as frequently as the givers. However, our prior work studied the gaze behaviors only in the reach phase of human-to-human handovers. To the best of our knowledge, there is no prior work that studies both the giver's and the receiver's gaze in all three phases of the handover process: reach, transfer, retreat. This gap is addressed in Sect. 3.3.

Past research revealed that robot gaze affects the subjective experience and timing of robot-to-human handovers [2–5,15]. A "turn-taking gaze" in which the robot switched its gaze from the handover location to the receiver's face halfway through the handover was favoured [2]. In a followup study, results revealed that the participants reached for the object sooner when the robot exhibited a "face gaze" i.e. continuously looked at receiver's face, as opposed to a shared attention gaze [3]. Fischer et al. [4] assigned a robot to retrieve parts according to participants' directions and compared two robot gaze behaviors during this task. They found that when the robot looked at the person's face instead of looking at it's own arm, participants were quicker to engage with the robot, smiled more often, and felt more responsible for the task. In a similar study, [5] it was found that when the robot looked at the participant's face while approaching them with an object, it significantly increased the robot's social presence, perceived intelligence, animacy, and anthropomorphism. Admoni et al. [15] used the robot's gaze behavior to instruct the human to place the handed-over object at a specific location. They showed that delays in the robot's release of an object draws human attention to the robot head and gaze and increases the participants' compliance with the robot's gaze behavior. In our prior work [6], we found that observers of a human-to-robot handover preferred a transition gaze in which the robot initially looked at their face and then at their hand during the reach phase. For participants in human-to-robot handovers, a face gaze was almost equally preferred as a transition gaze, though the evidence was statistically weaker.

A common limitation of these prior studies is that they do not investigate the effect of the object or the human's posture on the human's preference of robot gaze. Therefore, in the current study, as described in Sections 4–5, human preferences towards robot gaze behaviors in human-to-robot handovers for four different object types and two human postures is compared.

3 Methodology

3.1 Overview

This research aims to investigate human preferences for robot gaze behaviors in human-to-robot handovers for all three phases of the handover process (reach, transfer and retreat). To obtain possible options for robot gaze behaviors we first studied gaze behaviors in human-to-human handovers. A data set of videos of human-human handovers was analyzed, and the most common gaze behaviors of receivers were identified. Informed by this analysis, we conducted two user studies of the robot's gaze while receiving the object from the human in different situations. We investigated whether different object types or giver's postures affect the human preferences of robot gaze in human-to-robot handovers.

3.2 Hypotheses

The research hypotheses are:

- *H1*: People prefer certain robot gaze behaviors over others in terms of likability, anthropomorphism and timing communication.
- *H2*: Object size affects the user's ratings of the robot's gaze in a human-to-robot handover.
- *H3*: Object fragility affects the user's ratings of the robot's gaze in a human-to-robot handover.
- *H4*: User's posture (standing and sitting) affects the user's ratings of the robot's gaze in a human-to-robot handover.
- *H5*: Observers of a handover and participants in a handover have different preference ratings of the robot's gaze in a human-to-robot handover.

H1 is motivated by prior work which found evidence for different user preference ratings for robot gaze behaviors. We do not have a-priori hypothesis about the preference order of gaze behaviors. **H2** and **H3** are based on the intuition that the object's size and fragility could affect the preferred gaze behavior of a receiver. For example, when receiving large or fragile objects, the robot could be expected to convey attentiveness by looking at the giver's hand, whereas, when receiving small or non-fragile objects, the robot could be better off looking at the giver's face to convey friendliness. **H4** is based on the intuition that a standing giver may have different preferred gaze behavior of a receiver than a sitting giver. For example, a standing person could like the robot gaze at their face as their eyes are at the same level, whereas a sitting person could feel uncomfortable with the robot gazing down at their face. **H5** results from our previous finding that observers of a handover and participants in a handover had different preference ratings of robot gaze behaviors in the reach phase [6]. This research examines whether this holds true for robot gaze behaviors in all three phases of a handover and for handovers with different object types and giver postures.

3.3 Analysis of Gaze in Human-Human Handovers

We analyzed gaze behaviours in human-to-human handovers by annotating all three phases of each handover in a public dataset of human-human handovers [16], similar to our previous work [6]. A frame-by-frame video encoding was performed followed by annotating the giver's and receiver's gaze locations in each phase in each frame with the following discrete variables{G: Giver, R: Receiver}[1]:

1) **G's gaze:** R's face/R's hand/Own Hand/Other
2) **G's phase:** Reach/Transfer/Retreat
3) **R's gaze:** G's face/G's hand/Own Hand/Other
4) **R's phase:** Reach/Transfer/Retreat

[1] The annotations are available at: https://github.com/alapkshirsagar/handover-gaze-annotations/.

(a) (b) (c)

Fig. 1 Examples of gaze annotations of the human-human handovers dataset [16]. On the left is the giver and on the right the receiver: **a** Reach phase : The giver is gazing at the other's face while the receiver is gazing at the other's hand, **b** Transfer phase : Both the giver and receiver are gazing at the other's hand, **c** Retreat phase: Both the giver and the receiver are gazing at the other's face

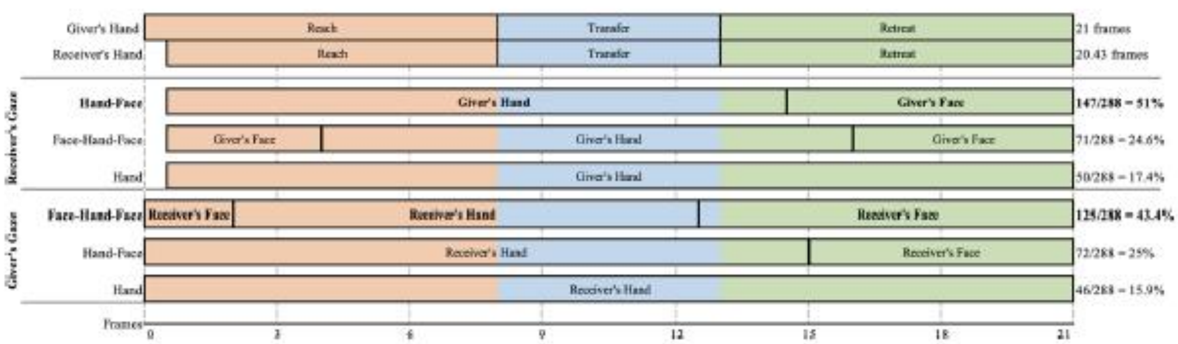


Fig. 2 Analysis of gaze behaviors in the reach, transfer and retreat phases of human-human handovers. Time flows left to right. Background colors (labeled on top two rows) correspond to each phase of a handover: red: reach; blue: transfer; green: retreat. The bottom six rows show one handover behavior each, three for the receiver and three for the giver. Boundaries correspond to average length of each phase. Prevalence of each behavior is noted at the right edge of the row. Givers and receivers have dissimilar frequently observed gaze behaviors

Figure 1 shows some examples of gaze annotations in the three phases of handovers. The analysis (Fig. 2) revealed that the most common gaze behaviors employed by people during handovers are:

1) *Hand-Face* gaze: The person continuously looks at the other person's hand during the reach and the transfer phases, and then looks at the other person's face during the retreat phase. The transition from hand to face happens slightly after the beginning of the retreat phase. More than 50% of receivers showed this behavior, whereas, only 25% of the givers in those videos exhibited this behavior.

2) *Face-Hand-Face* gaze: During the reach phase, the person initially looks at the other person's face and then at the other person's hand. They then continue looking at the other person's hand during the transfer phase. Finally they look at the other person's face during the retreat phase. The transition from face to hand occurs halfway through the reach phase, while the transition from hand to face occurs halfway through the retreat phase. More than 40% of givers exhibited this gaze, whereas only 25% of receivers did.

3) *Hand* gaze: Continuously looks at the other person's hand. The least frequent gaze, only 17.4% of receivers and 15.9% of givers showed this behavior.

3.4 Human-Robot Handover Studies

Two within-subject studies were conducted, a video study and an in-person study. The video study aimed to investigate an observer's preferences of robot gaze behaviors, whereas the in-person study aimed to investigate a giver's preferences of robot gaze behaviors.

A total of 144 undergraduate industrial engineering students participated in the experiment (72 in each study) and were compensated with one bonus point to their grade in a course for their participation. The average participation time was about 25 minutes. In the video study, there were 34 females and 38 males aged 23-29. In the in-person study, there were 36 females and 36 males aged 23-30. The study design was approved by the Human Subjects Research Committee at the Department of Industrial Engineering and Management, Ben-Gurion University of the Negev.

The following three gaze behaviors were implemented on a Sawyer cobot based on insights from the human-human handover analyses:

i. *Hand-Face* gaze: The robot's eyes continuously looked in the direction of the giver's hand during the reach and transfer phases. After the robot started to retreat, the eyes transitioned to look at the giver's face. Both the hand gaze and the face gaze were programmed manually to fixed locations.

ii. *Face-Hand-Face* gaze: The robot's eyes looked at the giver's face during the reach phase, giver's hand during the transfer phase and giver's face during the retreat phase.

iii. *Hand* gaze: The robot's eyes continuously looked in the direction of the giver's hand.

Given that the human gaze behavior was tied to the handover phase, as described above, we did not use fixed timings for the robot trajectory. Instead, the robot was programmed to use sensor information to initiate the handovers and gaze behaviors depending on the phase of the handover. The robot arm was programmed to reach a predefined position once the giver started the handover which was detected using a range sensor. The robot's gripper was equipped with an infrared proximity sensor, and it grasped the object when the object was close enough. The robot retreated to its home position after grasping the object. The robot was programmed in the Robot Operating System (ROS) environment with Rethink Robotics' Intera software development kit (SDK). The sensors were interfaced with the robot using an Arduino micro-controller.

Figure 3a shows a snapshot of a video recording illustrating the experimental setup.[2]

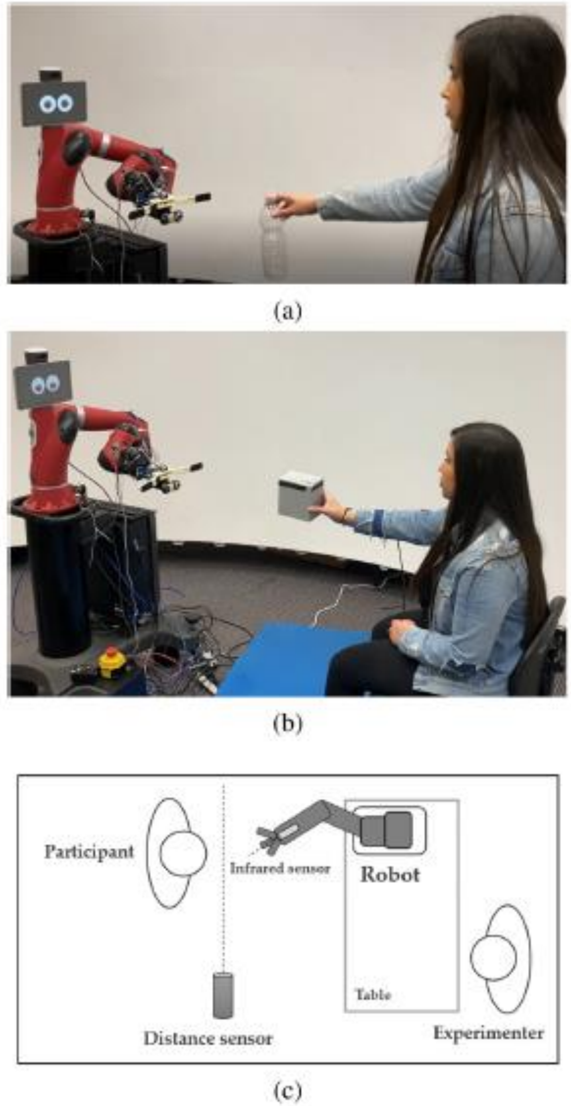


Fig. 3 Experimental Setup: Video frames of an actor handing over an object to the robot, used in the video study: **a** "Standing" posture **b** "Sitting" posture **c** Diagram of the setup for the in-person study

4 Video Study of Human-to-Robot Handovers

4.1 Experimental Procedure and Evaluation

The study was conducted remotely, and each participant received links to the videos, electronic consent form, and online questionnaires with study instructions. After signing the consent form and reading the instructions, they completed a practice session followed by 12 study sessions. Each ses-

[2] The videos are available at: https://youtu.be/9dD1YHG2Nco.

sion included one of the six pairing of the gaze patterns listed in Table 2, for a single condition out of the three listed in Table 1[3]. So that each participant watched all six pairs of gaze patterns twice, one for condition *a* and one for condition *b*. To reduce the recency effect of participants forgetting the previous conditions counterbalanced pairwise comparisons were performed instead of three-way comparisons. All six pairwise comparisons were combined into a ranked ordered list of three gaze patterns [18]. In each session, they watched two handover videos, consecutively.The different objects and postures used in the experiment are shown in Figs. 4 and 3 respectively.

The instructions at the start of the experiment, as well as the caption for each video, stated that participants should pay close attention to the robot's eyes in the video. After every two videos, the participants were asked to fill out a questionnaire which collected subjective measures as detailed below. The questionnaire was identical to the one used in our previous study [6] and in Zheng et al.'s study [3]. Questions 1 and 2 measure the metric *likability* (Cronbach's $\alpha = 0.83$). Questions 3 and 4 measure the metric *anthropomorphism* (Cronbach's $\alpha = 0.91$). Question 5 measures the metric *timing communication*.

1) Which handover did you like better? (1st or 2nd)
2) Which handover seemed more friendly? (1st or 2nd)
3) Which handover seemed more natural? (1st or 2nd)
4) Which handover seemed more humanlike? (1st or 2nd)
5) Which handover made it easier to tell when, exactly, the robot wanted the giver to give the object? (1st or 2nd)
6) Any other comments (optional)

4.2 Experimental Design

The experiment was designed as a between-within experiment, using likability, anthropomorphism, timing communication as the dependent variables. The participants were divided into three groups of 24 participants. Each group per-

(a)

(b)

Fig. 4 The objects used in the experiments: **a** Object size (small box and large box), **b** Object fragility (plastic bottle and glass bottle)

Table 1 Study Conditions (24 participants per condition)

Condition 1: Object Size	*a*. Small Box
	b. Large Box
Condition 2: Object fragility	*a*. Plastic Bottle
	b. Glass Bottle
Condition 3: User's Posture	*a*. Sitting
	b. Standing

Table 2 Six pairings of the three gaze patterns and their reverse order for each object or posture. Each participant experienced two versions (a/b of a single condition) of these pairings, for a total of 12 pairings

First Handover	Second Handover
Hand-Face	Face-Hand-Face
Hand-Face	Hand
Face-Hand-Face	Hand
Face-Hand-Face	Hand-Face
Hand	Hand-Face
Hand	Face-Hand-Face

[3] To represent objects of different fragility a plastic bottle and a glass bottle were used. In order to examine people's perception about the fragility of these objects, we conducted an online survey. This survey was conducted post experiment based on reviewers' feedback. A total of 24 participants responded to the survey. The participants were undergraduate students from the Department of Industrial Engineering and Management at Ben-Gurion University, similar to the students who participated in our video and in-person experiments. The participants were told that this study deals with object handovers between a human and a robot.

The survey included 10 pictures of objects, made from different materials. The plastic bottle and the glass bottle used in our experiment were among these objects. Each picture was followed by a yes or no question: "Do you perceive this object to be fragile?". Results revealed that all of the 24 participants perceived the plastic bottle to be non-fragile. 23 out of 24 participants perceived the glass bottle to be fragile. Additionally, when asked the same question for three other different plastic and glass bottles, 24 participants denoted the plastic bottles as non-fragile and 23 denoted the glass bottles as fragile. Details about this survey are available in [17]. This supports our decision to choose plastic and glass bottles to represents objects of different fragility.

Table 3 Combined preferences of gaze behaviors in the video study for the small and large object conditions

		Small Object					Large Object				
		Hand-Face	Face-Hand-Face	Hand	a_i	P_i	Hand-Face	Face-Hand-Face	Hand	a_i	P_i
Likability	Hand-Face	0	5.25	21	26.25	0.2	0	5.75	21	26.75	0.25
	Face-Hand-Face	18.75	0	23.5	42.25	0.77	18.25	0	21	39.25	0.70
	Hand	3	0.5	0	3.5	0.03	3	3	0	6	0.05
Anthropomorphism	Hand-Face	0	5	20.5	25.5	0.21	0	7.75	20.5	28.25	0.28
	Face-Hand-Face	19	0	22.5	41.5	0.75	16.25	0	20.75	37	0.65
	Hand	3.5	1.5	0	5	0.04	3.5	3.25	0	6.75	0.07
Timing Communication	Hand-Face	0	7.5	19	26.5	0.29	0	6	20	26	0.31
	Face-Hand-Face	16.5	0	21.5	38	0.63	18	0	17.5	35.5	0.57
	Hand	5	2.5	0	7.5	0.08	4	6.5	0	10.5	0.12

Table 4 Combined preferences of gaze behaviors in the video study for the non-fragile object and fragile object conditions

		Non-Fragile Object					Fragile Object				
		Hand-Face	Face-Hand-Face	Hand	a_i	P_i	Hand-Face	Face-Hand-Face	Hand	a_i	P_i
Likability	Hand-Face	0	6.25	19.5	25.75	0.27	0	7.25	21	28.25	0.32
	Face-Hand-Face	17.75	0	20.5	38.25	0.65	16.75	0	21	37.75	0.62
	Hand	4.5	3.5	0	8	0.08	3	3	0	6	0.06
Anthropomorphism	Hand-Face	0	6	18.25	24.25	0.27	0	7.5	19.5	27	0.32
	Face-Hand-Face	18	0	19	37	0.62	16.5	0	19.75	36.25	0.59
	Hand	5.75	5	0	10.75	0.11	4.5	4.25	0	8.75	0.09
Timing Communication	Hand-Face	0	7	19.5	26.5	0.34	0	7.5	19.5	27	0.35
	Face-Hand-Face	17	0	17	34	0.53	16.5	0	17	33.5	0.52
	Hand	4.5	7	0	11.5	0.13	4.5	7	0	11.5	0.13

Table 5 Combined preferences of gaze behaviors in the video study for the standing and sitting conditions

		Standing					Sitting				
		Hand-Face	Face-Hand-Face	Hand	a_i	P_i	Hand-Face	Face-Hand-Face	Hand	a_i	P_i
Likability	Hand-Face	0	4.5	21	25.5	0.23	0	8.25	19.75	28	0.35
	Face-Hand-Face	19.5	0	21	40.5	0.72	15.75	0	19.5	35.25	0.55
	Hand	3	3	0	6	0.05	4.25	4.5	0	8.75	0.10
Anthropomorphism	Hand-Face	0	7.25	20.5	27.75	0.33	0	9.25	19	28.25	0.37
	Face-Hand-Face	16.75	0	19.75	36.5	0.59	14.75	0	18.25	33	0.50
	Hand	3.5	4.25	0	7.75	0.08	5	5.75	0	10.75	0.13
Timing Communication	Hand-Face	0	6.5	20	26.5	0.31	0	7.5	18	25.5	0.30
	Face-Hand-Face	17.5	0	19	36.5	0.60	16.5	0	19.5	36	0.59
	Hand	4	5	0	9	0.09	5	4.5	0	9	0.11

formed one of the three study conditions listed in Table 1. The order of the 12 sessions were randomized and counterbalanced among the subjects.

4.3 Analysis

The participants' ratings for the likability and anthropomorphism of the gaze behaviors were measured by averaging their responses to Questions 1-2 and 3-4 respectively. The one-sample Wilcoxon signed-rank test was used to check if participants exhibited any bias towards selecting the first or the second handover. Similar to our previous work [6] and Zheng et. al's work [3], the Bradley-Terry model [19] was used to evaluate participants' rankings of the likeability, anthropomorphism and timing communication of gaze behaviors. To evaluate the hypothesis **H1**, i.e. $P_i \neq P_j \forall i \neq j$, where P_i is the probability that one gaze condition is preferred over others, the χ^2 values for each metric were computed, as proposed by Yamaoka et. al [20]:

$$B = n\sum_{i<j} log(P_i + P_j) - \sum_i a_i log P_i, \tag{1}$$

$$\chi^2 = ng(g-1)ln2 - 2Bln10, \tag{2}$$

where, $g = 3$ is the number of gaze behaviors, n is the number of participants, a_i is the sum of ratings in each row of Tables 3-7 (Appendix).

In order to examine **H2-H4**, we conducted two series of tests for each measured metric (likability, anthropomorphism and timing communication), and for each study scenario:

- Binary proportion difference tests for matched pairs [21], in which the difference between the proportion of participants who chose one gaze condition p_b over other p_c was evaluated in each study scenario. The distribution of differences $p_b - p_c$ is:

$$p_b - p_c \sim \mathcal{N}(0, \sqrt{\frac{p_b + p_c - (p_b - p_c)^2}{n}}) \tag{3}$$

  where $n = 24$ is the number of participants in each scenario. The Z-score is calculated according to the following formula:

$$Z = \frac{(p_b - p_c)}{\sqrt{var(p_b - p_c)}} \tag{4}$$

  A low Z-score means that the distribution of differences has zero mean with high probability.
- Equivalence tests based on McNemar's test for matched proportions [22,23], in which the proportion of participants who changed their gaze preferences in each study

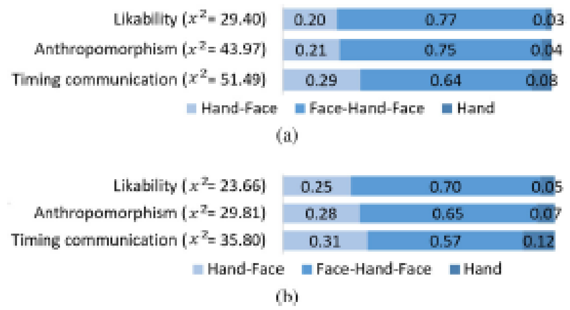


Fig. 5 χ^2 values and win-probabilities of gaze conditions in the video study for the three dependent measures: **a** Small object , **b** Large object

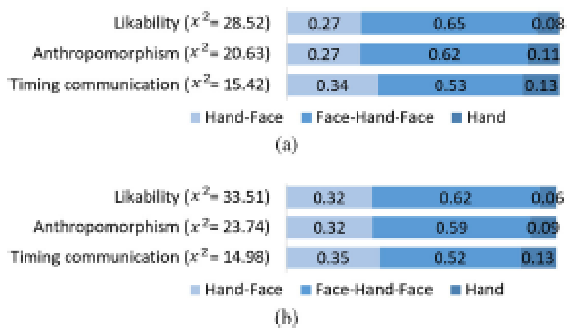


Fig. 6 χ^2 values and win-probabilities of gaze conditions in the video study for the three dependent measures: **a** Non-Fragile object , **b** Fragile object

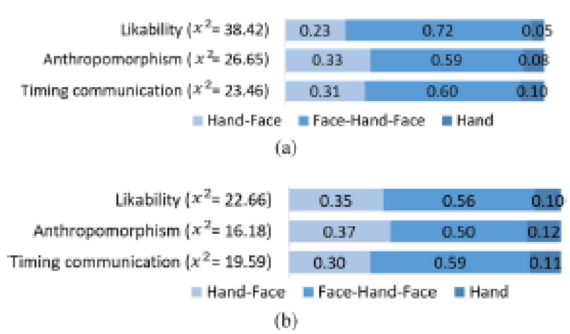


Fig. 7 χ^2 values and win-probabilities of gaze conditions in the video study for the three dependent measures: **a** Standing, **b** Sitting

scenario was compared within equivalence bounds of $\Delta = \pm 0.1$.

Table 6 Combined preferences of gaze behaviors in the in-person study for the small and large object conditions

		Small Object					Large Object				
		Hand-Face	Face-Hand-Face	Hand	q_i	p_i	Hand-Face	Face-Hand-Face	Hand	q_i	p_i
Likability	Hand-Face	0	6.5	20.75	27.25	0.29	0	7.25	20	27.25	0.31
	Face-Hand-Face	17.5	0	20.75	38.25	0.64	16.75	0	20.75	37.5	0.62
	Hand	3.25	3.25	0	6.5	0.07	4	3.25	0	7.25	0.07
Anthropomorphism	Hand-Face	0	6.5	21	27.5	0.3	0	7.5	19	26.5	0.29
	Face-Hand-Face	17.5	0	20.75	38.25	0.64	16.5	0	21.25	37.75	0.63
	Hand	3	3.25	0	6.25	0.06	5	2.75	0	7.75	0.08
Timing Communication	Hand-Face	0	7.5	22	29.5	0.37	0	8.5	18.5	27	0.34
	Face-Hand-Face	16.5	0	19	35.5	0.55	15.5	0	19	34.5	0.54
	Hand	2	5	0	7	0.08	5.5	5	0	10.5	0.12

Table 7 Combined preferences of gaze behaviors in the in-person study for the non-fragile object and fragile object conditions

		Non-Fragile Object					Fragile Object				
		Hand-Face	Face-Hand-Face	Hand	q_i	P_i	Hand-Face	Face-Hand-Face	Hand	q_i	P_i
Likability	Hand-Face	0	5.75	22	27.75	0.24	0	8	22.5	30.5	0.36
	Face-Hand-Face	18.25	0	23	41.25	0.73	16	0	21.25	37.25	0.59
	Hand	2	1	0	3	0.03	1.5	2.75	0	4.25	0.05
Anthropomorphism	Hand-Face	0	6.25	22.25	28.5	0.27	0	8	21.75	29.75	0.34
	Face-Hand-Face	17.75	0	22.75	40.5	0.70	16	0	21.5	37.5	0.61
	Hand	1.75	1.25	0	3	0.03	2.25	2.5	0	4.75	0.05
Timing Communication	Hand-Face	0	8	21.5	29.5	0.36	0	9	21	30	0.40
	Face-Hand-Face	16	0	20	36	0.57	15	0	18.5	33.5	0.50
	Hand	2.5	4	0	6.5	0.07	3	5.5	0	8.5	0.10

4.4 Results

4.4.1 Quantitative Results

To test for order effects, we checked, but did not find any bias towards selecting the first or the second handover [like: $z = -0.68$, $p = 0.50$; friendly: $z = 1.22$, $p = 0.22$; natural: $z = 0.20$, $p = 0.84$; humanlike: $z = 1.36$, $p = 0.17$; timing communication: $z = 1.23$, $p = 0.22$].

Tables 3 - 5 (Appendix) and Fig. 5-7 show the robot gaze preferences of the participants in terms of likability, anthropomorphism and timing communication.

Gaze conditions differ significantly in ratings (all χ^2 values are large ($p < 0.0001$)), supporting **H1**. Participants prefer the *Face-Hand-Face* transition gazes over *Hand-Face* and *Hand* gazes. *Hand* gaze is the least preferred condition.

Based on the binary proportion difference test, we did not find evidence that the proportion of observers of a handover preferring one gaze condition over the other is affected by object size (Table 9, Appendix), object fragility (Table 10, Appendix) and user's posture (Table 11, Appendix). Hypotheses **H2**, **H3** and **H4** are not supported (all p values are over 0.2).

However, based on the equivalence tests, we did not find evidence that the proportion of observers of a handover preferring one gaze condition over the other is equivalent for the two object sizes (Table 9, Appendix), object fragilities (Table 10, Appendix), or user's postures (Table 11, Appendix). Thus, hypotheses **H2**, **H3** and **H4** can also not be rejected (all p values are over 0.15).

4.4.2 Open-ended Responses

All open-ended responses are presented in [17] with major insights detailed below.

10 out of 72 participants gave at least one additional comment. Four out of the eight participants, who made *Hand-Face* gaze vs. *Face-Hand-Face* gaze comparisons, preferred *Face-Hand-Face* gaze over *Hand-Face* gaze due to the extended eye contact by the robot.

P059 - *"As much eye contact as possible."*
P048 - *"I preferred handover 2 (Face-Hand-Face gaze) because the robot looked more at the human"*

Two participants mentioned that they could not distinguish between *Face-Hand-Face* gaze and *Hand-Face* gaze, while two participants commented about the advantages and disadvantages of the two gaze patterns.

P041 - *"In handover 1 (Hand-Face gaze) you could tell that the robot was ready to receive the object. However, handover 2 (Face-Hand-Face gaze) felt more human-*

ized because the robot looked at the giver's eyes right until the transfer was made".

Four out of six participants, who commented on the comparison between *Hand-Face* gaze and *Hand* gaze, preferred *Hand-Face* gaze because of the eye movement.

P008 - *"In my opinion, the change in eye movement creates a better human-robot interaction."*
P009 - *"In the second handover (Hand-Face gaze) the eye movement, gave a good indication for the communication."*

Two participants mentioned that they could not distinguish between *Hand-Face* gaze and *Hand* gaze.

Six participants commented on *Face-Hand-Face* gaze vs. *Hand* gaze comparison. All of them said that they preferred *Face-Hand-Face* gaze over *Hand* gaze.

P009 - *"At handover 2 (Face-Hand-Face gaze), the robot looked at the object precisely when it wanted to take it, so it was perceived more understandable."*
P037 - *"In my opinion video 2 (Face-Hand-Face gaze) best simulated human-like behavior out of all the videos I have seen so far."*

5 In-person Study of Human-to-Robot Handovers

In the in-person study, another set of 72 participants were asked to perform object handovers with the Sawyer robot arm in a similar setup (Fig. 3c). The robot arm and the robot eyes were programmed in the same way as the video study described in Sect. 4.

5.1 Experimental Procedure, Design and Evaluation

The experiment was conducted during the COVID-19 pandemic. Therefore, several precautions were taken. The participants were asked to wash their hands with soap when they entered and exited the lab. The equipment was sterilized before and after each participant, and the experiment room's door remained open at all times. Only one participant was allowed at a time inside the room. Both the participant and conductor of the experiment wore masks and kept at least 2 meters distance between them.

After entering the experiment room, participants signed the electronic consent form, and answered a question on a computer: *How familiar are you with a collaborative robot (such as the one shown)?* Participants ranked this question on a scale from 1 - "Not at all familiar" to 5 - "Extremely familiar". The mean familiarity with this type of robot was found to be low (M=1.49, SD = 0.60, on a scale of 1-5).

Table 8 Combined preferences of gaze behaviors in the in-person study for the standing and sitting conditions

		Standing					Sitting				
		Hand-Face	Face-Hand-Face	Hand	q_j	P_j	Hand-Face	Face-Hand-Face	Hand	q_j	P_j
Likability	Hand-Face	0	7.5	17.5	25	0.31	0	6.75	18.75	25.5	0.28
	Face-Hand-Face	16.5	0	17.75	34.25	0.54	17.25	0	20	37.25	0.62
	Hand	6.5	6.25	0	12.75	0.15	5.25	4	0	9.25	0.10
Anthropomorphism	Hand-Face	0	8.25	16.5	24.75	0.31	0	7.5	17.75	25.25	0.28
	Face-Hand-Face	15.75	0	18	33.75	0.53	16.5	0	20.5	37	0.62
	Hand	7.5	6	0	13.5	0.16	6.25	3.5	0	9.75	0.10
Timing Communication	Hand-Face	0	9	18	27	0.35	0	6.5	19	25.5	0.31
	Face-Hand-Face	15	0	18.5	33.5	0.52	17.5	0	18	35.5	0.57
	Hand	6	5.5	0	11.5	0.13	5	6	0	11	0.12

Table 9 Results of binary proportion difference test and equivalence test for matched pairs comparing small object and large object user's preferences of robot gaze in handovers. Gaze condition in bold is the preferred choice in each pairwise comparison

	Metrics	Gaze Conditions	Binary Proportion Difference Test		Equivalence Test	
			Z score	P-value	Z score	P-value
Video Study	Likability	**Hand-Face** vs. Face-Hand-Face	−0.15	0.44	0.39	0.35
		Hand-Face vs. Hand	0.00	0.50	−0.92	0.18
		Face-Hand-Face vs. Hand	0.38	0.36	0.19	0.58
	Anthropomorphism	**Hand-Face** vs. Face-Hand-Face	−0.78	0.22	0.31	0.38
		Hand-Face vs. Hand	0.00	0.50	0.80	0.21
		Face-Hand-Face vs. Hand	0.27	0.40	−0.32	0.38
	Timing communication	**Hand-Face** vs. Face-Hand-Face	0.41	0.34	−0.75	0.23
		Hand-Face vs. Hand	−0.16	0.44	−0.25	0.60
		Face-Hand-Face vs. Hand	0.65	0.26	0.04	0.52
In-Person Study	Likability	**Hand-Face** vs. Face-Hand-Face	−0.20	0.42	0.51	0.30
		Hand-Face vs. Hand	0.12	0.45	0.45	0.33
		Face-Hand-Face vs. Hand	0.00	0.50	0.96	0.17
	Anthropomorphism	**Hand-Face** vs. Face-Hand-Face	−0.27	0.40	0.59	0.28
		Hand-Face vs. Hand	0.32	0.38	0.06	0.47
		Face-Hand-Face vs. Hand	−0.08	0.47	0.27	0.39
	Timing communication	**Hand-Face** vs. Face-Hand-Face	−0.25	0.40	0.27	0.39
		Hand-Face vs. Hand	0.55	0.29	0.36	0.36
		Face-Hand-Face vs. Hand	0.00	0.50	0.30	0.38

Table 10 Results of binary proportion difference test and equivalence test for matched pairs comparing fragile object and non-fragile object user's preferences of robot gaze in handovers. Gaze condition in bold is the preferred choice in each pairwise comparison

	Metrics	Gaze Conditions	Binary Proportion Difference Test		Equivalence Test	
			Z score	P-value	Z score	P-value
Video Study	Likability	**Hand-Face** vs. Face-Hand-Face	−0.27	0.39	0.29	0.38
		Hand-Face vs. Hand	−0.24	0.41	−0.57	0.28
		Face-Hand-Face vs. Hand	−0.08	0.47	−0.92	0.18
	Anthropomorphism	**Hand-Face** vs. Face-Hand-Face	−0.41	0.34	−0.67	0.25
		Hand-Face vs. Hand	−0.20	0.42	−0.34	0.37
		Face-Hand-Face vs. Hand	−0.12	0.45	0.83	0.20
	Timing communication	**Hand-Face** vs. Face-Hand-Face	−0.13	0.45	−0.03	0.51
		Hand-Face vs. Hand	0.00	0.50	−0.33	0.37
		Face-Hand-Face vs. Hand	0.00	0.50	0.31	0.38
In-Person Study	Likability	**Hand-Face** vs. Face-Hand-Face	−0.61	0.27	−0.20	0.58
		Hand-Face vs. Hand	−0.07	0.47	0.41	0.66
		Face-Hand-Face vs. Hand	0.26	0.40	0.32	0.62
	Anthropomorphism	**Hand-Face** vs. Face-Hand-Face	−0.47	0.32	0.43	0.33
		Hand-Face vs. Hand	0.08	0.47	0.56	0.71
		Face-Hand-Face vs. Hand	0.19	0.43	−0.23	0.41
	Timing communication	**Hand-Face** vs. Face-Hand-Face	−0.24	0.41	−0.03	0.51
		Hand-Face vs. Hand	0.08	0.47	−0.14	0.44
		Face-Hand-Face vs. Hand	0.24	0.41	−0.66	0.25

Table 11 Results of binary proportion difference test and equivalence test for matched pairs comparing sitting and standing user's preferences of robot gaze in handovers. Gaze condition in bold is the preferred choice in each pairwise comparison

	Metrics	Gaze Conditions	Binary Proportion Difference Test		Equivalence Test	
			Z score	P-value	Z score	P-value
Video Study	Likability	**Hand-Face** vs. Face-Hand-Face	−1.08	0.15	−0.63	0.26
		Hand-Face vs. Hand	0.20	0.42	−0.04	0.52
		Face-Hand-Face vs. Hand	0.24	0.41	−0.05	0.52
	Anthropomorphism	**Hand-Face** vs. Face-Hand-Face	−0.49	0.31	0.71	0.24
		Hand-Face vs. Hand	0.24	0.41	−0.65	0.74
		Face-Hand-Face vs. Hand	0.24	0.41	−0.60	0.72
	Timing communication	**Hand-Face** vs. Face-Hand-Face	−0.27	0.40	0.44	0.33
		Hand-Face vs. Hand	0.33	0.37	0.16	0.43
		Face-Hand-Face vs. Hand	−0.08	0.47	−0.21	0.58
In-Person Study	Likability	**Hand-Face** vs. Face-Hand-Face	0.20	0.42	−0.28	0.39
		Hand-Face vs. Hand	−0.21	0.42	−0.51	0.31
		Face-Hand-Face vs. Hand	−0.37	0.36	0.72	0.24
	Anthropomorphism	**Hand-Face** vs. Face-Hand-Face	0.19	0.43	−0.20	0.42
		Hand-Face vs. Hand	−0.21	0.42	−0.39	0.35
		Face-Hand-Face vs. Hand	−0.40	0.35	0.06	0.48
	Timing communication	**Hand-Face** vs. Face-Hand-Face	0.64	0.26	−0.45	0.33
		Hand-Face vs. Hand	−0.16	0.44	0.48	0.32
		Face-Hand-Face vs. Hand	0.08	0.47	−0.32	0.38

The study instructions were given orally by the experimenter. Participants then completed a practice session followed by 12 randomly assigned study sessions. In each session, the participants performed two sequential handovers with the robot. The 12 sessions consisted of the same pairings of gaze behaviors as in the video experiment, followed by the same questionnaire questions. The only difference was in Question 5, which was "Which handover made it easier to tell when, exactly, the robot wanted you to give the object? (1st or 2nd)". The experimental design was also same as the video study.

5.2 Analysis

The hypotheses **H1-H4** were evaluated using the same procedure as described in Sect. 4.3.

To evaluate hypothesis **H5**, we conducted two series of tests for each measured metric (likability, anthropomorphism and timing communication), and for each study scenario. These tests are different from the tests for "matched pairs" which we performed for testing **H2-H4**, since for testing **H5** we need to compare two different participants' groups:

- Binary proportion difference tests for unmatched pairs [24], in which the difference between the proportion of participants who chose one gaze condition over other in each study scenario for the video p_b and in-person p_c studies was evaluated. The distribution for the differences $p_b - p_c$ is:

$$p_b - p_c \sim \mathcal{N}(0, \sqrt{p_d(1-p_d)(\frac{1}{n_b} - \frac{1}{n_c}}) \quad (5)$$

where $n_b = 24$ and $n_c = 24$ are the number of participants in each scenario of the video study and in-person study respectively, and p_d is the pooled proportion calculated as follows:

$$p_d = \frac{X_b + X_c}{n_b + n_c} \quad (6)$$

where X_b and X_c are the number of participants who preferred one gaze condition over the other (shown in Tables 3 - 8, Appendix) in the video and in-person study respectively. Then, the Z-score is calculated same as equation (4).

- Equivalence tests for unmatched proportions [25], in which the proportion of participants who chose one gaze condition over other in each study scenario for the video p_b and in-person p_c studies was tested for equivalence within the bounds of $\Delta = \pm 0.1$.

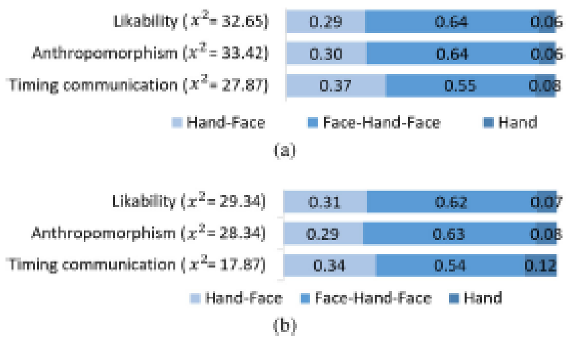


Fig. 8 χ^2 values and win-probabilities of gaze conditions in the in-person study for the three dependent measures: **a** Small object, **b** Large object

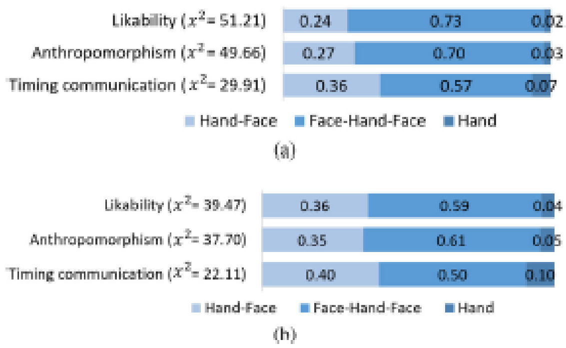


Fig. 9 χ^2 values and win-probabilities of gaze conditions in the in-person study for the three dependent measures: **a** Non-Fragile object, **b** Fragile object

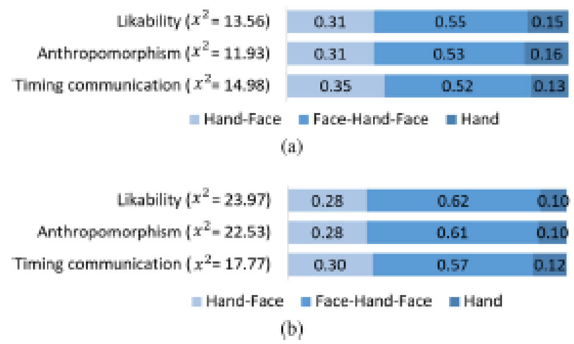


Fig. 10 χ^2 values and win-probabilities of gaze conditions in the in-person study for the three dependent measures: **a** Standing, **b** Sitting

5.3 Results

5.3.1 Quantitative Results

There was no bias towards selecting the first or the second handover [like: z =-0.88, p = 0.38; friendly: z = -0.27, p = 0.79; natural: z =-0.48, p = 0.63; humanlike: z = -1.16, p = 0.25; timing communication: z =0.34, p = 0.73]. Tables 6-8 (Appendix) and Fig. 8- 10 show the robot gaze preferences of the participants in terms of likability, anthropomorphism and timing communication. In all six experimental conditions, the gaze conditions differ significantly in ratings ($p < 0.0001$), supporting **H1**. As in the video study, participants preferred the *Face-Hand-Face transition* gazes over *Hand-Face* and *Hand* gazes. *Hand* gaze was the least preferred ($p < 0.0001$).

Based on the binary proportion difference test, the proportion of participants in a handover preferring one gaze condition over other can not be claimed to be affected by object size (Table 9, Appendix), object fragility (Table 10, Appendix) and user's posture (Table 11, Appendix), contradicting hypotheses **H2**, **H3** and **H4**. The proportion of participants in a handover preferring one gaze condition over other (Table 12, Appendix) also cannot be claimed to be affected by the interaction modality (video or in-person), contradicting **H5**.

However, based on the equivalence tests, we did not find evidence that the proportion of participants in a handover preferring one gaze condition over the other is equivalent for the two object sizes (Table 9, Appendix), object fragilities (Table 10, Appendix), or user's postures (Table 11, Appendix). Thus, hypotheses **H2**, **H3** and **H4** can also not be rejected (all p values are over 0.15). We also did not find evidence that the proportion of participants in a handover preferring one gaze condition over other (Table 12, Appendix) is equivalent for the two interaction modalities (video or in-person). Thus hypothesis **H5** can also not be rejected.

5.3.2 Open-Ended Responses

14 out of 72 participants gave additional comments.

Seven participants made *Hand-Face* gaze vs. *Face-Hand-Face* gaze comparisons. Two of these participants stated that they preferred *Face-Hand-Face* over *Hand-Face* gaze because they preferred longer eye contact by the robot.

> P020 - *"I preferred handover 1 (Face-Hand-Face gaze) because the robot stared at me before and after the handover, and I felt accompanied by it during the entire handover."*

Four participants mentioned that they could not distinguish between the two conditions, while one participant mentioned that *Face-Hand-Face* gaze pattern didn't feel natural.

Four out of the seven participants who commented on the comparison between *Hand-Face* gaze and *Hand* gaze, said that they preferred *Hand-Face* gaze.

P014 - *"In the first handover (Hand-Face gaze) the robot looked straight at me after the handover and seemed to be more friendly."*
P050 - *"In the first handover (Hand-Face gaze), the robot's eye movement was fully accompanied by the handover movement, and therefore it seemed more natural."*

Three participants mentioned that they could not distinguish between *Hand-Face* gaze and *Hand* gaze.

Seven out of eight participants, who commented on the comparison between *Face-Hand-Face* gaze and *Hand* gaze gazes, said that they preferred *Face-Hand-Face* gaze over *Hand* gaze because of a longer eye contact by the robot.

P014 - *"In the first handover (Hand gaze), the robot focused only on the object, and in the second handover (Face-Hand-Face gaze) it focused on me too, so it felt more natural."*
P016 - *"I preferred the second handover (Face-Hand-Face gaze) mainly because the robot looked me in the eyes at the beginning and the end."*

6 Discussion

Prior works studying robot gaze in handovers did so either for a robot as giver, or—in our own prior work on robot receiver gaze—for a small and non-fragile object, and one specific posture of the human. However, for a robot receiver, the object type or giver posture might influence preferences of robot gaze behavior. This raises the question whether the findings in the prior work generalize over variations in the handover task. In this work we investigated the effect of different object types and giver postures on preferred robot gaze behavior in a human-to-robot handover. We did not find evidence that the participants' gaze preference for a robot receiver in a handover is affected by small, large, fragile and non-fragile objects, standing or sitting postures, and the interaction modality i.e. video or in-person. However, in our study, the proportion of participants preferring one gaze condition over other is not statistically equivalent. Thus we cannot completely reject the effect of these scenarios over gaze preferences. In addition, the above-mentioned prior work [6] studied the robot receiver's gaze behaviors only in the reach phase of human-to-robot handovers. The work presented in this paper **extends the empirical evidence by studying the gaze patterns for all three phases of the handover: reach, transfer and retreat**.

As in the previous study [6], results revealed that **the most preferred gaze behavior for a robot receiver was different from the observed most frequent behavior of a human receiver**. When a person receives an object from another person, the most frequent gaze behavior is a *Hand-Face* gaze, in which the receiver looks at the giver's hand throughout the reach and transfer phases, and then at the giver's face in the retreat phase. This indicates that receivers must keep their gaze focused on the task and thus sacrifice the social benefits of the face gaze. The previous findings [6] had revealed that a robot receiver can utilize the flexibility of its perception system to incorporate a face-oriented gaze for social engagement. This finding is reinforced by our current study as the participants preferred a *Face-Hand-Face* transition gaze behavior, in which, the robot initially looked at their face, then transitioned its gaze to their hand during the reach phase, continued to look at their hand during the transfer phase, and finally transitioned its gaze back to again look at their face during the retreat phase. Open-ended responses suggested that **people preferred the robot looking at their face at the beginning and the end of the handover, and the robot's eyes following the object during the transfer phase**. This gaze behavior complemented the robot's handover motion, and thus portrayed the robot as more human-like, natural, and friendly. Another possible explanation is that the social aspects of a human receiver are implicit, whereas a robot has to establish its social agency for a better handover experience. Based on these findings, **we recommend to HRI designers to implement a *Face-Hand-Face* transition gaze when the robot receives an object from a human, regardless of human posture and characteristics of the object being handed over.**

There are several limitations of this study which could motivate future work. The results are limited by the sample size and the specific cultural and demographic makeup of its participants. Larger population samples of different age groups, backgrounds, and cultures should be investigated to help generalize the findings of our experiments. Moreover, as with any experimental study, there is a question of external validity. A handover that is part of a more complex collaborative or assistive task might elicit different expectations of the robot's gaze, a fact that should be considered by designers of HRI systems. To better understand these contextual requirements, additional realistic scenarios of assistive and collaborative tasks should be considered.

7 Conclusion

Video watching studies and in-person studies of robot gaze behaviours in human to robot handovers, revealed that:

- The participants preferred a gaze pattern in which the robot initially looks at their face and then transitions its gaze to their hand and then transitions its gaze back to look at their face again.

Table 12 Results of binary proportion difference test and equivalence test for unmatched pairs comparing video and in-person user's preferences of robot gaze in handovers. Gaze condition in bold is the preferred choice in each pairwise comparison. L: Likability, A: Anthropomorphism, T: Timing communication

	Study Scenario	Metrics	Gaze Conditions	Binary		Equivalence Test	
				Test Statistic	Pvalue	Test Statistic	Pvalue
Object Size	Small Object	L	**Hand-Face** vs. Face-Hand-Face	−0.42	0.67	0.39	0.35
			Hand-Face vs. Hand	0.11	0.91	−0.92	0.18
			Face-Hand-Face vs. Hand	1.48	0.14	0.19	0.58
		A	**Hand-Face** vs. Face-Hand-Face	−0.51	0.61	0.31	0.38
			Hand-Face vs. Hand	−0.21	0.83	0.80	0.21
			Face-Hand-Face vs. Hand	0.85	0.40	−0.32	0.38
		T	**Hand-Face** vs. Face-Hand-Face	0.00	1.00	−0.75	0.23
			Hand-Face vs. Hand	−1.23	0.22	−0.25	0.60
			Face-Hand-Face vs. Hand	0.99	0.32	0.04	0.52
	Large Object	L	**Hand-Face** vs. Face-Hand-Face	−0.49	0.62	0.29	0.38
			Hand-Face vs. Hand	0.41	0.68	−0.57	0.28
			Face-Hand-Face vs. Hand	0.11	0.91	−0.92	0.18
		A	**Hand-Face** vs. Face-Hand-Face	0.08	0.94	−0.67	0.25
			Hand-Face vs. Hand	0.57	0.57	−0.34	0.37
			Face-Hand-Face vs. Hand	−0.22	0.83	0.83	0.20
		T	**Hand-Face** vs. Face-Hand-Face	−0.79	0.43	−0.03	0.51
			Hand-Face vs. Hand	0.54	0.59	−0.33	0.37
			Face-Hand-Face vs. Hand	−0.51	0.61	0.31	0.38
Object Stiffness	Non-Fragile Object	L	**Hand-Face** vs. Face-Hand-Face	0.17	0.87	−0.63	0.26
			Hand-Face vs. Hand	−1.05	0.29	−0.04	0.52
			Face-Hand-Face vs. Hand	−1.24	0.21	−0.05	0.52
		A	**Hand-Face** vs. Face-Hand-Face	−0.08	0.94	0.71	0.24
			Hand-Face vs. Hand	−1.59	0.11	−0.65	0.74
			Face-Hand-Face vs. Hand	−1.61	0.11	−0.60	0.72
		T	**Hand-Face** vs. Face-Hand-Face	−0.31	0.76	0.44	0.33
			Hand-Face vs. Hand	−0.82	0.41	0.16	0.43
			Face-Hand-Face vs. Hand	−1.03	0.30	−0.21	0.58
	Fragile Object	L	**Hand-Face** vs. Face-Hand-Face	−0.23	0.82	0.51	0.30
			Hand-Face vs. Hand	−0.74	0.46	0.45	0.33
			Face-Hand-Face vs. Hand	−0.11	0.91	0.96	0.17
		A	**Hand-Face** vs. Face-Hand-Face	−0.15	0.88	0.59	0.28
			Hand-Face vs. Hand	−0.93	0.35	0.06	0.47
			Face-Hand-Face vs. Hand	−0.73	0.47	0.27	0.39
		T	**Hand-Face** vs. Face-Hand-Face	−0.46	0.65	0.27	0.39
			Hand-Face vs. Hand	−0.60	0.55	0.36	0.36
			Face-Hand-Face vs. Hand	−0.49	0.62	0.30	0.38
User's Posture	Standing	L	**Hand-Face** vs. Face-Hand-Face	−1.00	0.32	−0.20	0.58
			Hand-Face vs. Hand	1.27	0.20	0.41	0.66
			Face-Hand-Face vs. Hand	1.19	0.23	0.32	0.62
		A	**Hand-Face** vs. Face-Hand-Face	−0.31	0.76	0.43	0.33
			Hand-Face vs. Hand	1.37	0.17	0.56	0.71
			Face-Hand-Face vs. Hand	0.62	0.54	−0.23	0.41
		T	**Hand-Face** vs. Face-Hand-Face	−0.77	0.44	−0.03	0.51
			Hand-Face vs. Hand	0.71	0.48	−0.14	0.44

Table 12 continued

Study Scenario	Metrics	Gaze Conditions	Binary		Equivalence Test	
			Test Statistic	Pvalue	Test Statistic	Pvalue
		Face-Hand-Face vs. Hand	0.17	0.87	−0.66	0.25
Sitting	L	**Hand-Face** vs. Face-Hand-Face	0.47	0.64	−0.28	0.39
		Hand-Face vs. Hand	0.36	0.72	−0.51	0.31
		Face-Hand-Face vs. Hand	−0.19	0.85	0.72	0.24
	A	**Hand-Face** vs. Face-Hand-Face	0.53	0.60	−0.20	0.42
		Hand-Face vs. Hand	0.43	0.67	−0.39	0.35
		Face-Hand-Face vs. Hand	−0.82	0.41	0.06	0.48
	T	**Hand-Face** vs. Face-Hand-Face	0.32	0.75	−0.45	0.33
		Hand-Face vs. Hand	−0.34	0.73	0.48	0.32
		Face-Hand-Face vs. Hand	0.52	0.60	−0.32	0.38

- The participants' gaze preference did not change for changes in the object size, object fragility, or the user's posture. However, the gaze preferences were also not statistically equivalent for different object size, object fragility, or the user's posture.

These results could help the design of non-verbal cues in human-to-robot object handovers, which are integral to collaborative and assistive tasks in the workplace and at home.

Acknowledgements This research was supported by Ben-Gurion University of the Negev through the Agricultural, Biological and Cognitive Robotics Initiative, the Helmsley Charitable trust, the Marcus Endowment Fund, and the Rabbi W. Gunther Plaut Chair in Manufacturing Engineering. We would like to acknowledge Hahn-Robotics for their donation of the Sawyer robot. We also thank Prof. Israel Parmet, Ben-Gurion University of the Negev, for his guidance in statistical analysis.

Funding This research was supported by Ben-Gurion University of the Negev through the Helmsley Charitable Trust, the Agricultural, Biological and Cognitive Robotics Initiative, the Marcus Endowment Fund, and the Rabbi W. Gunther Plaut Chair in Manufacturing Engineering. We would like to acknowledge Hahn-Robotics for their donation of the Sawyer robot.

Availability of data and material: Our annotations of gazes in human-human handovers are available at: https://github.com/alapkshirsagar/handover-gaze-annotations/ The videos of robot gaze conditions used in our studies are available at: https://youtu.be/9dD1YHG2Nco

Declarations

Conflicts of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Human and animal rights The study design was approved by the Human Subjects Research Committee at the Department of Industrial Engineering and Management, Ben-Gurion University of the Negev. Informed consent was obtained from the participants.

Appendix

Tables 3 - 8 show the robot gaze preferences of the participants in terms of Likability, Anthropomorphism and Timing Communication. The values in the first three columns indicate the number of "wins" of a row condition over a column condition i.e. the number of participants who preferred a row condition over a column condition. For example, in Table 3 a Likability rating of 21 in the small object, "Hand-Face" row and "Hand" column shows that 21 participants liked the *Hand-Face* gaze over the *Hand* gaze. We obtained these ratings by averaging the participants' responses for both ordered pairwise comparisons, and thus some of these values are fractions. The values in a_i column show the sum of the ratings for each row. The probability that a row condition is preferred over other conditions was calculated using an iterative estimation algorithm [18] and the probability values are shown in P_i column.

Tables 9-11 show the results of binary proportion difference tests and equivalence tests for matched pairs which we used to evaluate **H2-H4**. We evaluated the user's preferred gaze behavior in terms of Likability, Anthropomorphism, and Timing Communication for different study conditions. The values in "Z-score" column represent the test statistic. For example, in Table 9, a Z-score of 0.00 and a P-value of 0.5 for Likability in *Hand-Face* vs. *Hand* gaze conditions means that the proportion of participants in the video study who liked *Hand-Face* over *Hand* condition for both small and large object is not statistically different. However, for the same scenario, a Z-score of -0.92 and a P-value of 0.18 for the Equivalence Test indicates that the proportions are not statistically equivalent as well.

Table 12 show the results of binary proportion difference tests for unmatched pairs which we used to evaluate **H5**.

References

1. Kshirsagar A, Kress-Gazit H, Hoffman G (2019) Specifying and synthesizing human-robot handovers, In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau
2. Moon A, Troniak D, Gleeson B, Pan M, Zheng M, Blumer B, MacLean K, Croft E (2014) Meet me where I'm gazing: How shared attention gaze affects human-robot handover timing, In ACM/IEEE International Conference on Human-robot Interaction (HRI). Bielefeld, Germany
3. Zheng M, Moon A, Croft E, Meng M (2015) Impacts of robot head gaze on robot-to-human handovers. Int J Soc Robot 7(5):783–798
4. Fischer K, Jensen L, Kirstein F, Stabinger S, Erkent Ö, Shukla D, Piater J (2015) The effects of social gaze in human-robot collaborative assembly, In International Conference on Social Robotics (ICSR). France, Paris
5. Kühnlenz B, Wang Z.-Q, Kühnlenz K (2017) Impact of continuous eye contact of a humanoid robot on user experience and interactions with professional user background, In IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man), Lisbon, Portugal
6. Kshirsagar A, Lim M, Christian S, Hoffman G (2020) Robot gaze behaviors in human-to-robot handovers. IEEE Robot Autom Lett 5(4):6552–6558
7. Koay KL, Sisbot EA, Syrdal DS, Walters ML, Dautenhahn K, Alami R (2007) Exploratory study of a robot approaching a person in the context of handing over an object, In AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics. Stanford CA
8. Edsinger A, Kemp CC (2007) Human-robot interaction for cooperative manipulation: Handing objects to one another, In IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man). Jeju, South Korea
9. Cakmak M, Srinivasa SS, Lee MK, Forlizzi J, Kiesler S (2011) Human preferences for robot-human hand-over configurations, In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). San Francisco
10. Cakmak M, Srinivasa S. S, Lee M. K, Kiesler S, Forlizzi J (2011)Using spatial and temporal contrast for fluent robot-human hand-overs, In 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)
11. zu Borgsen S. M, Bernotat J, Wachsmuth S, (2017) Hand in hand with robots: differences between experienced and naive users in human-robot handover scenarios, In International Conference on Social Robotics (ICSR). Tsukuba, Japan
12. Pan M. K, Croft E. A, Niemeyer G (2018) Evaluating social perception of human-to-robot handovers using the robot social attributes scale (rosas), In ACM/IEEE International Conference on Human-Robot Interaction (HRI), Chicago, USA
13. Flanagan J, Johansson R (2003) Action plans used in action observation. Nat 424(6950):769
14. Strabala K, Lee M, Dragan A, Forlizzi J, Srinivasa S (2012) Learning the communication of intent prior to physical collaboration, In IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man). France, Paris
15. Admoni H, Dragan A, Srinivasa S, Scassellati B (2014) Deliberate delays during robot-to-human handovers improve compliance with gaze communication, In ACM/IEEE International Conference on Human-Robot Interaction (HRI). Bielefeld, Germany
16. Carff A, Foglino F, Bruno B, Mastrogiovanni F (2019) A multi-sensor dataset of human-human handover. Data Brief 22:109–117
17. Faibish T (2021) Human-robot handovers: Human preferences and robot learning, Master's thesis, Department of Industrial Engineering and Management, Ben-Gurion university of the Negev, Beer Sheva, Israel
18. Hunter D (2003) MM algorithms for generalized bradley-terry models. Annals Statist 32(1):384–406
19. Bradley R, Terry M (1952) Rank analysis of incomplete block designs: I. the method of paired comparisons. Biometrika 39(3/4):324–345
20. Yamaoka F, Kanda T, Ishiguro H, Hagita N (2006) How contingent should a communication robot be? In ACM SIGCHI/SIGART Conference on Human-robot Interaction (HRI). Salt Lake City, USA
21. May WL, Johnson WD (1997) The validity and power of tests for equality of two correlated proportions. Statist Med 16(10):1081–1096
22. McNemar Q (1947) Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12(2):153–157
23. Morikawa T, Yanagawa T, Endou A, Yoshimura I (1996) Equivalence tests for pair-matched binary data. Bull Inf Cybern 28(1):31–46
24. Illowsky B, Dean S (2018) Introductory statistics, pp 579–584
25. Lakens D, Scheel AM, Isager PM (2018) Equivalence testing for psychological research: a tutorial. Adv Methods Pract Psychol Sci 1(2):259–269

Chapter 4. Implementation and Evaluation of Guided Policy Search for Robot Reaching Towards Moving Targets

4.1 Introduction

This study evaluates and implements Guided Policy Search (GPS) on a robot controller to perform human-robot object handovers. Handovers play an essential role in the current use of collaborative and assistive robots alongside humans, for instance, in household chores, elderly care, collaborative assembly, and surgical assistance. Out of the three phases comprising a handover: reach, transfer and retreat (Kshirsagar et al., 2019), we focus on the first phase - the reach phase. In this phase, both participants spread their arms towards the handover location. Previous HRI studies have suggested several online (Kshirsagar et al., 2021; Yang et al., 2020; Kshirsagar et al., 2019; Pan et al., 2019; Scimmi et al., 2019; Pan et al., 2018; Vogt et al., 2018; Zhao et al., 2018; Medina et al., 2016; Maeda et al., 2014; Bdiwi et al., 2013; Yamane et al., 2013; Micelli et al., 2011) and offline (Rasch et al., 2019; Peternel et al., 2017; Moon et al., 2014; Sisbot & Alami, 2012; Cakmak et al., 2011;Cakmak et al., 2011b) controllers for the reach phase of handovers. Yet, these methods require precise robot's dynamics models and\or human kinematics models. Lately, Guided Policy Search (GPS) (Levine et al., 2014; Levine et al., 2015; Levine et al., 2016), a model-based reinforcement learning algorithm, has become the focus of interest for many researchers. The GPS algorithm has been used to learn controllers without known robot dynamics and showed encouraging success in several autonomous tasks (Levine et al., 2016; S Levine et al., 2015; Levine & Abbeel, 2014), but with no human interaction. This method uses an iterative adaptation of local controllers, a dynamic model, and a global policy to optimize a policy over repeated trials, without prior knowledge of the robot dynamics.

To the best of our knowledge, most GPS algorithms have been tested on autonomous manipulation (Chebotar et al., 2017; Levine et al., 2016; Levine et al., 2015; Levine & Abbeel, 2014), and locomotion tasks (Zhang et al., 2016; Levine & Abbeel, 2014;Levine & Koltun, 2013, Levine & Koltun, 2013b). No work has used GPS for HRI tasks, like object handovers on a real robot (Kshirsagar et al., 2021). GPS has also been used for learning manipulation tasks, i.e. placing a hanger on a bar, inserting shapes into a sorting cube, inserting a hammer underneath a nail, screwing a bottle cap stacking small blocks, assembling toys, inserting rings on wooden pegs (Levine & Abbeel, 2014). The use of fixed targets, fixed robot dynamics, and small variations in the test locations is common to all GPS applications discussed above.

An object handover task is different, requiring novel tools to handle it. First, it requires motion planning for a moving (non-fixed) target, i.e., the human's hand. Second, the robot dynamics is not fixed due to the diverse objects being handed over. Finally, the training target trajectories and the testing target trajectories could differ substantially when human's unpredictable behavior is taken into consideration.

Prior research of this field included simulation testing with a robot arm substituting for the human, generating the variability and movement of the handover target location (Kshirsagar et al., 2021). Despite providing important insights, their application to a real-world environment is limited. In this work, we evaluated a robot controller that uses Guided Policy Search with a physical robot, with training conducted both in a simulation environment and directly on the physical robot.

4.2 Related work

In this section, we provide a brief summary of existing controllers for the reach phase of human-robot handovers, and previous works using GPS.

4.2.1 Human-Robot handover reach phase controllers

Numerous offline or online controllers have been suggested for refining human-robot interaction during the reach phase of human-robot handovers. Offline controllers (Rasch et al., 2019; Peternel et al., 2017; Moon et al., 2014; Sisbot & Alami, 2012; Cakmak et al., 2011;Cakmak et al., 2011b) encompass a few disadvantages. Lack of adaptability is one distinct disadvantage. The robot's motion plan is computed before the reach phase initiation and does not update simultaneously to the changing human's actions during this phase. Therefore, offline controllers may not be desirable, in particular, if the human operator is preoccupied with other tasks and does not pay his/her undivided attention to the handover, thus, possibly resulting in an unsuccessful handover. We propose here an online controller that continuously updates the robot's motion plan throughout the reach phase while observing the momentary state of the human operator.

The visual servoing approach, i.e., directing a robot towards the human's hand, is the simplest approach used in online controllers for the reach phase of handovers (Pan et al., 2018; Bdiwi et al., 2013; Micelli et al., 2011). This controller generates velocities proportional to the distance between the position of the human's hand and the robot's gripper, allowing it to continuously update the robot's motion plan. Other velocity profiles and motion planners have been used to direct the robot towards the predicted handover location. Pan et al. (Pan et al., 2019) attempted to achieve smooth minimum-jerk trajectories using Bézier curves. Scimmi et al. (Scimmi et al., 2019) applied a predefined smooth velocity profile. Kshirsagar et al. (Kshirsagar et al., 2019) investigated the possibility of specifying the robot's handover behavior by synthesizing handovers automatically. All of these controllers lack human effortlessness and fluency in motion.

With the objective to imitate the reaching phase in human handovers, several online controllers have used various movements primitives, i.e., Dynamic Movement Primitives (DMPs) (Prada et al., 2014), Probabilistic Movement Primitives (ProMPs) (Maeda et al., 2014), and triadic interaction meshes (IMs) (Vogt et al., 2018). Other approaches have used human demonstrations to implement the reaching phase in robots by using dynamical systems (Medina et al., 2016), look-up tables (Yamane et al., 2013), or neural networks (Yang et al., 2020; Zhao et al., 2018). Some researchers have proposed reinforcement learning to learn online controllers for the reach phase from human feedback (Kupcsik et al., 2018; Riccio et al., 2016).

However, all these existent controllers require the robot dynamics. Dynamical parameters may be difficult to obtain for proprietary claimed commercial robots and custom-built robots. One possibility to address this challenge is the use of system identification methods and learning of dynamical models. However, this requires extensive training data as global and complex dynamical models need to be learned. On the contrary, GPS builds local control models, integrates them with a global policy trained by the local controllers through supervised learning, and is thereby data-efficient.

4.2.2 Guided policy search for human-robot handovers

Most of the existing controllers for human-robot handovers require precise robot kinematic/dynamic models. Moreover, many require controller parameters which are non-intuitive and difficult to tune, such as weights of movement primitives or velocity tracking gains. In contrast, "Guided Policy Search (GPS)" (Levine et al., 2014; Levine et al., 2015; Levine et al., 2016) can be used to generate an online handover controller which does not require tuning of control parameters or the robot's dynamic/kinematic models. A few of the algorithm's compelling features include generalizability, sample efficiency, and local minima avoidance (Du et al., 2021; Kshirsagar et al., 2021). It combines learned local dynamic models with a global optimal control policy, and by the use of deep neural networks, it can generalize from local policies.

4.2.2.1 Guided policy search

Of the various Reinforcement Learning methods, policy search methods focus on discovering suitable parameters for a given policy parameterization (Deisenroth et al., 2013). Since policy search methods depend on trial and error to optimize their parameters, they are prone to get stuck in local minima, in particular, for policies with a large number of parameters. To address this issue, prior studies have suggested “Guided Policy Search”, a policy search method that allows the combination of supervised learning of the policy with local trajectory optimization.

The first “Guided Policy Search” algorithm proposed by Levine and Koltun (Levine & Koltun, 2013) was comprised of differential dynamic programming as a means to produce locally optimal controllers which guided, with a large number of parameters, the supervised learning of neural network policies. They used demonstrations to initialize trajectories and applied importance sampling to generate new samples of optimized trajectories to each gradient step. They applied their method by learning locomotion tasks, for instance, walking, running, hooping and planar swimming. In order to combine policy search with trajectory optimization, Levine and Koltun applied in another study (Levine & Koltun, 2013b) variational decomposition of a maximum likelihood objective rather than using their previous importance sampling method. They proved that with respect to the previously considered locomotion tasks (like walking, running, hopping, and planar swimming), this method surpasses the importance sampling GPS method. Levine and Koltun (Levine & Koltun, 2014) utilized a policy agreement constraint for the guidance of policy search with trajectory optimization. To solve the constrained optimization problem, differential dynamic programming (DDP) and dual gradient descent was used. They showed that in comparison to their prior importance sampling and variational GPS algorithms, the constrained GPS algorithm yielded better results on the locomotion tasks. In addition, they were able to learn complicated tasks like walking on uneven terrain and bipedal push recovery. In all these GPS variants, knowledge of the system dynamics was required.

In a subsequent study, Levine and Abbeel (Levine & Abbeel, 2014) suggested a method for trajectories optimization with unknown system dynamics. They used Levine and Koltun's (Levine & Koltun, 2014) constrained GPS algorithm and extended it by refitting locally linear dynamics models iteratively. They showed that their method required fewer samples compared to model-free methods and eliminated the need to learn global models, which is challenging for complex systems. The method was evaluated by simulating robotic locomotion tasks, such as, swimming and walking, and robotic manipulation tasks, such as peg insertion. Levine et al. (Levine et al., 2015) used the constrained GPS algorithm and adjusted it to study manipulation skills on a real robot with unknown dynamics. An adaptive scheme was added for choosing the number of samples and step size, and an augmentation method for policy training with synthetic samples. They performed various experiments with a PR-2 robot (a two-arm robotic system with 7 DOF in each arm), such as assembling toys, stacking Lego blocks, inserting a shoe tree, screwing bottle caps, and inserting rings on wooden pegs to demonstrate their algorithm.

In another study, Levine et al. (Levine et al., 2016) provided an end-to-end algorithm using GPS to transform sensory input (raw images) into motor output (joint torques). They formulated the constrained GPS algorithm as an instance of Bregman-Alternating Direction Method of Multipliers (BADMM). They examined their method on different tasks requiring visual and control close coordination, e.g., placing a hanger on a bar, inserting shapes into a sorting cube, inserting a hammer underneath a nail, and screwing a bottle cap. Zhang et al. (Zhang et al., 2016) enhanced the original GPS algorithm to form training data without disastrous failures by adding a model predictive control (MPC) scheme. During the training phase, they used an instrumented setup to gain full state observations and trained a deep neural network policy with samples produced by MPC. During the testing phase, partial system observations were sufficient for the policy to produce control inputs. They showed that their enhanced GPS algorithm with MPC was comparable to the original GPS algorithm without model errors. They also showed that the enhanced GPS algorithm exceeded the original one with the introduction of model errors. Chebotar et al. (Chebotar et al., 2017) provided another modification to the GPS algorithm. Rather than using the former iterative linear quadratic regulator (iLQR) to generate local controllers, they added a model-free local optimizer based on path integral (PI) stochastic optimal control.

Furthermore, contrary to Levine and Koltun GPS algorithms, which generate training data by local controllers, Chebotar et al., ran global policy on new sets of task cases in each iteration to generate the training samples. They first configured the local policies using kinesthetic teaching and initialized the global policy by performing numerous standard GPS iterations with local policy sampling using PI. The algorithm performed better than iLQR-based GPS, on tasks which included intermittent and variable contacts (contacts at different changing spatial locations) as well as discontinuous cost functions.

As depicted in Fig. 1, the GPS algorithm alternates between generating optimal trajectories for each initial condition (local iLQR controllers) and training a global policy supervised by the local controllers. The global policy's role is to improve the local controllers, retaining them close to the

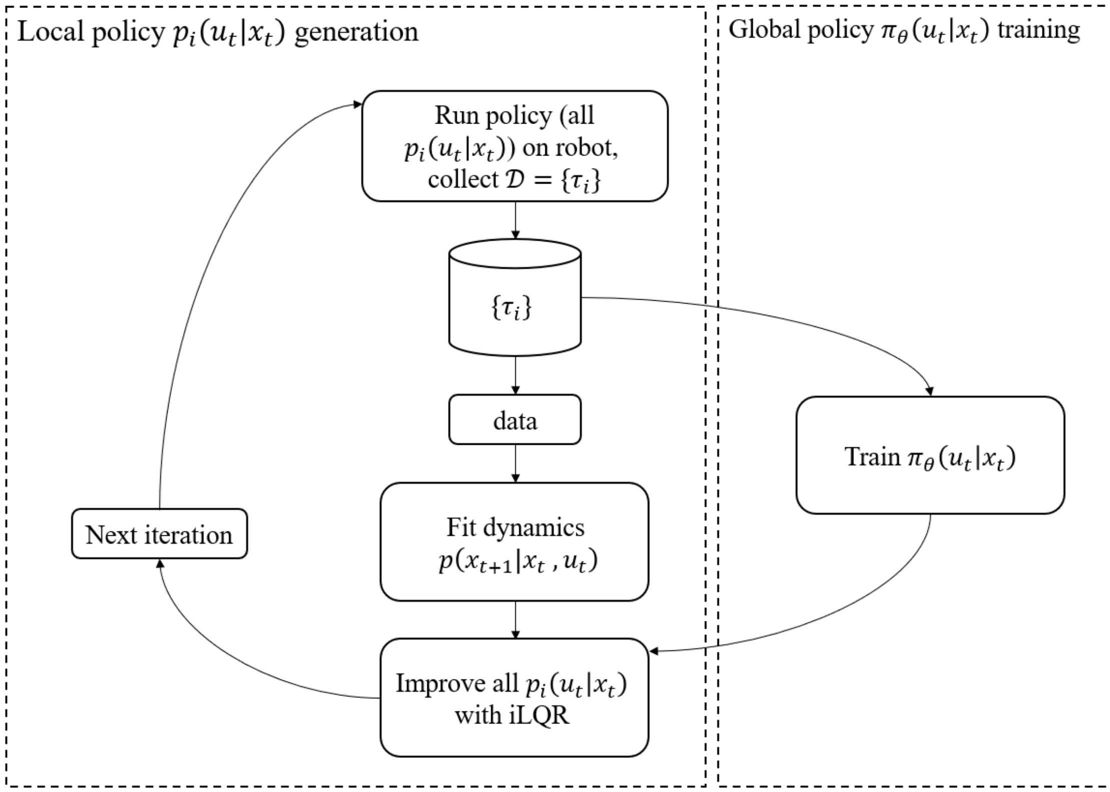


Figure 1: Guided policy search algorithm. Iteratively updates the local controllers (local policies) and the global policy. The local policies serve as the "experts" for supervised learning of the global policy. The local policies are also updated to avoid drifting away from the global policy.

global policy. This algorithm does not require knowledge of the dynamics model as it utilizes the training data with locally linear models to approximates the dynamics.

Most GPS algorithms have been tested on autonomous manipulation (Chebotar et al., 2017; Levine et al., 2016; Levine et al., 2015; Levine & Abbeel, 2014), and locomotion tasks (Zhang et al., 2016; Levine & Abbeel, 2014;Levine & Koltun, 2013, Levine & Koltun, 2013b). A recently published study by Kshirsagar et al, evaluated the potential of GPS to train a robot controller for human-robot object handovers (Kshirsagar et al., 2021) and explored the sensitivity of GPS to different state representations. Three different system state representation were investigated *(FULL, RELATIVE, REDUCED).*

The *full* state representation consisted of the robot joint angles, the robot joint velocities, the human arm joint angles, the human arm joint velocities, the positions and velocities of three points on the object, the human hand, and the robot gripper, and the robot gripper's width. This study has shown that a policy trained with *Relative* state representation (does not include the human's joint angles and velocities from the state representation, and expressing the human hand's position and velocity in a reference frame attached to the robot gripper) has a better overall performance, and therefore we used the *Relative* state representation.

They also showed that the use of GPS creates a global policy that does not perform well for target test locations that are spatially too distant from target training locations. This issue can be mitigated by adding local controllers trained over target locations within the high error regions. More efficient reaching trajectories can be obtained by training on moving targets, although it results in higher worst-case errors. Lastly, they found that changes in the robot's end-effector mass, inducing changes in robot dynamics, are well tolerated and adjusted by the global policy. In this study, training and testing was conducted in a simulated environment. In this thesis, we repeated the study on a physical robot and tested GPS for object handovers in a real-world context. To the best of our knowledge, GPS has not been applied to object handovers on a real robot (Kshirsagar et al., 2021). This gap has been addressed in this study.

4.3 Policy search formulation of handover

We start by briefly describing the GPS algorithm and then we formalize the handover task's reach phase as a reinforcement learning problem.

4.3.1 Guided policy search algorithm

Policy search algorithms aim to discover a policy $\pi_\theta(\boldsymbol{u}_t|x_t)$ that will minimize the execution cost $E_{\pi_\theta}[\sum_{t=1}^{T} l(x_t, \boldsymbol{u}_t)]$ of the desired task. Here, θ indicates the policy parameters, for instance, the weights of a neural network. The system at time *t* is defined by state x_t (for example, the joints angles, joints velocities, end-effector angles, end-effector velocities, and object's positions), control inputs $\boldsymbol{u}_t$ (for example, motor torque commands) and a cost function $(x_t, \boldsymbol{u}_t)$.

When trying to solve this minimization problem using reinforcement learning, large amounts of training data are required, and the algorithm is susceptible to local minima. Guided policy search algorithms surmount these concerns by using "local" controllers $p_i(\boldsymbol{u}_t|x_t)$ (guiding distributions) to train a "global" policy $\pi_\theta(\boldsymbol{u}_t|x_t)$ through supervised learning. The training of local controllers could be via trajectory optimization methods like iLQR. Hence, GPS is formulated in terms of a constrained optimization problem, given by

$$min_{p,\theta}\ E_{\pi_\theta}[\textstyle\sum_{t=1}^{T} l(x_t, \boldsymbol{u}_t)] \quad s.t \quad p(\boldsymbol{u}_t|x_t) = \pi_\theta(\boldsymbol{u}_t|x_t)\ \forall t \quad (1),$$

where $p(\boldsymbol{u}_t|x_t)$ is a guiding distributions mixture $p_i(\boldsymbol{u}_t|x_t)$. The cost is minimized with respect to $p(\tau) = p(x_1)\prod_{t=1}^{T} p(x_{t+1}|x_t, \boldsymbol{u}_t)\, p(\boldsymbol{u}_t|x_t)$ over trajectories $\tau = \{x_1, \boldsymbol{u}_1, \dots, x_T, \boldsymbol{u}_T\}$ with dynamic model of the system given by $p(x_{t+1}|x_t, \boldsymbol{u}_t)$.

As Section 4.2.2 details, few GPS algorithm variants require knowledge of the robot dynamic models, whereas others, iteratively learn locally linear dynamics models using training data. In this study, we use an algorithm introduced by Levine et al. (Levine et al., 2016), which does not require knowledge of the robot dynamics and uses the Bregman-Alternating Direction Method of Multipliers (BADMM). This algorithm represents the local controllers $p_i(\boldsymbol{u}_t|x_t)$ and the dynamics $p_i(x_{t+1}|x_t, \boldsymbol{u}_t)$ as linear, time-varying Gaussians:

$$p_i(\boldsymbol{u}_t|x_t) = \mathcal{N}\big(\boldsymbol{K}_{t,i}x_{t,i} + \boldsymbol{k}_{t,i}, \boldsymbol{C}_{t,i}\big), \qquad (2)$$

$$p_i(x_{t+1}|x_t, \boldsymbol{u}_t) = \mathcal{N}\big(f_{xt,i}x_t + f_{\boldsymbol{u}t,i}\boldsymbol{u}_t + f_{ct,i}\boldsymbol{F}_{t,i}\big). \qquad (3)$$

These type of controllers may serve as an appropriate choice for guiding distribution optimization, as they can be efficiently learned, using a small number of real-world samples. For each training target trajectory (in our case: the reach motion of the human), a different set of controller and dynamics parameters are suited. However, all of the local controllers supervise a single global policy, making it generalizable to various test target trajectories. To make the constraint in (1) tractable, Levine et al. (Levine et al., 2016) proposed modifying the constraint by multiplying with $p(x_t)$ and applying it to expected action:

$$min_{p,\theta}\ E_p[\sum_{t=1}^{T} l(x_t, \boldsymbol{u}_t)] \quad s.t \quad E_{p(\boldsymbol{u}_t|x_t)}[\boldsymbol{u}_t] = E_{p(x_t)\pi_\theta(\boldsymbol{u}_t|x_t)}[\boldsymbol{u}_t] \quad \forall t \ . \quad (4)$$

The GPS algorithm alternates between training a global policy supervised by the local controllers and generating optimal trajectories for each local controller using iLQR. Additional use of the global policy is to improve the local controllers, so that the local controllers stay close to the global policy. Thus, GPS alternates minimization of θ and p as follows:

$$\theta \leftarrow argmin_\theta \sum_{t=1}^{T} E_{p(x_t)\pi_\theta(\boldsymbol{u}_t|x_t)}\left[\boldsymbol{u}_t^T \lambda_{\mu t}\right] + v_t E_{p(x_t)}[D_{KL}\left(p(\boldsymbol{u}_t|x_t)\middle|\middle|\pi_\theta(\boldsymbol{u}_t|x_t)\right)], \quad (5)$$

$$P \leftarrow argmin_P \sum_{t=1}^{T} E_{p(x_t,\boldsymbol{u}_t)}\left[l(x_t, \boldsymbol{u}_t)] - \boldsymbol{u}_t^T \lambda_{\mu t}\right] + v_t E_{p(x_t)}[D_{KL}\left(\pi_\theta(\boldsymbol{u}_t|x_t)\middle|\middle|p(\boldsymbol{u}_t|x_t)\right)], \quad (6)$$

$$\lambda_{\mu t} \leftarrow \lambda_{\mu t} + \alpha v_t (E_{p(x_t)\pi_\theta(\boldsymbol{u}_t|x_t)}[\boldsymbol{u}_t] - E_{p(x_t)p(\boldsymbol{u}_t|x_t)}[\boldsymbol{u}_t]), \quad (7)$$

where $\lambda_{\mu t}$ is the Lagrange multiplier on the expected action at time t, v_t is the weight of the Kullback–Leibler divergence term that serves to keep $p(\boldsymbol{u}_t|x_t)$ close to $\pi_\theta(\boldsymbol{u}_t|x_t)$. For a more comprehensive description of GPS algorithms, see (Levine et al., 2016)

4.3.2 System state representation

As discussed in Sec. 4.2.2.1 we used the *Relative* system state representation (Kshirsagar et al., 2021). The *Relative* state representation consists of the robot joint angles θ_r, the robot joint velocities $\dot{\theta}_r$, the positions and velocities of the object in the robot end-effector frame $(p_o^r, \dot{p}_o^r)$, the positions and velocities of the human hand in the robot end-effector frame $(p_h^r, \dot{p}_h^r)$:

$$x_t = [\theta_r, \dot{\theta}_r, p_o^r, p_h^r, \dot{p}_o^r, \dot{p}_h^r]_t, \quad (8)$$

The robot's control input u_t consists of the robot joint torques τ and the force applied by the gripper's actuator f_g, constrained by $u_{min} \leq u_t \leq u_{max}$:

$$u_t = [\tau, f_g]_t. \quad (9)$$

To consider the robot's dynamics we used torques rather than a kinematic model in terms of velocities or positions as control inputs. By that, the need for tuning low-level position/velocity controllers is eliminated. Moreover, position or velocity controllers might apply considerable impact forces on the human., and thus, endanger human safety.

4.3.3 Cost function

The task of the robot (moving its gripper towards the human's hand in the reach phase of handovers) is described in terms of the following cost function:

$$c_{reach} = [||p_r - p_h||^{2} + \ln(||p_r - p_h||^{2} + \alpha_{reach})], \quad (10)$$

where p_r is the position of the robot and p_h is the position of the human hand. This cost function penalizes and encourages the robot according to the following conditions: This cost function first penalizes the robot for spatial distance from the human's hand. Second, it encourages the robot for accurate placement owing to its concave shape, as described in (Levine et al., 2015). In other words, this cost function encourages the robot to quickly and accurately reach the human's hand. α_{reach} is the parameter that determines the penalty in the target's surroundings. As in Levine et al. (2015), we set $\alpha_{reach} = 1e - 5$ (Sec. 4.5).

4.4 Implementation

To evaluate a robot controller that uses Guided Policy Search with a physical Panda (Franka Emika) robot, we train a collaborative robot to perform handover reaching motions, in both a simulation environment (sim-to-real) and directly on the physical robot (real-to-real) over repeated trials.

4.4.1 MuJoCo simulation environment

We build upon the BADMM-GPS implementation by Finn et al. (Finn et al., 2016) and Kshirsagar et al. (Kshirsagar et al., 2021). The collaborative robot in the handover task is simulated in MuJoCo (Multi-Joint dynamics with Contact) (Todorov et al., 2012). MuJoCo is a physics engine aiming to facilitate research and development where fast and accurate simulation is needed. MuJoCo provides a unique combination of speed, accuracy and modeling power. MuJoCo was used to train the robot in simulation. We imported the Panda URDF file to MuJoCo in order to simulate the Panda robot.

Fig. 2 shows the MuJoCo simulation environment was built for the previous study (Kshirsagar et al., 2021), and used for this study. Fig. 2 shows a Panda robot with 7 degrees-of-freedom (DOF), equipped with a two fingered gripper. The environment also includes a pseudo-robot arm with two DOF and a mass rigidly attached to its end-effector, substituting the human operator.

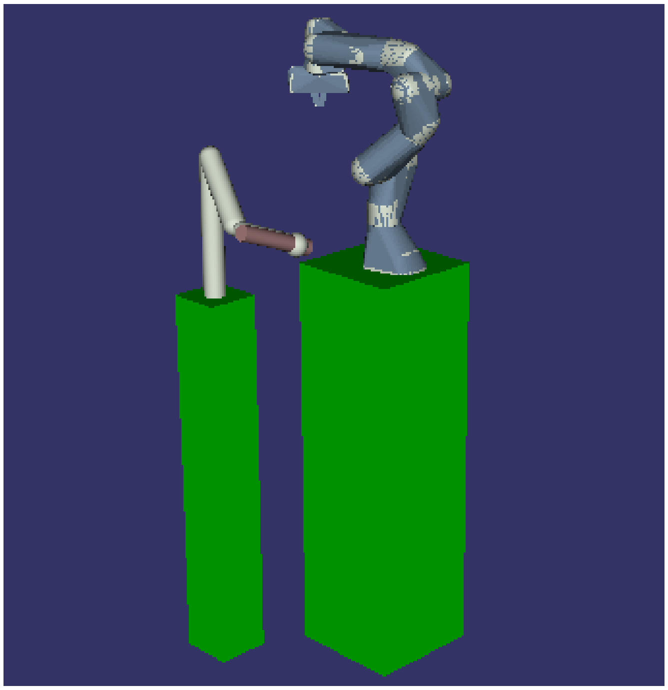


Figure 2: MuJoCo (Multi-Joint dynamics with Contact) simulation environment for human handover tasks. A Panda robot (right) was trained in simulation on reaching movements in a human-to-robot handover task. The human operator is represented by a pseudo-robot (left).

4.4.2 The Panda robot

We trained a Panda robot (Fig. 3) to perform handover reaching motions. The Panda Robot is a 7 DOF anthropomorphic arm with torque sensors at each joint, allowing adjustable stiffness/compliance and advanced torque control. It weighs 17.8 kg, has a payload of 3 kg, a reach of 855 mm and a workspace coverage of 94.5%.
We used a Panda robot as it facilitates to conduct research due to its add-on Franka Control Interface (FCI), allowing to study control and motion algorithms, grasping strategies, interaction scenarios and machine learning, FCI allows a fast low-level bidirectional connection to the robot's arm and hand. The FCI provides the current status of the robot and enables its direct control at a rate of 1kHz (Franka Emika GmbH, 2020).

libfranka is a C++ software library that implements the client-side interface of the FCI, i.e. the drivers implementing the 1 kHz UDP-based communication with the robot. It also gives access to the robot model library, which provides the kinematic and dynamic model of the robot. franka_ros connects Panda with the entire ROS ecosystem. It integrates libfranka into ROS Control, and includes URDF models and detailed 3D meshes of the robot and end-effector for visualization (e.g. RViz) and kinematic simulations.

In the beginning of our training attempts, the robot was fixed to a table. Then, in an attempt to reach the object, during one of the training sessions, the robot bumped into the table and took a hit that caused a permanent offset in the torque sensing module of joint six (which affects the torque sensing capabilities and control of the arm). We had to use another robotic arm, since Franka-Emika did not offer a repair service for the arm. To avoid such situations in the future, we designed a different mount for the robot as shown, in Fig 3. A plate that was attached to the robot's base and mounted on a pillar fixed to the floor. Hence, the robot had no possibility to crash.

Figure 3: Panda robot developed by Franka-Emika connected to a designed floating position.

4.4.3 OptiTrack motion tracking system

The OptiTrack motion tracking system was used to track the positions of the human's hand and the robot end effector. Since it is not practical to have a human trainer/tester perform exactly the same handover motion in all training/testing iterations, we use recorded human hand motions during the training process. The OptiTrack system used 8 multiple synchronized 2D cameras in our setup (Fig. 4), capturing images of reflective markers. To compute the markers' 3D positions these 2D positions are superimposed and triangulation is used. The mocap_optitrack ROS package was used to stream OptiTrack mocap data to tf. This package contains a node that translates motion capture data from an OptiTrack rig to tf transforms, poses and 2D poses. The node receives packets that are streamed by a NatNet compliant source, decodes them and broadcasts the poses of configured rigid bodies as tf transforms, poses, and/or 2D poses.

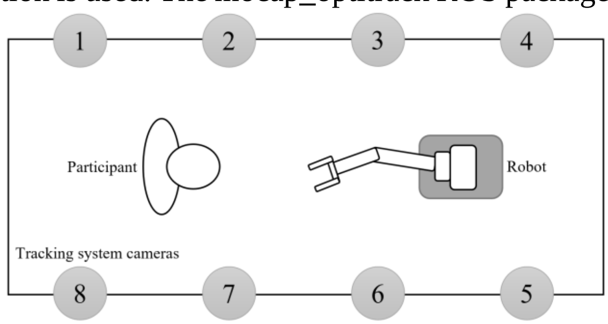


Figure 4: Diagram of the setup for the experiments. Our setup consisted of Panda robot and OptiTrack motion tracking system with 8 cameras.

4.4.4 Robot operating Ssystem

The Robot Operating System (ROS) was used. ROS is a collection of tools, code libraries, and protocols providing a flexible framework for writing robot programs (Casañ et al., 2015). It offers a messaging interface that allows communication between different code elements. The topic interface is anonymous and asynchronous, allowing fast and convenient data transfer and processing.

In this project, ROS was used to operate the Panda robot and to get messages from the OptiTrack motion tracking system. We used the distributed computing capabilities of ROS and ran different ROS packages/nodes on different machines as shown in Fig 5.

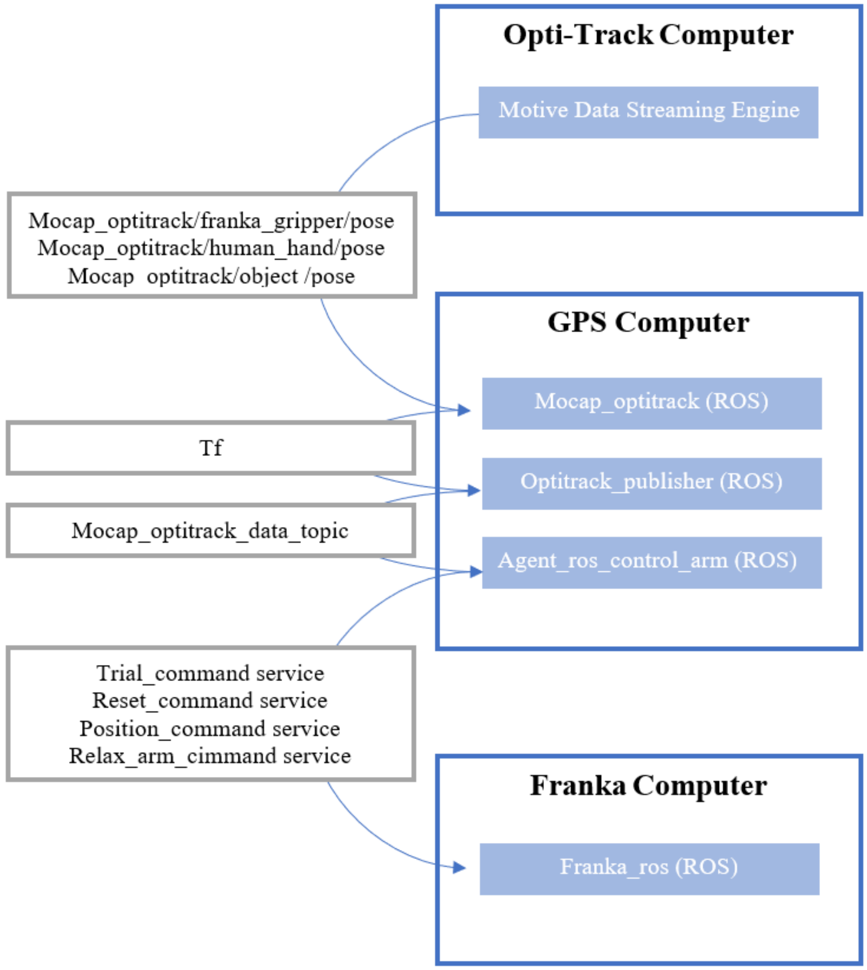


Figure 5: Distribution of ROS nodes across different computers and connections between them using ROS topics

4.4.5 Guided policy search suite

The BADMM-GPS implementation used in this thesis work was written by Chelsea Finn (Finn et al., 2016), at that time a researcher in Levine’s group. The addition of a ROS controller and GPS agent for a KUKA robot was developed by Jack White (White, 2018). Using Jack White's additions, we added an interface for the Panda robot.

4.4.5.1 GPS agent

In Finn's GPS implementation, the agent is the central component. Following loading up of the experimental configuration, the agent handles the running of the general policy training, the local policy generator/optimizer, and communicates with the controller. An agent class, stemming from Finn's base class, must be composed to communicate with the controller (to transmit actions to the controller and accept the state's transmission from the controller). However, selecting a controller (and robot or another process) fully depends on the user.

In the case of the Panda robot (similarly to the KUKA by Jack White), ROS topics are used to transmit and receive these quantities. Therefore, the agent must register as a ROS node and establish publishers for the GPS commands. The GPS sends to the robot the following commands:

- Get data: sends a request to the controller for the latest state and expects a response
- Relax arm: tells the controller to stop sending torques to the robot
- Reset arm: tells the controller to return the robot to the initial position specified for this round of trajectory optimizations—does not expect a response
- Trial command: sends the controller a policy and expects the return of a trajectory.

The GPS agents, implemented for different robots, vary merely in the communication method between the controller and the agent. In Panda's agent (similarly to the KUKA by Jack White), the only difference from the PR-2 controller was that this agent did not send commands to the passive arm and did not expect replies from the passive arm (as a part of general updates) since that the Panda and the KUKA are one-armed robots as opposed to the two-armed PR-2 robot.

4.4.5.2 GPS controller

In the context of the GPS algorithm, the GPS controller serves as an interface between the different kinds of robots or another conceptual control layer and the GPS agent. The controller created by Finn, named *RobotPlugin*, is a complex base class written in C++. Since this base class does not have any prior knowledge of the robot, or the hardware abstraction that will be used, separate classes, instantiated in *RobotPlugin*, abstract components in the following ways:

- Sensors:
  - The sensors are responsible for abstractions of physical state sensors, e.g., the states of the joints arriving from the robot (the robot joint angles, the robot joint velocities)
- Controllers:
  - Trial controller - commands the robot to perform sequential trials and return a set of trajectories.
  - Position controller - commands the robot to rest by an in-built PID controller.

Finn's software remained partially completed, regardless of the efforts invested in abstraction. The most notable gap is rooted in the fact that the *RobotPlugin* class assumes two physical robots. This assumption is made because it was originally implemented on a PR-2 robot, which has two arms. An actual PID controller runs on one arm and trial torques are sent, whereas, on the other arm, a dummy PID controller runs and no torques are sent. PID controllers and torque commands are required for any ROS controller derived directly from *RobotPlugin*.

In the suite provided by Finn is a derivative C++ class of *RobotPlugin*, called *PR2Plugin*, specifically designed for the PR-2 robot. Instead of managing two instances of a one-arm *RobotPlugin* class, *RobotPlugin* contains the code for two arms and *PR2Plugin* merely extends this with more PR-2-specific code. Directly inheriting the *RobotPlugin* class for a single robot arm is not possible due to the use of two arm trial controllers. Trying to implement GPS for a KUKA LWR4+ robot, Jack White implemented *one* arm based code by writing a new GPS controller, which is based on the *PR2Plugin* class and derived from the *RobotPlugin* class. He added an intermediate class, *SingleArmPlugin*, between *RobotPlugin* and *KUKALWRPlugin* to feed the variables related to the passive trial arm from the *RobotPlugin* class and to force it to not expect updates. We use the same GPS controller as in Jack White’s work, but modify it to work with the Panda robot and OptiTrack motion capture system. The structure of our Panda controller is depicted in Fig. 6. Our full experimental configuration is described in Appendix D. The changes are summarized below:

1. Tune PID parameters- A PID joint position controller is used to reset the arm before beginning the GPS trial/test. The controller commands the robot to move to a predefined position, defined

in terms of the joint angles. The ``proportional'' part of the controller applies control input proportional to the error between the current position and the target position. The ``integral'' part of the controller applies control input depending on the integration of error between the current position and the target position. The ``derivative'' part of the controller applies control input depending on the difference between the derivatives of the current position and the target position. The default parameters of this PID controller did not work with the real Panda robot. With the default parameters, the robot's joints did not move at all, only the tip joint would barely turn. We tuned the PID parameters to work with the Panda robot (Table 1). A large variety of different PID parameters were tried, but the integral and derivative gains had little effect. Making the proportional gain too high resulted in the robot crashing or abruptly halting as it exceeded the joint velocity limits.

Table 1: PID parameters before and after changes

Joint Number	LWR Values [Jack White reference]				Panda Values			
	P	I	D	I_clamp	P	I	D	I_clamp
Joint 1	2400	0	18	4	6	3	3	1
Joint 2	1200	0	20	4	6	3	3	1
Joint 3	1000	0	6	4	6	3	3	1
Joint 4	700	0	4	4	6	3	3	1
Joint 5	300	0	6	2	2.5	1	1	1
Joint 6	300	0	4	2	2.5	1	1	1
Joint 7	300	0	2	2	2.5	1	1	1

2. Tune the initial local controllers- The initial local controllers used in the GPS training process are linear gaussian controllers which try to hold the robot's initial position. The initial controller gains are computed with LQR, defined by the parameters described below. It is important to initialize these parameters to ensure that the robot starts the learning process while maintaining stability. The default parameters used in Finn's code for PR-2 or Jack White's code for LWR did not work with the Panda robot. With these parameters, the robot did not

move at all. The initial controller values that worked with the Panda robot were obtained by trial-and-error (Table 2).

a. Robot joint gains: - A vector of scalar gains, one for each torque/joint of the robot. These are used to guess the initial dynamics of the robot by LQR. The initial local controllers are extremely sensitive to these gains; a too high gain leads to exceed the joint limits, whereas a too low gain prevents the joint from moving at all.
b. Initial variance, stiffness, stiffness velocity– These three values are used to compute the Hessian of the loss with respect to trajectory at a single timestep. The initial variance affects the state-space explored by the robot in the initial training step. A higher initial variance results in larger explored state-space but with higher control inputs, which might exceed the joint limits in some cases causing the robot to halt. A lower initial variance results in smaller control inputs, but also a smaller explored state-space causing the robot to not learn the task.

Table 2: Initial controller values of PR-2, LWR, Panda robots.

Parameter	PR-2	LWR	Panda
Joint 1 gain	3.09	24	0.1
Joint 2 gain	1.08	12	0.1
Joint 3 gain	0.393	10	0.1
Joint 4 gain	0.674	7	0.1
Joint 5 gain	0.111	3	0.001
Joint 6 gain	0.152	3	0.001
Joint 7 gain	0.098	6	0.001
Initial variance	1	30	0.5
Stiffness	0.5	60	1.0
Stiffness velocity	0.25	0.25	0.5

3. Use Franka HW interface (franka_ros ROS package) instead of LWR HW interface (kuka_lwr ROS package)
4. Feed OptiTrack data to FrankaPlugin through ROSTopic sensor abstraction of GPS controller
5. Replace the former PC with a more powerful PC for controlling the robot- during our implementation trials of the GPS algorithm on the real Panda robot (Fig. 2), we encountered communication constraints violation errors. To overcome these, we tried replacing the network (network speed was 1000Mb/s) and the networks cables, but without success. Then, we tried using a local network (with no internet communication), which also had no effect, and also ran all the network tests, which showed no network issues. Finally, according to a consultation with the Franka support team, we replaced the PC with a more powerful PC (with an upgraded CPU. Further details are attached in Appendix E), which solved the communication constraints violation errors.
6. Clamp the torques- following plenty of trials and errors, we realized that the torques sent to the robot's joints needed to be limited to a maximum range to work without velocity or joint's position violation errors. First, we tried implementing penalization for out-of-limits velocities and torques, but with no success; The torques generated by the local controllers remained high. Therefore, we created a clamp function to limit the torques to a range of [-3,3] and added it to the trial controller class before the torques were sent to the robot's joints.
7. A new report ROS publisher- we re-encountered communication problems with the initiation of the training phase. These problems were manifested by sudden stops of the robot's movement during the training process. After substantial debugging efforts, we realized that the robot did not receive the published result of a trial of completion, sent from the real-time report ROS publisher (RobotPlugin). To address this issue, we wrote a non-real time ROS publisher which replaced the real-time report publisher written by (Finn et al., 2016).

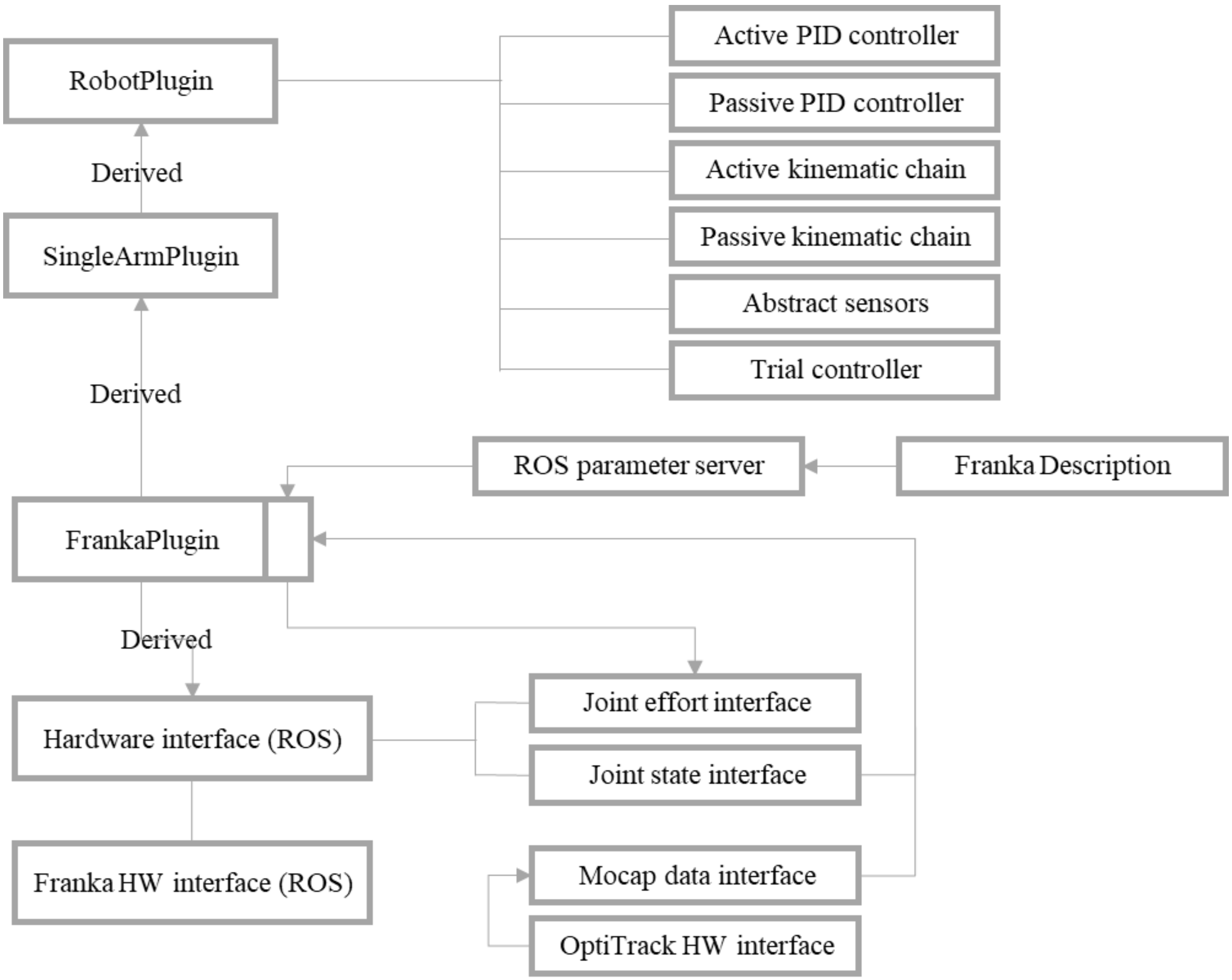


Figure 6: The structure of a Panda controller similar to LWR controller of (White, 2018).

4.5 Evaluation

We evaluated the performance of a robot controller that uses Guided Policy Search on a physical robot. Training is conducted both in a simulation environment and directly on the physical robot. The performance of the global policy was measured by calculating the error between the end-effector's position and the human hand's position at the last time step. To do so, we conducted two experiments:

4.5.1 Sim-to-Real

In the first experiment, we trained the Panda robot to perform handovers over repeated trials in a simulation environment and tested it on a real Panda robot on novel target trajectories. We found that the policy trained in simulation could not be transferred to the real robot, because the simulation model of the robot is different from the real robot in three ways:

1. **Torques and joints velocity limits.** Limited torques can be generated on the real robot (for joints 1-4: $-87[Nm] \leq \tau \leq 87[Nm]$ and for joints 5-7: $-12[Nm] \leq \tau \leq 12[Nm]$). The learned global policy in simulation generated very high torques (hundreds and even thousands Nm), and thus, could not be tested on the real robot. To reduce the torques computed in the simulation, we tried the following:
    - add to the cost function a penalization term for out of limit velocities and out of limit torques. The torques were reduced slightly but not enough to run the global policy on the real robot.
    - clamp the torques before the torques were sent to the robot's joints. In that case, the robot could not learn at all and barely moved from its initial position.
2. **Robot mass**. When the mass of the robot in the MuJoCo model was set the same as the mass of the real robot, MuJoCo required much higher values of joint torques to move the robot. One possible explanation is that the joint damping values in MuJoCo model were different from the real robot. We tried to scale down the mass of the MuJoCo model, but unfortunately, without success in improving the learning process in simulation. If the mass is too low, the robot overshoots the target (the robot arm seems like flying), and if the mass is high, the robot barely moves from its initial position. In both cases, the robot did not learn at all.

3. **Damping values**. As described above, the joint damping values in MuJoCo model were different from the real robot. The correct values were not provided by Franka-Emika. We tried to tune the damping to improve the MuJoCo model. If the damping values are too low, less friction acts on the joints and the robot does not hold its initial position and falls down. For high damping values, the robot does not move much.

Overall, tuning the MuJoCo model parameters to match the real robot parameters proved to be an infeasible solution. Thus, we decided to train the physical robot instead of a simulated robot, with a simulated target.

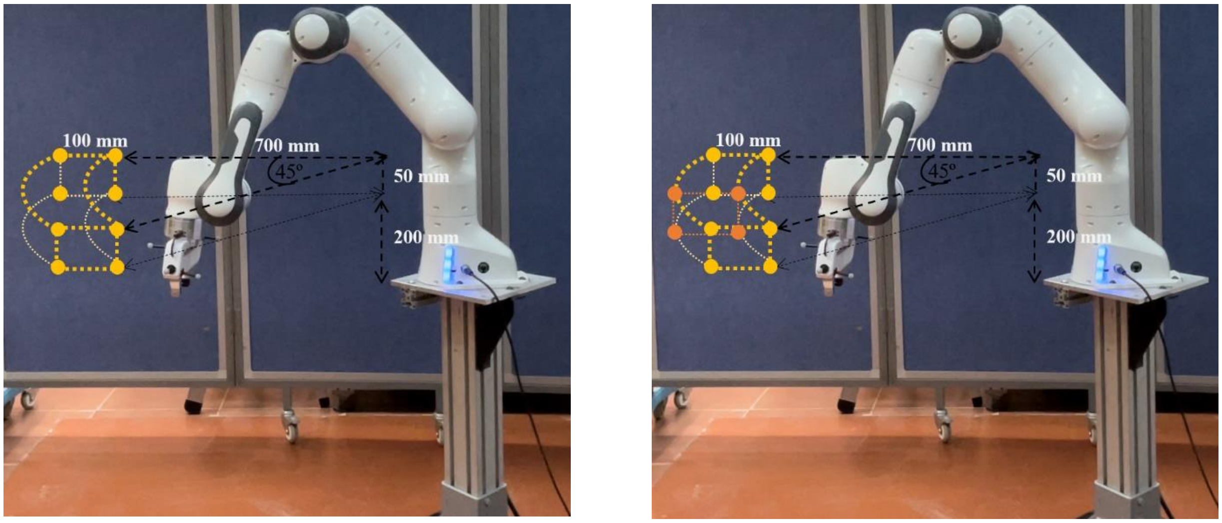
4.5.2 Real-to-Real

In the remaining text, we denote the Panda robot the “learner”, and the human is denoted as the “trainer” when we are in the training phase or the “tester” when we are in the testing phase.

In the second experiment we train and test the real Panda robot to perform handovers over repeated trials for two scenarios: large variations in target locations and moving targets. Since it is not practical to have a human trainer/tester perform exactly the same handover motion in all training iterations, we use recorded human hand motions during the training/testing process.

The first research question examined in our study is the spatial generalizability of the learned global policy, i.e., how does the global policy perform for significant spatial differences between training and testing locations.

To answer this question, we tried to test the learnt global policy at different locations of a static tester on a region around the learner robot, as shown in Fig. 7. The dimensions of the region are: Inner radius= 700 mm, Outer radius = 800 mm, Min height = 200 mm, Max height = 250 mm, Min angle = 0 º, max angle = 45º, measured from the robot’s base. This region was selected by trial and error to ensure that the robot does not run into joint position/velocity limits in the training/testing process. For each angle in 5deg increments, we test on a grid of 3×3 targets, resulting in 90 test locations. We compared two scenarios of local controllers: one with 8 local controllers and another with 12 local controllers. The global policy was trained with these local controllers for 11 iterations. Both the learner and the trainer/tester commenced their movement in each trail simultaneously. The learner's movement lasted 5 seconds, while the trainer/tester's movement lasted 1 second (which corresponds with the movement duration of humans in the reach phase of a handover). The test performance is measured as the mean error between the learner's gripper position and the tester's hand position over the last time step of each trial.



(a) (b)

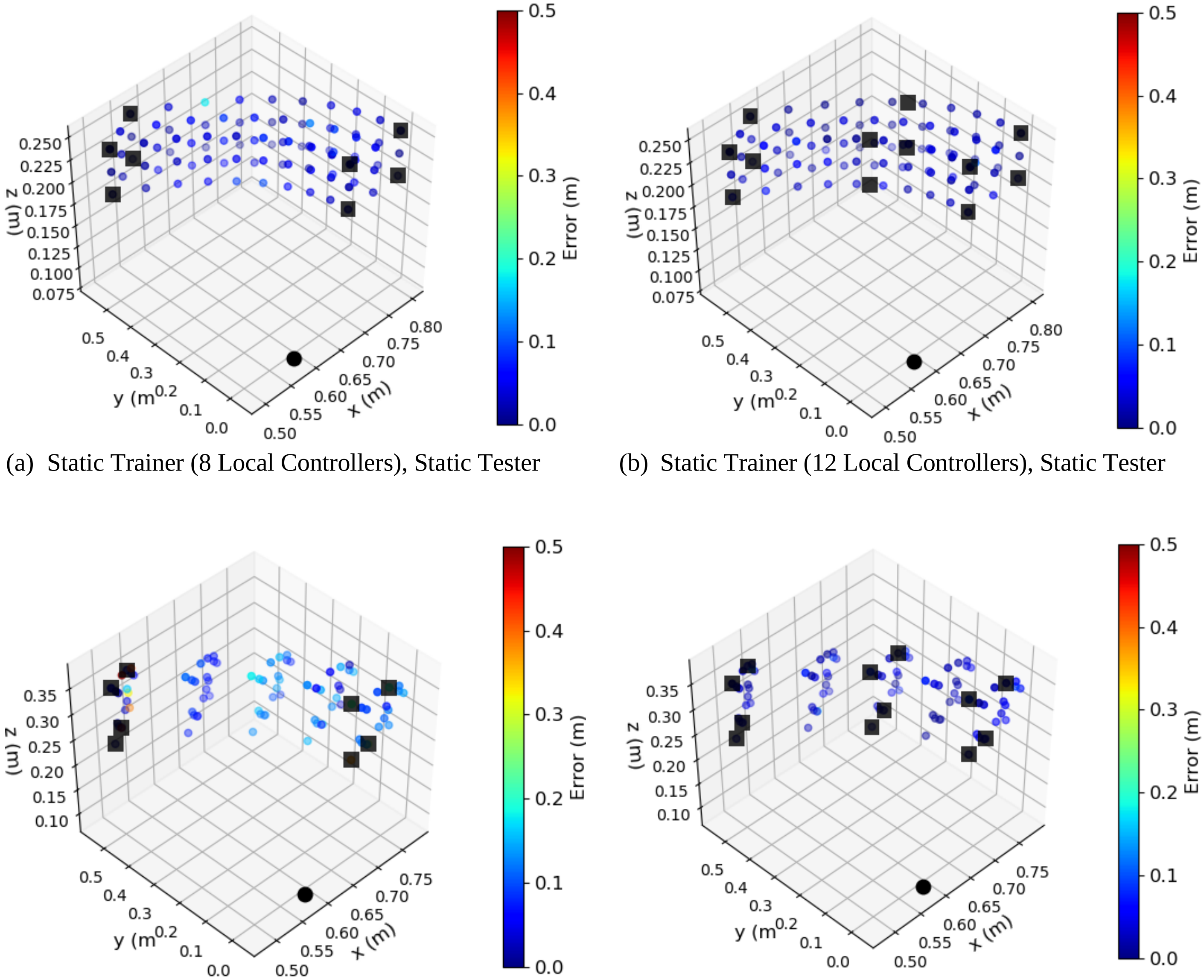
Figure 7: The training and testing region for: (a) 8 local controllers, and for (b) 12 local controllers. The yellow circles represent the initial 8 training locations, and the orange circles represent the additional 4 training locations that were located in a vertical plane dividing the workspace. This region was selected by trial and error to ensure that the robot does not run into joint position/velocity limits in the training/testing process.

The performance of the learned global policy is presented in Fig. 8(a). The black circle represents the learner's gripper's initial position, and the black squares represent the training locations. Mean error, range, and standard deviation are presented in Fig. 9 (left). The mean testing error (41.71 mm) is about twice as large as the mean training error (22.67 mm). As the test error can be reduced by adding more local controllers (Kshirsagar et al., 2021), we added 4 additional local controllers. They were located in a vertical plane dividing the workspace (Fig. 8(b)). The mean and standard deviation of the testing error of the global policy, trained with 12 local controllers, was reduced to 29±18 mm.

Next, we investigated how GPS performs when the target is moving. First, we used the same global policy shown in Fig. 8(a) (static training), but instead of a static tester, we used a moving target encoded in a recorded human reaching motion. The final position of the motion was in a region similar to the one shown in Fig. 7. The robot generated highly inefficient trajectories and reached areas outside of its Cartesian position limits, and thus, could not execute these trajectories.

A possible way to address this issue, as found in our previous study (Kshirsagar et al., 2021), is to train the controller with a moving target. We trained the robot with recorded human reaching motions, and tested the policy on another set of recorded human reaching motions. Some samples of these reaching motions are shown in the video attachment. Figures 8(c) and 8(d) show the performance of the global policy for various final positions of the tester's gripper, defined as in previous trials. Fig. 9 (right) shows error distributions.

For the global policy trained with a moving trainer and 8 local controllers (Fig. 8(c)), the mean testing error is 124.28 mm. Although the test errors are high as compared to the static tester scenario, the robot stayed within the joint and Cartesian limits. Moreover, the variance over target location is high, and the worst-case error is 791.11 mm, 442% higher than the maximum error for static tester condition (179 mm). Surprisingly, this maximum error occurred for a test motion close to one of the training motions. This could be attributed to the highly non-linear nature of the global policy. Interestingly, GPS did not converge to a low training error, which was 123.23 mm, 544% higher than for static training (22.67 mm). Training the global policy with a moving trainer and 12 local controllers (Fig. 8(d)), reduced the mean testing error to 37.93 mm. The worst-case error also improved (138.71 mm). An inspection of the generated trajectories and torques shows that this approach results in trajectories and torques similar to those achieved with static targets. Distributions of training and testing performance for each target scenario are presented in Fig. 9. Each point is the mean error between the learner's gripper position and the tester's hand position over the last time step of a trial. Error bars show one standard deviation around the mean of each distribution.



(a) Static Trainer (8 Local Controllers), Static Tester

(b) Static Trainer (12 Local Controllers), Static Tester

(c) Moving Trainer (8 Local Controllers), Moving Tester

(d) Moving Trainer (12 Local Controllers), Moving Tester

Figure 8: Global policy evaluation for different types of trainers and testers. The black circle represents the learner's gripper's initial position, and the black squares represent the training locations. In the 'static' case, the trainer/tester stays in a fixed configuration. In the 'moving' case, the trainer/tester moves with a human-like trajectory (that were recorded in advanced) and reaches the locations given by colored dots. Thus, each point corresponds to the final position of the tester's gripper in a trial. Error between the learner's gripper position and the tester's gripper position is calculated over the last time step of each trial.

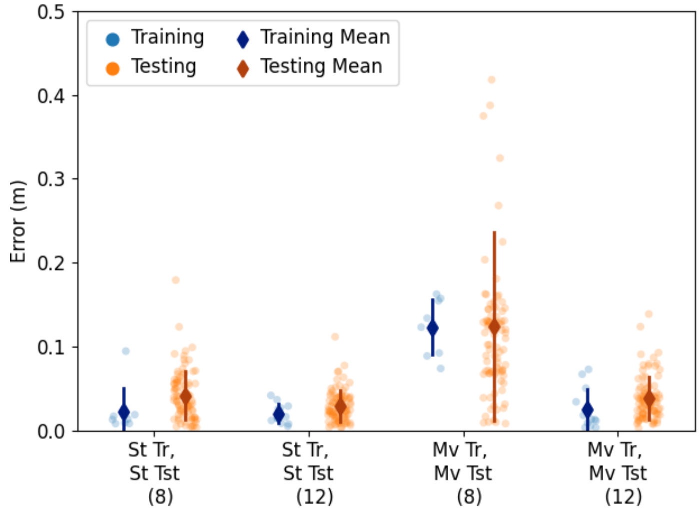


Figure 9: Distributions of training and testing performance for each target scenario. Each point is the mean error between the learner's gripper position and the tester's hand position over the last time step of a trial. Error bars show one standard deviation around the mean of each distribution.

4.6 Conclusions and future work

Our work evaluated the feasibility of GPS as a learning method for human-robot handovers in a real-world environment for large variations in target locations and for moving targets. Training was analyzed both in a simulation environment and directly on a physical robot. We used a variant of GPS that does not require prior knowledge of robot dynamics. Instead, it learns locally linear dynamics models from the training data (Levine et al., 2016). Prior studies used GPS for autonomous manipulation (Chebotar et al., 2017; Levine et al., 2016; Levine et al., 2015; Levine & Abbeel, 2014) and locomotion tasks (Zhang et al., 2016; Levine & Abbeel, 2014;Levine & Koltun, 2013, Levine & Koltun, 2013b) which are characterized by small variations in target locations and a static environment. However in a handover task, the robot operates in a dynamic environment due to unpredictable and non-static human behavior, resulting in a wider spread of target locations. These challenges have been addressed in a recently published study by (Kshirsagar et al., 2021). In this study, the potential of GPS to train a robot controller for human-robot object handovers in a simulation environment has been explored. Despite uncovering important insights, their application to a real-world environment is limited, as this study showed.

Unlike a real-world environment, which warrants constant human supervision (for resetting experiments, monitoring hardware status, and ensuring safety), data can be continuously obtained with no need for human intervention in simulation. Hence, a simulation environment is faster, cheaper, and safer than experimenting on a real robot. However, the reality gap is a significant obstacle, preventing learning to robotic's applications. In simulations, for instance, the robots can learn to perform bicycle stunts (Tan et al.,2014), while in the real world, it is still challenging to teach robots basic tasks like walking. To fully exploit robotic's potential benefits, bridging the reality gap is crucial. This bridging would result in a better simulation benchmark for robotics, focusing the research efforts on the most pressing robot learning challenges. In this study, we first tried to learn the policy in a simulation environment and then deployed it to the real robot. It was found to be an infeasible solution, as the MuJoCo model parameters did not match the real robot parameters. Thus, we decided to train directly on the physical robot.

We found that it was not possible to train the physical robot for the same target locations used in (Kshirsagar et al., 2021) as the Panda robot always ran into some joint velocity or Cartesian position limits during the training process. Thus we had to reduce the robot's target workspace by trial-and-error to avoid these limits. In this reduced workspace (Fig. 7), we found that when the robot was trained to reach only static target locations, the global policy performance could be slightly improved by adding local controllers in regions with highest test errors (in the middle of the working plane) (Fig. 8(a) compared to Fig. 8(b)).

When evaluating the global policy trained with static targets on a moving test target, the robot generated highly inefficient trajectories and reached areas outside of its Cartesian position limits. To overcome this issue, we trained the global policy with moving targets. Nevertheless, this solution was not free of drawbacks. It successfully reduced the mean error and resulted in more efficient and low-torque trajectories, but resulted in a high-variance (unreliable) global policy with significantly larger worst-case errors. This issue can be addressed by adding local controllers to the training phase, improving the global policy performance (Fig. 8(d)).

This study introduces preliminary steps toward implementing GPS in a real-world environment for human-robot handovers. Nevertheless, we did not take into account numerous essential aspects of handovers, such as the robot's movement legibility and the human's adaptation to the robot's movements. Our studies were also conducted in a limited workspace. The workspace selected for the learning process was relatively small because the robot ran into joint or cartesian limits in the training phase of a larger workspace. The robot's low-level controller had inbuilt safety stops that interfered with the robot controller whenever it reached any joint/Cartesian position/velocity/torque limits (Appendix F). It was not possible to override these limits, which made it difficult to train the robot. To examine the GPS algorithm for a larger workspace, we recommend to use a robot that allows overriding these limits. Also, there is a need to develop GPS algorithms that will train local controllers and global policy while obeying these limits. Despite these aforementioned limitations, this study contributes to the understanding of the challenges and applicability of GPS in a real-world context. Also, it demonstrates the potential benefits and drawbacks of GPS as an algorithmic tool to further develop the field of human-robot collaboration in general and the area of human-robot handovers in specific.

Chapter 5. Summary

In recent years, we have witnessed a substantial shift towards a more direct human-robot collaboration in the industry. The technological advances in robot hardware have enabled researchers from the industry to envision an entirely shared environment. Robots would interact with and act on their surroundings in this foreseen environment, including other agents like human workers and robots. The recent COVID-19 pandemic has highlighted the need for developing independently operating as well as collaborative robots to use in additional fields such as the medical field (hospitals and care homes). In this context, robots must be developed with the abilities to exchange objects for successful cooperation and to collaborate in manipulation tasks.

In this thesis, we focused on two main aspects regarding human-to-robot handovers. In the first study we analyzed what are the most frequent gaze behaviors in a human-human handover. we found that the most common gaze behaviors of receivers were: hand-face, face-hand-face and hand gaze. Then, with the purpose of implementing these behaviors on a collaborative robot, we investigated whether and to what extent the user's preference of the robot's gaze, when it is receiving an object from the human, and is this dependent on the object size and type and on different human-robot configurations. We performed two types of user studies (video and in-person) with a collaborative robot that exhibited these gaze behaviors while receiving an object from a human. To investigate the effect of object's size, object's fragility or the human's posture on human's preferences for the robot gaze, objects of different sizes (a small box and a large box), different fragility (a plastic bottle and a glass bottle) and different giver's posture (standing and sitting) were used. The results of both studies were similar. The participants preferred the gaze behavior in which the robot initially looked at their face, then transitioned its gaze to their hand (during the reach phase and the transfer phase) and then transitioned its gaze back to look at their face again (during the retreat phase). Open-ended responses suggested that the change between looking at the giver's face and then at the giver's hand and then back at the giver's face portrayed the robot as more humanlike, natural, and friendly. Also, they felt that this behavior complemented the robot's handover. people preferred the robot looking at their face at the beginning and the end of the handover, and the robot's eyes following the object during the transfer phase. This gaze behavior complemented the robot's handover motion, and thus portrayed the robot as more human-like, natural, and friendly. Another possible explanation is that the social aspects of a human receiver are implicit, whereas a robot has to establish its social agency for a better handover experience. Based on these findings, we recommend to HRI designers to implement a Face-Hand-Face transition gaze when the robot receives an object from a human, regardless of human posture and characteristics of the object being handed over. There are several limitations of this study which could motivate future work. The results are limited by the sample size and the specific cultural and demographic makeup of its participants. Larger population samples of different age groups, backgrounds, and cultures should be investigated to help generalize the findings of our experiments. It would also be interesting to evaluate if the specific gazes are dependent on the population sample (age group, experience with technology, extrovert vs. introvert), task (time critical, entertainment), environment (industry/hospital/restaurant) and robot (e.g., reliability, motion smoothness).

Moreover, as with any experimental study, there is a question of external validity. A handover that is part of a more complex collaborative or assistive task might elicit different expectations of the robot's gaze, a fact that should be considered by designers of HRI systems. To better understand these contextual requirements, additional realistic scenarios of assistive and collaborative tasks should be considered.

According to the results of our first study, with correlation to the relevant literature, we discover that other key components of HRI, which may influence human’s acceptance of robotics, are the perceived naturalness and smoothness of the robot's movements. Therefore, we decided to pursue our second study regarding human-robot handover, implementing an online controller to produce reaching motion of the robot to further develop the acceptance and practical use of collaborative robots in the industry. In the second study we developed a robot controller that uses Guided Policy Search (GPS) to perform object handovers and evaluated the effect of different training scenarios (simulation and physical robot) on performance. We evaluated the controller with a physical robot while the training was conducted both in a simulation environment and directly on the physical robot. In the first experiment, in an attempt to bridge the reality gap from simulation, we wanted to test the policy learnt in the simulation environment on the real robot. To do so, we tried to tune the MuJoCo model parameters to match the real robot parameters. It was proved to be an infeasible solution because the learned global policy in simulation generated very high torques (hundreds and even thousands Nm), and thus, could not be tested on the real robot. Therefore, we decided to train the physical robot instead of a simulated robot, with a simulated target.

In the second experiment, we train and test the real Panda collaborative robot to perform handovers over repeated trials for two scenarios: large variations in target locations and moving targets. The first research question examined in our study was how does the GPS perform for significant spatial differences between training and testing locations. We found the global policy performance slightly improved by using 12 local controllers. The second research question examined in our study was how does the GPS perform with moving targets. First, we used the global policy trained with static targets, but instead of a static tester, we used a recorded human reaching motion. In this case, the robot generated highly inefficient trajectories and reached areas outside of its cartesian position limits. To address this issue, we trained and test the robot with moving targets. It successfully reduced the mean error and resulted in more understandable and low-torque efficient trajectories, but resulted in a more high-variance (unreliable) global policy with significantly larger worst-case errors. This issue can be addressed by adding local controllers to the training phase, improving the global policy performance.

This study contributes to the knowledge regarding the applicability of GPS in a real-world context. Also, it demonstrates the potential benefits and the drawbacks of GPS as a tool to further develop the field of human-robot collaboration. We did not take into account numerous essential aspects of handovers, such as the robot's movement legibility and the human's adaptation to the robot's movements. Our studies were also conducted in a limited workspace. During the training process the Panda robot ran into some joint velocity or cartesian position, so we had to reduce the robot's target workspace by trial-and-error to avoid these limits. To examine the GPS algorithm for a larger workspace, we recommend to use a robot that allows overriding these limits. Also, there is a need to develop GPS algorithms that will train local controllers and global policy while obeying these limits.

References

Argyle, M., & Cook, M. (1976). Gaze and mutual gaze. In *Gaze and mutual gaze.* Oxford, England: Cambridge U Press.

Argyle, M., & Ingham, R. (1972). Gaze, Mutual Gaze, and Proximity. *Semiotica*, Vol. 6, p. 32. https://doi.org/10.1515/semi.1972.6.1.32

Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. *Portions of This Paper Were Presented at the Society for Research in Child Development Conference, New Orleans, Mar 1993; the British Psychological Society, Welsh Branch," Faces" Conference, U Wales Coll of Cardiff, Sep 1993; and the British Society for T*. Lawrence Erlbaum Associates, Inc.

Basili, P., Huber, M., Brandt, T., Hirche, S., & Glasauer, S. (2009). Investigating Human-Human Approach and Hand-Over. In H. Ritter, G. Sagerer, R. Dillmann, & M. Buss (Eds.), *Human Centered Robot Systems: Cognition, Interaction, Technology* (pp. 151–160). https://doi.org/10.1007/978-3-642-10403-9_16

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*(3), 412–414.

Bdiwi, M., Kolker, A., Suchý, J., & Winkler, A. (2013). Automated assistance robot system for transferring model-free objects from/to human hand using vision/force control. *International Conference on Social Robotics*, 40–53. Springer.

Cakmak, M., Srinivasa, S. S., Min Kyung Lee, Forlizzi, J., & Kiesler, S. (2011a). Human preferences for robot-human hand-over configurations. *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1986–1993. https://doi.org/10.1109/iros.2011.6094735

Cakmak, M., Srinivasa, S. S., Min Kyung Lee, Forlizzi, J., & Kiesler, S. (2011b). *Human preferences for robot-human hand-over configurations*. (May 2014), 1986–1993. https://doi.org/10.1109/iros.2011.6094735

Cakmak, Maya, Srinivasa, S. S., Kyung Lee, M., Kiesler, S., & Forlizzi, J. (2011). Using spatial and temporal contrast for fluent robot-human hand-overs. *HRI 2011 - Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction*, 489–496. https://doi.org/10.1145/1957656.1957823

Cakmak, Maya, Srinivasa, S. S., Lee, M. K., Kiesler, S., & Forlizzi, J. (2011). Using Spatial and Temporal Contrast for Fluent Robot-Human Hand-Overs. *Proceedings of the 6th International Conference on Human-Robot Interaction*, 489–496. https://doi.org/10.1145/1957656.1957823

Cañigueral, R., & Hamilton, A. F. d. C. (2019). The role of eye gaze during natural social interactions in typical and autistic people. *Frontiers in Psychology*, *10*(MAR), 1–18.

https://doi.org/10.3389/fpsyg.2019.00560

Carfì, A., Foglino, F., Bruno, B., & Mastrogiovanni, F. (2019). A multi-sensor dataset of human-human handover. *Data in Brief*, *22*, 109–117. https://doi.org/10.1016/j.dib.2018.11.110

Casa, G. A., Cervera, E., Moughlbay, A. A., Alemany, J., & Martinet, P. (2015). *ROS-based online robot programming for remote education and training ROS-based Online Robot Programming for Remote Education and Training **. (February 2019). https://doi.org/10.1109/ICRA.2015.7140055

Cassell, J. (2001). Embodied conversational agents: representation and intelligence in user interfaces. *AI Magazine*, *22*(4), 67.

Chebotar, Y., Kalakrishnan, M., Yahya, A., Li, A., Schaal, S., & Levine, S. (2017). Path integral guided policy search. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 3381–3388. IEEE.

Cherubini, A., Passama, R., Crosnier, A., Lasnier, A., & Fraisse, P. (2016). Collaborative manufacturing with physical human–robot interaction. *Robotics and Computer-Integrated Manufacturing*, *40*, 1–13. https://doi.org/10.1016/j.rcim.2015.12.007

Cook, M. (1977). Gaze and Mutual Gaze in Social Encounters: How long—and when—we look others "in the eye" is one of the main signals in nonverbal communication. *American Scientist*, *65*(3), 328–333. Retrieved from http://www.jstor.org/stable/27847843

Dehais, F., Sisbot, E. A., Alami, R., & Causse, M. (2011). Physiological and subjective evaluation of a human-robot object hand-over task. *Applied Ergonomics*, *42*(6), 785–791. https://doi.org/10.1016/j.apergo.2010.12.005

Deisenroth, M. P., Neumann, G., & Peters, J. (2013). A survey on policy search for robotics. *Foundations and Trends in Robotics*, *2*(1–2), 388–403.

Du, J., Fu, J., & Li, C. (2021). Guided Policy Search Methods: A Review. *Journal of Physics: Conference Series*, *1748*(2). https://doi.org/10.1088/1742-6596/1748/2/022039

Duan, F., Tan, J. T. C., & Arai, T. (2011). A new human-robot collaboration assembly system for cellular manufacturing. *Proceedings of the 30th Chinese Control Conference*, *CCC 2011*, 5468–5473.

Duan, F., Tan, J. T. C., Tong, J. G., Kato, R., & Arai, T. (2012). Application of the assembly skill transfer system in an actual cellular manufacturing system. *IEEE Transactions on Automation Science and Engineering*, *9*(1), 31–41. https://doi.org/10.1109/TASE.2011.2163818

Edsinger, A., & Kemp, C. C. (2007). Human-robot interaction for cooperative manipulation: Handing objects to one another. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 1167–1172. https://doi.org/10.1109/ROMAN.2007.4415256

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, *24*(6), 581–604.

Finn, C., Zhang, M., Fu, J., Tan, X., McCarthy, Z., Scharff, E., & Levine, S. (2016). Guided policy search code implementation, 2016. *Software Available from Rll. Berkeley. Edu/Gps*.

Fitzgerald, C. (2013). Developing Baxter. *IEEE Conference on Technologies for Practical Robot Applications , TePRA*, pp 1–6.

Franka Emika GmbH. (2020). *Panda's Instruction Handbook*.

Gibson, J. J., & Pick, A. D. (1963). Perception of another person's looking behavior. *The American Journal of Psychology*, *76*(3), 386–394.

Glasauer, S., Huber, M., & Basili, P. (2010). Interacting in time and space: Investigating human-human and human-robot joint action. *RO-MAN*, *2010*, 252–257.

Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, *136*, 359–364. https://doi.org/10.1016/j.cognition.2014.11.040

Goffman, E. (1963). *Behavior in public places. Glencoe*. Free Press.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, *26*(3), 245–256.

Hentout, A., Aouache, M., Maoudj, A., & Akli, I. (2019). Human–robot interaction in industrial collaborative robotics: a literature review of the decade 2008–2017. *Advanced Robotics*, *33*(15–16), 764–799. https://doi.org/10.1080/01691864.2019.1636714

Hoffman, G., & Breazeal, C. (2007). Cost-based anticipatory action selection for human-robot fluency. *IEEE Transactions on Robotics*, *23*(5), 952–961. https://doi.org/10.1109/TRO.2007.907483

Hoffman, G., & Breazeal, C. (2009). Effects of anticipatory perceptual simulation on practiced human-robot tasks. *Autonomous Robots*, *28*(4), 403–423. https://doi.org/10.1007/s10514-009-9166-3

Huber, M., Rickert, M., Knoll, A., Brandt, T., & Glasauer, S. (2008). Human-robot interaction in handing-over tasks. *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, 107–112. https://doi.org/10.1109/ROMAN.2008.4600651

Ibarz, J., Tan, J., Finn, C., Kalakrishnan, M., Pastor, P., & Levine, S. (2021). How to train your robot with deep reinforcement learning: lessons we have learned. *The International Journal of Robotics Research*, *40*(4–5), 698–721. https://doi.org/10.1177/0278364920987859

Kaelbling, L. P. (2020). The foundation of efficient robot learning. *Science*, *369*(6506), 915–916. https://doi.org/10.1126/science.aaz7597

Kiesler, S. (2005). Fostering common ground in human-robot interaction. *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, 729–734. IEEE.

Koay, K. L., Sisbot, E. A., Syrdal, D. S., Walters, M. L., Dautenhahn, K., & Alami, R. (2007). Exploratory study of a robot approaching a person in the context of handing over an object. *AAAI Spring Symposium - Technical Report*, *SS-07-07*, 18–24.

Krüger, J., Lien, T. K., & Verl, A. (2009). Cooperation of human and machines in assembly lines. *CIRP Annals - Manufacturing Technology*, *58*(2), 628–646. https://doi.org/10.1016/j.cirp.2009.09.009

Kshirsagar, A., Hoffman, G., & Biess, A. (2021). Evaluating Guided Policy Search for Human-Robot Handovers. *IEEE Robotics and Automation Letters*, *6*(2), 1–1. https://doi.org/10.1109/lra.2021.3067299

Kshirsagar, A., Kress-Gazit, H., & Hoffman, G. (2019). Specifying and Synthesizing Human-Robot Handovers. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5930–5936. IEEE.

Kshirsagar, A., Lim, M., Christian, S., & Hoffman, G. (2020). *Effects of Robot Gaze Behaviors in Human-to-Robot Handovers*.

Kuo, C.-H. (2020). *Robotics and mechatronics : proceedings of the 6th IFToMM International Symposium on Robotics and Mechatronics (ISRM 2019)* (1st ed. 20).

Kupcsik, A., Hsu, D., & Lee, W. S. (2018). Learning dynamic robot-to-human object handover from human feedback. In *Robotics research* (pp. 161–176). Springer.

Leal, D., & Yihun, Y. (2019). Progress in Human-Robot Collaboration for Object Handover. *2019 IEEE International Symposium on Measurement and Control in Robotics (ISMCR)*, C3-2. IEEE.

Levine, S, Wagener, N., & Abbeel, P. (2015). Learning contact-rich manipulation skills with guided policy search (2015). *ArXiv Preprint ArXiv:1501.05611*.

Levine, Sergey, & Abbeel, P. (2014). Learning neural network policies with guided policy search under unknown dynamics. *Advances in Neural Information Processing Systems*, *27*, 1071–1079.

Levine, Sergey, Finn, C., Darrell, T., & Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *The Journal of Machine Learning Research*, *17*(1), 1334–1373.

Levine, Sergey, & Koltun, V. (2013a). Guided policy search. *International Conference on Machine Learning*, 1–9.

Levine, Sergey, & Koltun, V. (2013b). Variational policy search via trajectory optimization. *Advances in Neural Information Processing Systems*, *26*, 207–215.

Levine, Sergey, & Koltun, V. (2014). Learning complex neural network policies with trajectory optimization. *International Conference on Machine Learning*, 829–837.

Maeda, G., Ewerton, M., Lioutikov, R., Amor, H. Ben, Peters, J., & Neumann, G. (2014). Learning interaction for collaborative tasks with probabilistic movement primitives. *2014 IEEE-RAS International Conference on Humanoid Robots*, 527–534. IEEE.

Magrini, E., Ferraguti, F., Ronga, A. J., Pini, F., De Luca, A., & Leali, F. (2020). Human-robot coexistence and interaction in open industrial cells. *Robotics and Computer-Integrated Manufacturing*, *61*, 101846. https://doi.org/https://doi.org/10.1016/j.rcim.2019.101846

Mainprice, J., Gharbi, M., Simeon, T., & Alami, R. (2012). Sharing effort in planning human-robot handover tasks. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 764–770. https://doi.org/10.1109/ROMAN.2012.6343844

Medina, J. R., Duvallet, F., Karnam, M., & Billard, A. (2016). A human-inspired controller for fluid human-robot handovers. *IEEE-RAS International Conference on Humanoid Robots*, 324–331. https://doi.org/10.1109/HUMANOIDS.2016.7803296

Micelli, V., Strabala, K., & Srinivasa, S. (2011). *Perception and control challenges for effective human-robot handoffs*.

Moon, Aj., Troniak, D. M., Gleeson, B., Pan, M. K. X. J., Zheng, M., Blumer, B. A., … Croft, E. A. (2014). Meet Me Where i'm Gazing: How Shared Attention Gaze Affects Human-Robot Handover Timing. *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, 334–341. https://doi.org/10.1145/2559636.2559656

Mutlu, B. (2009). *Designing Gaze Behavior for Humanlike Robots*. https://doi.org/TR-CMU-HCII-09-101

Nass, C., & Steuer, J. (1993). Voices, boxes, and sources of messages: Computers and social actors. *Human Communication Research*, *19*(4), 504–527.

Pan, M. K. X. J., Croft, E. A., & Niemeyer, G. (2018). Exploration of geometry and forces occurring within human-to-robot handovers. *2018 IEEE Haptics Symposium (HAPTICS)*, 327–333. IEEE.

Pan, M. K. X. J., Knoop, E., Bächer, M., & Niemeyer, G. (2019). Fast handovers with a robot character: Small sensorimotor delays improve perceived qualities. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 6735–6741. IEEE.

Pcr, A., & Kit, L. (2012). Data Sheet Data Sheet. 고생물학회지, (September 2004), 0–1. Retrieved from http://www.papersearch.net/view/detail.asp?detail_key=10000715

Peternel, L., Kim, W., Babič, J., & Ajoudani, A. (2017). Towards ergonomic control of human-robot co-manipulation and handover. *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, 55–60. IEEE.

Prada, M., Remazeilles, A., Koene, A., & Endo, S. (2014). Implementation and experimental validation of Dynamic Movement Primitives for object handover. *IEEE International Conference on Intelligent Robots and Systems*, 2146–2153. https://doi.org/10.1109/IROS.2014.6942851

Rasch, R., Wachsmuth, S., & König, M. (2019). An evaluation of robot-to-human handover configurations for commercial robots. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 7588–7595. IEEE.

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge university press.

Riccio, F., Capobianco, R., & Nardi, D. (2016). Learning human-robot handovers through π-STAM: Policy improvement with spatio-temporal affordance maps. *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, 857–863. IEEE.

Risko, E. F., Richardson, D. C., & Kingstone, A. (2016). Breaking the Fourth Wall of Cognitive Science: Real-World Social Attention and the Dual Function of Gaze. *Current Directions in Psychological Science*, *25*(1), 70–74. https://doi.org/10.1177/0963721415617806

Robotics, I. F. of. (2019). World Robotics 2019 Industrial Robots. *Executive Summary World Robotics 2019 Industrial Robots*, 13–16. Retrieved from https://ifr.org/downloads/press2018/Executive Summary WR 2019 Industrial Robots.pdf

Sayfeld, L., Peretz, Y., Someshwar, R., & Edan, Y. (2017). Evaluation of Human-Robot Collaboration Models for Fluent Operations in Industrial Tasks. *ArXiv Preprint ArXiv:1708.04790*.

Scimmi, L. S., Melchiorre, M., Mauro, S., & Pastorelli, S. (2019). Experimental real-time setup for vision driven hand-over with a collaborative robot. *2019 International Conference on Control, Automation and Diagnosis (ICCAD)*, 1–5. IEEE.

Simmel, G. (1921). Sociology of the senses: Visual interaction. *Introduction to the Science of Sociology*, *3*.

Sisbot, E. A., & Alami, R. (2012). A human-aware manipulation planner. *IEEE Transactions on Robotics*, *28*(5), 1045–1057.

Snyder, M., Grather, J., & Keller, K. (1974). Staring and compliance: A field experiment on hitchhiking. *Journal of Applied Social Psychology*, *4*(2), 165–170.

Someshwar, R. (2017). Human-robot synchronization for time-critical tasks, PhD thesis. Dept. of Industrial Engineering and Management, Ben-Gurion University of the Negev.

Someshwar, R., & Edan, Y. (2017). Investigating Joint-Action in Short-Cycle Repetitive Handover Tasks: The Role of Giver Versus Receiver and its Implications for Human-Robot Collaborative System Design. *International Journal of Social Robotics*, 1–16. https://doi.org/10.1007/s12369-017-0424-9

Someshwar, R., Meyer, J., & Edan, Y. (2012). Models and methods for HR synchronization. *IFAC Proceedings Volumes*, *45*(6), 829–834.

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction*, *11*(2), 97–124.

Strabala, K., Lee, M. K., Dragan, A. D., Forlizzi, J. L., Srinivasa, S., Cakmak, M., & Micelli, V. (2013). Towards Seamless Human-Robot Handovers. *Journal of Human-Robot Interaction*, *2*(1), 112–132. https://doi.org/10.5898/jhri.2.1.strabala

Tan, J., Gu, Y., Liu, C. K., & Turk, G. (2014). Learning bicycle stunts. *ACM Transactions on Graphics (TOG)*, *33*(4), 1–12.

Tan, J. T. C., Duan, F., Zhang, Y., Watanabe, K., Kato, R., & Arai, T. (2009). Human-Robot Collaboration in Cellular Manufacturing: Design and Development. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009*, 29–34. https://doi.org/10.1109/IROS.2009.5354155

Tantawi, K. H., Sokolov, A., & Tantawi, O. (2019). Advances in industrial robotics: From industry 3.0 automation to industry 4.0 collaboration. *2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-ICON)*, 1–4. IEEE.

Todorov, E., Erez, T., & Tassa, Y. (2012). Mujoco: A physics engine for model-based control. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 5026–5033. IEEE.

Tsarouchi, P., Spiliotopoulos, J., Michalos, G., Koukas, S., Athanasatos, A., Makris, S., & Chryssolouris, G. (2016). A Decision Making Framework for Human Robot Collaborative Workplace Generation. *Procedia CIRP*, *44*, 228–232. https://doi.org/10.1016/j.procir.2016.02.103

Umbrico, A., Cesta, A., Cortellessa, G., & Orlandini, A. (2020). A holistic approach to behavior adaptation for socially assistive robots. *International Journal of Social Robotics*, *12*(3), 617–637.

Unhelkar, V. V, Siu, H. C., & Shah, J. A. (2014). Comparative Performance of Human and Mobile Robotic Assistants in Collaborative Fetch-and-Deliver Tasks. *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, 82–89. https://doi.org/10.1145/2559636.2559655

Vogt, D., Stepputtis, S., Jung, B., & Amor, H. Ben. (2018). One-shot learning of human–robot handovers with triadic interaction meshes. *Autonomous Robots*, *42*(5), 1053–1065.

White, J. (2018). *Guided policy search for a lightweight industrial robot arm*.

Wong, K. K., Atikhah, N., Samah, A., Sahimi, M. S., & Othman, W. A. F. W. (2019). Development of Reverse Vending Machine using Recycled Materials and Arduino Microcontroller. *International Journal of Engineering Creativity and Innovation (IJECI)*, *1*(1), 7–16.

Yamane, K., Revfi, M., & Asfour, T. (2013). Synthesizing object receiving motions of humanoid robots with human motion database. *Proceedings - IEEE International Conference on Robotics and Automation*, 1629–1636. https://doi.org/10.1109/ICRA.2013.6630788

Yang, W., Paxton, C., Cakmak, M., & Fox, D. (2020). Human grasp classification for reactive human-to-robot handovers. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 11123–11130. IEEE.

Zhang, T., Kahn, G., Levine, S., & Abbeel, P. (2016). Learning deep control policies for autonomous aerial vehicles with mpc-guided policy search. *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 528–535. IEEE.

Zhao, X., Chumkamon, S., Duan, S., Rojas, J., & Pan, J. (2018). Collaborative human-robot motion generation using LSTM-RNN. *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, 1–9. IEEE.

Zlatanov, N. (2016). Arduino and Open Source Computer Hardware and Software. *Journal of Water, Sanitation and Hygiene for Development*, *10*(11), 1–8. https://doi.org/10.13140/RG.2.1.1071.7849

Appendices

Appendix A- Robotic system development

System description

The system (Fig. 10) includes a robot arm receiving an object from a human, a distance sensor to detect the giver's movement, and an infrared proximity sensor placed on the robot arm to detect the object distance from the robot gripper. The sensors are controlled by an Arduino microcontroller, which transmits the data to the robot.

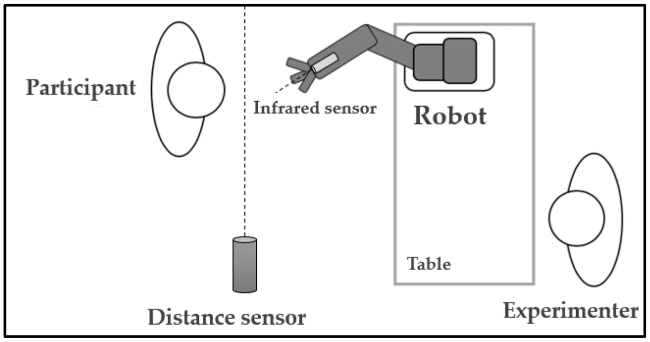


Figure 10: The experimental setup

Hardware

The system consists of several components connected to a single computer. The components include a Sawyer robot, distance sensor, photoelectric sensor, and Arduino nano development board.

Sawyer robot

A Sawyer robot was used for the experiments (Fig. 11). The robot arm is autonomous and programmed to reach a predefined position once the handover begins. The robot grasps the object when the object gets close enough. Finally, the robot retreats to its home position after the human releases the object and starts to retreat.

Figure 11: The Sawyer robot

Category: Small robots

Lifting load: up to 4 kg

Number of degrees of freedom: 7

Self-weight: 19 kg

Arm speed: 7.2 km/h

DS35 Mid-Range Distance Sensor

A SICK DS35 mid-range photoelectric distance sensor (Fig. 12) was used.

Figure 12: DS35 Mid-Range Distance Sensor

This sensor uses HDDM (high definition distance measurement) technology to ensure maximum reliability and accuracy for distance measurement detection.

Target: Natural objects

Resolution: 0.1 mm

Accuracy: ±10 mm

In this project, this sensor was used to detect when the giver starts reaching.

Contrinex LTK-1180-103 Photoelectric Sensor

A Contrinex standard photoelectric sensor was placed on the robot arm and used to detect the object distance from the robot gripper. The Contrinex through-beam (Fig. 13) photoelectric sensor utilizes infrared, visible and laser light sources to detect targets, reliably and repeatably, at extended distances (Pcr & Kit, 2012).

Figure 13: Contrinex LTK-1180-103 Photoelectric Sensor

- Setting range: 40-600mm
- Time delay before availability: 60msec

Arduino Nano

Arduino is a single-board microcontroller with open-source hardware (Zlatanov, 2016), enabling to connect inputs and outputs (Fig. 14) (Wong et al., 2019). For this purpose, we used the Arduino programming language (based on Wiring), and the Arduino Software (IDE), based on Processing. The microcontroller can be programmed using C and C++ programming languages. In addition to using traditional compiler toolchains, the Arduino project provides an integrated development environment (IDE) based on the Processing language project.

Figure 14: Arduino Nano

A breadboard was used to connect the distance sensor and the photoelectric sensor to the Arduino. Additionally, we connected a battery (to provide a power source), resistors, and the ground to the breadboard, as depicted in Fig. 15 and Fig. 16. The Arduino, which is connected to the main computer, continuously receives signals from the sensors, and transfers them to the sawyer robot via analog input (A1) and digital input (D1)

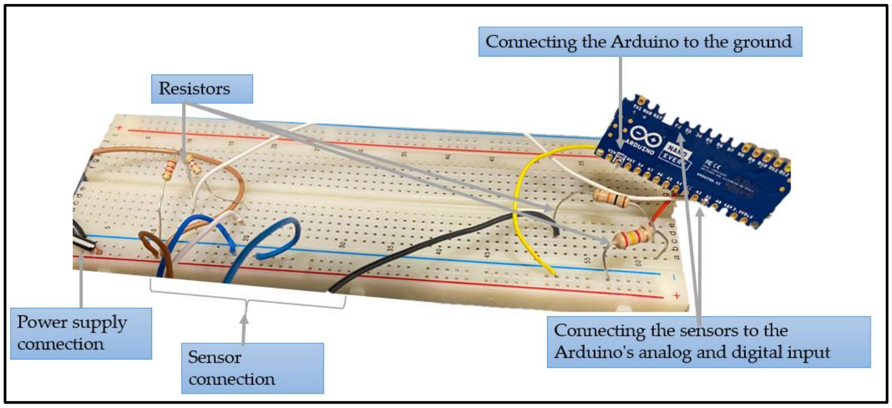


Figure 15: Circuit implementation

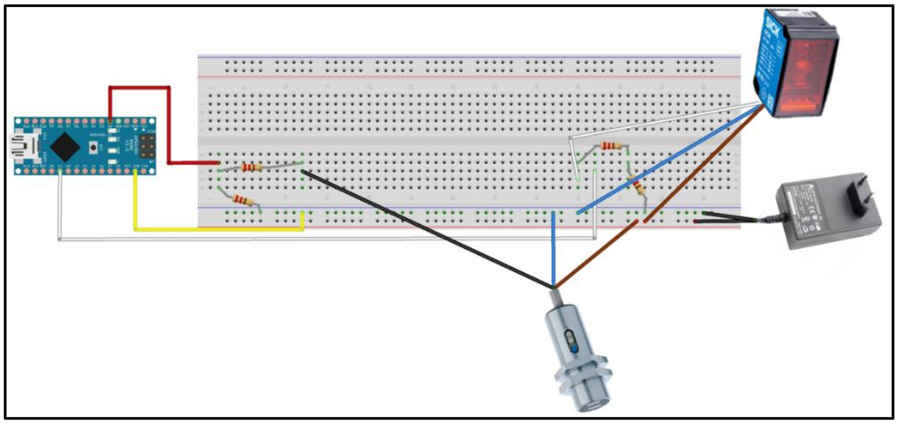


Figure 16: Schematic hardware connection

Software

This section reviews the developed system (Fig. 17) and explains the code, which consists of three Python classes and one class in C++. Full code is detailed in **שגיאה! מקור ההפניה לא נמצא.**.

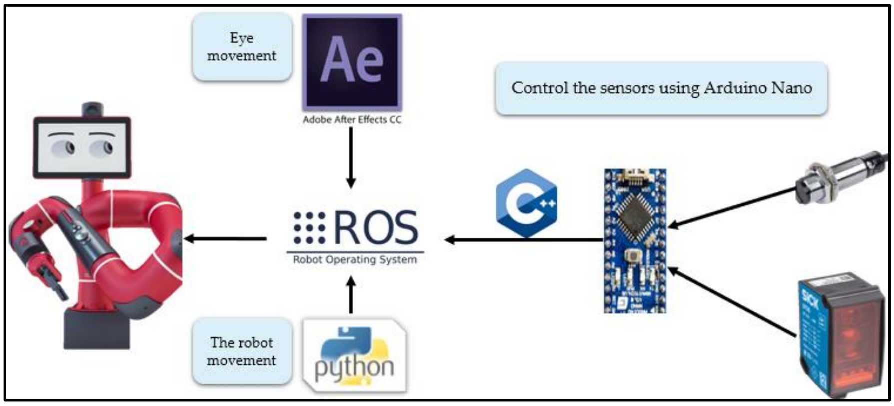


Figure 17: The Developed System

ROS

In this project, ROS was used to control the robot movement, to get messages from the Arduino indicating the sensors' measurements, and to control the robot screen.

Robot movement- python

The robot class, sends motion commands to the robot. In order to assess the beginning of the handover, the robot gets signals from the distance sensor, indicating on the position of the participant's arm. The robot arm is autonomous and programmed to reach a predefined position once the handover begins. Using a photoelectric sensor to assess the distance from the object, the robot grasps the object when the object gets close enough. Finally, the robot retreats to its home position after the human releases the object and starts to retreat.

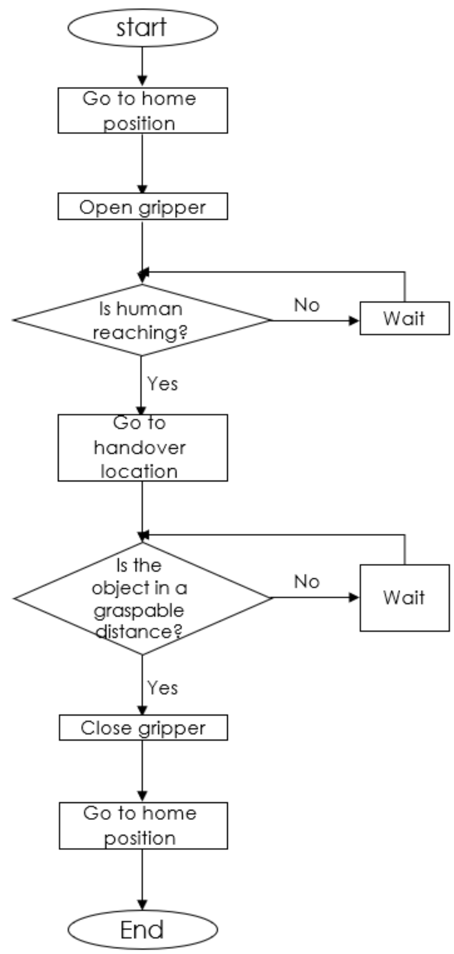


Figure 18: Flow chart of the robot movement

Eye gaze

The robot's eye gaze was created in three steps: first, creating animations for eye movements. Second, programming a Python class that controls and projects the robot's eye gaze on the robot's screen. Finally, programming a Python class that was responsible for keeping the robot's head straight, instead of moving correspondingly with the robot's arm axis.

1. **Eye movement animation- Adobe After Effects**: The eye movement animations were created in Adobe After Effects. Adobe After Effects is a digital visual-effects, motion graphics, and compositing application developed by Adobe Systems and used in the post-production process of film making, video games, and television production. In this project, 3D eye movement animations were created, which simulated the three most common gaze (Hand-Face, Face-Hand-Face, Hand gaze) that were identified in human-human handovers analysis. In order to discover human's preferable gaze pattern, these animations were projected during our experiments on the Sawyer robot's built-in screen.

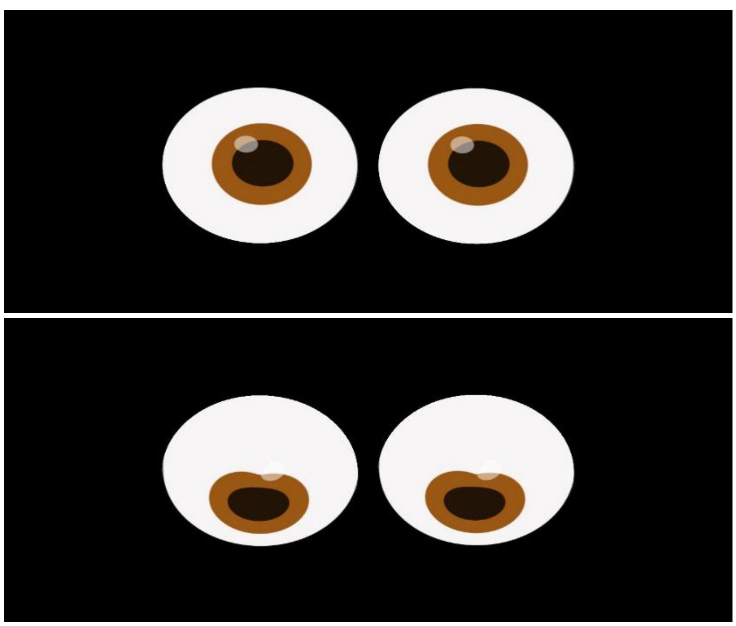


Figure 19: Pictures from the 3D eye movement animation. The top figure portrays the robot's eyes as they look toward the participant's eyes, and the bottom figure portrays the robot's eyes as they look toward the participant's hand.

2. **Eye gaze- Python class**: This class controls and projects the robot's eye gaze on the robot's screen. The three gaze patterns that were implemented are: Hand-Face, Face-Hand-Face, Hand gaze (which were identified in human-human handovers analysis). In this code, the eye gaze pattern for the session was chosen. Each eye gaze pattern has a defined function that runs a different eye movement animation depending on the handover phase.

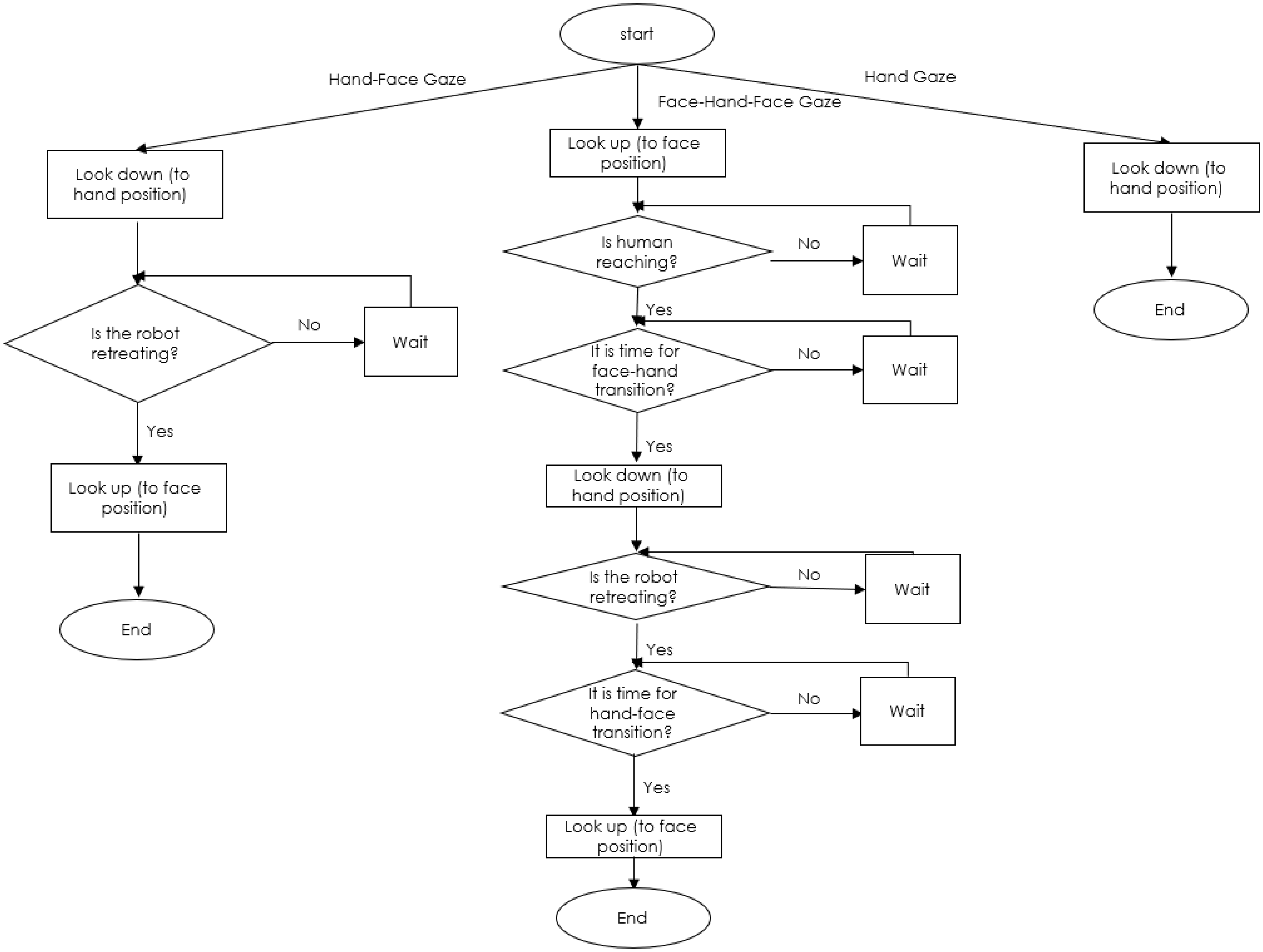


Figure 20: Flow chart of the robot's eye gaze

3. **The robot's head- Python class**: For the eye gazes to look more natural and human, the robot's head needed to stay straight. By default, when the robot moves its hand, the head rotates along with the base joint. This Python class is responsible for keeping the head straight toward the subject's face by countering the base rotation.

Arduino- C++

In this class, the signals are regularly received from the distance sensor and the photoelectric sensor. After retrieving the signals information from the sensors, it sends it to ROS via the ROS-Arduino interface.

The algorithm consist of the following steps:

1. Power ON the system which includes the microcontroller and sensors
2. Initialise the system
3. Read data from the sensors and send it to ROS via the ROS-Arduino interface

Appendix B- Online survey – people's perception of objects' fragility

To represent objects of different fragility a plastic bottle and a glass bottle were used. In order to examine people's perception about the fragility of these objects, we conducted an online survey. This survey was conducted post experiment based on reviewers' feedback.

A total of 24 participants responded to the survey. The participants were undergraduate students from the Department of Industrial Engineering and Management at Ben-Gurion University, similar to the students who participated in our video and in-person experiments. The participants were told that this study deals with object handovers between a human and a robot.

The survey included 10 pictures of objects, made from different materials (Fig. 21). The plastic bottle and the glass bottle used in our experiment were among these objects. Each picture was followed by a yes or no question: "Do you perceive this object to be fragile?".

Results revealed that all of the 24 participants perceived the plastic bottle to be non-fragile. 23 out of 24 participants perceived the glass bottle to be fragile. Additionally, when asked the same question for three other different plastic and glass bottles, 24 participants denoted the plastic bottles as non-fragile and 23 denoted the glass bottles as fragile. This supports our decision to choose plastic and glass bottles to represents objects of different fragility.

Figure 21: 10 pictures of objects, made from different materials presented in the online survey.

4.1 Appendix C- Open-ended responses

Video study of human-to-robot handovers

10 out of 72 participants gave at least one additional comment. Eight participants made Hand-Face gaze vs. Face-Hand-Face gaze comparisons. Two participants mentioned that they could not distinguish between them, saying, *"I did not see a significant change between the two videos"*, *"They looked the same to me*". Two participants preferred Face-Hand-Face gaze over Hand-Face gaze owing to the extended robot's eye contact, saying, *"As much eye contact as possible"*, *"I preferred handover 2 (Face-Hand-Face gaze) because the robot looked more at the human"*. Two participants thought that Face-Hand-Face gaze seemed more natural and human-like, and mentioned: *"Handover 1 (Face-Hand-Face gaze) felt more human-like"*, *"I preferred the second handover (Face-Hand-Face gaze) because it was more natural for the robot to look down as he extends his arm"*. Nevertheless, two participants said that they found advantages and disadvantages in both of the gaze patterns, and said: *"It is easier when the robot looks at the object, so the giver could know when it is required to hand the object over. Yet, not looking in the eyes may be considered rude"*, *"In handover 1 (Hand-Face gaze) you could tell that the robot was ready to receive the object. However, handover 2 (Face-Hand-Face gaze) felt more humanized because the robot looked at the giver's eyes right until the transfer was made"*.

While, two out of six participants, who commented on comparing Hand-Face gaze vs. Hand gaze, mentioned that they could not distinguish between them, saying, *"There is no difference"*, *"The 2 handovers looked the same"*. The other four participants said that they preferred Hand-Face gaze, saying, *"In my opinion, the change in eye movement creates a better human-robot interaction"*, *"In the second handover (Hand-Face gaze) the eye movement, gave a good indication for the communication"*, *"It is easier to understand the robot "willingness" to receive the box when the robot's eyes move as its arm progresses"*, *"It's nice that the robot looks straight at you after delivering an object"*

Six participants made Face-Hand-Face gaze vs. Hand gaze comparisons. They said that they preferred Face-Hand-Face gaze over Hand gaze because they preferred much eye contact as possible and they thought that the Face-Hand-Face gaze was clearer, saying for instance, *"At handover 2 (Face-Hand-Face gaze), the robot looked at the object precisely when it wanted to* *take it, so it was perceived more understandable"*, *"In my opinion video 2 (Face-Hand-Face gaze) best simulated human-like behavior out of all the videos I have seen so far"*

In-person study of human-to-robot handovers

14 out of 72 participants gave at least one additional comment. Seven participants made Hand-Face gaze vs. Face-Hand-Face gaze comparisons. Four participants mentioned that they could not distinguish between them, saying, *"Felt quite the same*", "*I didn't notice a difference*". Two participants stated that they preferred Face-Hand-Face over Hand-Face gaze because they preferred longer eye communication, saying, *"In Handover number 2 (Face-Hand-Face gaze) the robot looked at me for the longest amount of time, and it was the best handover so far*", *"I preferred handover 1 (Face-Hand-Face gaze) because the robot stared at me before and after the handover, and I felt accompanied by it during the entire handover*". Nevertheless, one participant argued that in his opinion Face-Hand-Face gaze pattern didn't feel natural, and used the following words, *"handover number 2 (Face-Hand-Face gaze) did not feel natural"*

While three out of seven participants, who commented on comparing Hand-Face gaze vs. Hand gaze, mentioned that they could not distinguish between them, saying, *"They looked the same to me*", *"Indifference between first and second handover*". Four of them said that they preferred Hand-Face gaze, and mentioned: *"In the first handover (Hand-Face gaze) the robot looked straight at me after the handover and seemed to be more friendly*", *"In the second handover (Hand-Face gaze) the robot looked directly at me, and it felt more human-like*", *"In the first handover (Hand-Face gaze), the robot's eye movement was fully accompanied by the handover movement, and therefore it seemed more natural*"

Seven out of eight participants, who commented on comparing Face-Hand-Face gaze vs. Hand gaze gazes, said that they preferred Face-Hand-Face gaze over Hand gaze because they preferred much eye contact, and some of them described that Face-Hand-Face was more natural, saying, "*In the first handover (Hand gaze), the robot focused only on the object, and in the second handover (Face-Hand-Face gaze) it focused on me too, so it felt more natural"*, *"I preferred the second handover (Face-Hand-Face gaze) mainly because the robot looked me in the eyes at the beginning and the end"*. However, one participant said he felt that both handovers seemed to be unfriendly, and used the following words: *"In both handovers the robot looked down, unfriendly"*.

Appendix D- Agent configuration

Component/ Variable	Description	Default value(s)	
		Sim-to-Real	Real-to-Real
EE_POINTS	Two point offsets. To ascertain that the end-effector attain the correct orientation and not merely reaches the correct position, the GPS backend requires at least two point offset. At the end of the local policy training subtraction of world-space end-effector position from the positions of these points is conducted.	$[\begin{pmatrix} 0.22 \\ -0.025 \\ 0.55 \end{pmatrix}, \begin{pmatrix} 0.22 \\ -0.025 \\ -0.55 \end{pmatrix}]$	$[\begin{pmatrix} -0.07 \\ 0.06 \\ 0.013 \end{pmatrix}, \begin{pmatrix} 0.22 \\ -0.025 \\ -0.55 \end{pmatrix}]$
Panda _Gains	A vector of scalar gains, one for each torque/joint of the robot.	$(0.1, 0.01, 0.1, 0.01, 0.01, 0.01, 0.01)$	$(0.1, 0.1, 0.1, 0.1, 0.001, 0.001, 0.001)$
Agent	Top-level configuration and details of the agent that was used.		
type	Name of the agent.	AgentROSControlArm	AgentMuJoCo/ AgentROSControlArm
dt	Step size [s].	0.05	
T	The trajectory length [steps].	400	100
state_include	A list of the internal variables that represent the system state.	[Joint angles, joint velocities, end-effector points, end-effector point velocities]	
algorithm	The details of the policy-improvement algorithm to be used by GPS.		
type		AlgorithmBADMM	
iterations	Number of full iterations of optimization.	11	
init_traj_distr	Set-up for the differential dynamic programming initialization of the linear quadratic regulator.		
init_gains	The initial joint gains.	$\frac{1}{\text{Panda\_Gains}}$	
init_var	The variance of the initial trajectory distribution. The initial variance affects the state-space explored by the robot in	1000	0.5

	the initial training step. A higher initial variance results in larger explored state-space but with higher control inputs, which might exceed the joint limits in some cases causing the robot to halt. A lower initial variance results in smaller control inputs but also a smaller explored state-space causing the robot to not learn the task.		
stiffness	Initial stiffness of the joints. Important to get the joints turning in the initial distributions before the true dynamics begin to be discovered.	1	1
stiffness_vel	Initial velocity stiffness.	0.5	0. 5
cost	The weighted sum of the cost terms defined below.		
weights	The external weights of each cost term.	$[1, 1]$	
dynamics	Specifies the type of dynamic model prior used to optimize the trajectories.	Maximum 20—cluster Gaussian mixture model.	
fk_cost			
wp	The internal vector (length T) of weights per trajectory step.	$[1, 1, 1]$	
l1	The internal weight of the L1 norm sub-term	2	
l2	The internal weight of the L2 norm sub-term	2	
alpha		$1e-5$	
config	Connects above options, the optimization algorithm and the agent		
num_sapmles	The number of trajectory samples used on each iteration to improve the dynamic model.	5	

Appendix E- PC information

Power supply	HEC Cougar VTE600 // XPG Probe - Dual GPU compatible **power supply 600W** Active PFC, 12cm silent fans, **80 PLUS® Bronze certified**
Processor	**Intel® Core™ i9-10900, LGA1200 Package 2.8GHz** 10 Core with Hyper-Threading, 14nm, 65W, 20MB Cache L3 Intel® Max Turbo Boost Technology up to 5.1GHz, Enhanced Intel SpeedStep® SSE4.2, AVX2.0, TSX-NI, Secure Key. Intel Virtualization Technology VT-x/d Dual Channels of DDR4 2933MHz memory controller Integrated Intel® UHD Graphics 630 up to 1.2GHz, 3 displays up to 4096x2304 Antec C40 high effectivity, silent CPU cooler
Mother board	**ASUS TUF Gaming B460M-Plus** LGA1200 Socket, Intel® 10th generation Intel® Core™ processor ready. **B460 Chipset**. Type-A 2(+4)*USB2.0 & 4(+2)*USB3.2 G1. microATX. **4*DIMM** 240-pin Dual Channel DDR4 2133-2933MHz **up to 128GB** 2*PCI Exp. x16 v3.0 (1*x16 & 1*x4), 1*PCI Exp. x1 v3.0. Aura RGB strip headers B460 PCH 6*SATA-3, Matrix RAID (0,1,10,5) Smart Response, Optane™ memory 2 port M2 SATA3.0 or PCIe v3.0 up to 32Gb/s M-key up to 2280 Integrated: Realtek ALC S1200A High Definition Audio 7.1 codec. **Intel® I219-V 1.0Gbps** RJ45 Ethernet controller. Serial port header Video out ports: **DVI-D, HDMI 1.4b & DisplayPort 1.4 (4096x2160)**
Memory	**Kingston Hyper-X 32GB DDR-4 2933**(3200)**MHz Dual Channel (2x16GB)**
SSD HD	**WD Black SN750 Series SSD Drive 500GB** **PCIe NVMe 3.0 x4 M2** 2280, Read//Write up to **3430//2600MB/s.** AES 256-bit Encryption. 5 year warranty or 300TBW
Ethernet network adapter	**Intel® Ethernet Converded network adapter X550-T2** **Dual port**, RJ-45 **10 Gbps port**. **PCI Express x4 slot**. Low Profile and Full Height Virtual Machine Device Queues (VMDq) support

Appendix F- Joint space limits of the Panda robot

Name	**Joint 1**	**Joint 2**	**Joint 3**	**Joint 4**	**Joint 5**	**Joint 6**	**Joint 7**	**Unit**
q_{max}	2.8973	1.7628	2.8973	-0.0698	2.8973	3.7525	2.8973	rad
q_{min}	-2.8973	-1.7628	-2.8973	-3.0718	-2.8973	-0.0175	-2.8973	rad
$\dot{q}_{max}$	2.1750	2.1750	2.1750	2.1750	2.6100	2.6100	2.6100	$\frac{rad}{s}$
$\ddot{q}_{max}$	15	7.5	10	12.5	15	20	20	$\frac{rad}{s^2}$
$\dddot{q}_{max}$	7500	3750	5000	6250	7500	10000	10000	$\frac{rad}{s^3}$
$\tau_{j_{max}}$	87	87	87	87	12	12	12	Nm
$\dot{\tau}_{J_{max}}$	1000	1000	1000	1000	1000	1000	1000	$\frac{Nm}{s}$

Appendix F- Implementation and Evaluation of Guided Policy Search for Robot Reaching Towards Moving Targets

Implementation and Evaluation of Guided Policy Search for Robot Reaching Towards Moving Targets

Alap Kshirsagar[1†], Tair Faibish[2†], Guy Hoffman[1] and Armin Biess[2]

Abstract— **We investigate the performance of a model-based reinforcement learning (RL) method, Guided Policy Search (GPS), for generating reaching motions of a collaborative robot arm. We conduct this evaluation in the context of a robot controller for the reach phase of a human-robot handover. In a previous work, we evaluated GPS for the same task but only in a simulation environment. This paper provides new insights on the limitations of GPS on a physical robot platform. First, we find that a policy learnt in simulation does not transfer to the physical robot owing to differences in model parameters. Second, the robot's workspace needs to be severely reduced to successfully train with GPS owing to the joint-space limitations of the physical robot. Third, a policy trained with moving targets results in large worst-case errors even in regions spatially close to the training target locations. Our findings could motivate further research towards utilizing machine learning algorithms for physical human-robot collaboration.**

Index Terms— **Physical Human-Robot Interaction, Reinforcement Learning, Manipulation Planning**

I. Introduction

In this work, we evaluate the potential of Guided Policy Search (GPS) for reactive robot reaching motion towards a moving target. Several tasks in domestic and industrial environments require robots to reach towards moving targets. Examples include human-robot object handovers, manipulation of objects on conveyor belts and catching flying objects. We focus on the scenario of human-to-robot object handover in which the robot needs to reach towards a moving target i.e. the human's hand. Researchers have suggested several closed-loop controllers for the reach phase of human-to-robot object handovers [1]–[17]. However, these methods require prior knowledge of the robot's dynamics and/or human kinematics.

GPS-Bregman Alternating Direction Method of Multipliers (BADMM) [18] is a reinforcement learning algorithm that does not require prior knowledge of the robot/environment dynamics. It uses an iterative adaptation of local controllers, local dynamics model, and a global policy over repeated trials. GPS was initially proposed by Levine et. al [19]–[21], and since then researchers have proposed several variations of the GPS algorithm [22]. GPS algorithms have been successfully demonstrated for autonomous manipulation [18], [23]–[25], and locomotion tasks [19], [20], [24], [26]. However, to the best of our knowledge, GPS has not been tested on physical robots for tasks that require the robot to reach towards unpredictable moving targets, such as human-to-robot object handovers. We seek to address this gap by evaluating GPS for the reach-to-handover motion generation of a collaborative robot.

In our previous work [27], we used GPS-BADMM to train a robot arm to perform reach-to-handover motions in a simulation environment. We found that the policy learnt with GPS does not perform well for test locations that are spatially distant from training locations. This issue can be mitigated by adding more local controllers trained over target locations in those high error regions. Further, a policy trained with static targets generates high joint torques when tested with moving targets. More efficient reaching trajectories can be obtained by training on moving targets, although it results in higher worst-case errors. Despite providing important insights, our prior work is limited in it's application to a real-world environment. The goal of our present work is twofold: first to replicate our previous findings on a physical robot arm, and second, to provide new insights on the challenges associated with the real world implementation of GPS.

II. Policy Search Formulation of Handovers

We formalize the reach phase of a handover task as a reinforcement learning problem by specifying the state/action space, as well as a cost/reward function over the system states and control inputs.

A. State/Action Space

In our previous work [27], we explored the sensitivity of GPS to different state representations. We investigated three different system state representations: `FULL`, `RELATIVE`, `REDUCED`. We found that a policy trained with `RELATIVE` state representation had a better overall performance. Thus in this work, we use the `RELATIVE` state representation which consists of the robot joint angles θ_r, the robot joint velocities $\dot{\theta}_r$, the positions and velocities of the robot-end effector in the world frame attached to the base of the robot ($\mathbf{p}_r$,$\dot{\mathbf{p}}_r$), and the positions and velocities of the human hand in the robot end-effector frame ($\mathbf{p}_h^r$,$\dot{\mathbf{p}}_h^r$).

$$\mathbf{x}_t = [\theta_r, \dot{\theta}_r, \mathbf{p}_r, \mathbf{p}_h^r, \dot{\mathbf{p}}_r, \dot{\mathbf{p}}_h^r]_t. \quad (1)$$

Similar to our previous work, the robot's control input $\mathbf{u}_t$ consists of the robot joint torques τ and the force applied by the gripper's actuator f_g, constrained by $\mathbf{u_{min}} \leq \mathbf{u}_t \leq \mathbf{u_{max}}$:

† Alap Kshirsagar and Tair Faibish contributed equally to this work.

[1]Alap Kshirsagar (Corresponding Author, `ak2458@cornell.edu`) and Prof. Guy Hoffman (`hoffman@cornell.edu`) are with the Sibley School of Mechanical and Aerospace Engineering, Cornell University, USA.

[2] Tair Faibish (`tairse@bgu.ac.il`) and Dr. Armin Biess (Corresponding Author, `armin.biess@gmail.com`) are with the Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Be'er Sheva, Israel.

$$\mathbf{u}_t = [\tau, f_g]_t \tag{2}$$

B. Cost Function

We use a cumulative error cost function to describe the reach-to-handover motion of a robot,

$$c_{reach} = \sum_{t=0}^{T} \left[||\mathbf{p}_r - \mathbf{p}_h||^2 + \ln(||\mathbf{p}_r - \mathbf{p}_h||^2 + \alpha_{reach}) \right]_t , \tag{3}$$

where $\mathbf{p}_r$ is the position of the robot and $\mathbf{p}_h$ is the position of the human hand, and T is the duration of each trial. This cost function penalizes the robot for spatial distance away from the human's hand, and it encourages precise placement owing to its concave shape, as described in [25]. To be consistent with prior works [25], [27], we set $\alpha_{reach} = 1e{-}5$ in the evaluations described in the next section.

C. GPS-BADMM Algorithm

As depicted in Fig. 1, the GPS algorithm alternates between generating optimal trajectories for each initial condition (local iLQR controllers) and training a global policy supervised by the local controllers. The global policy's role is to improve the local controllers, retaining them close to the global policy. The BADMM variant of the GPS algorithm does not require knowledge of the dynamics model as it utilizes the training data with locally linear models to approximates the dynamics.

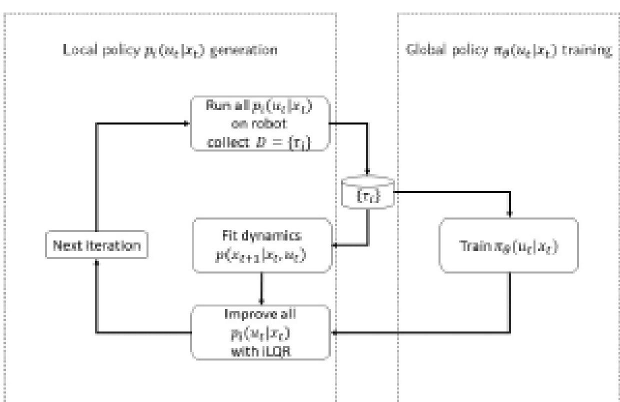


Fig. 1: Guided policy search algorithm iteratively updates the local controllers or policies and the global policy. The local policies serve as the "experts" for supervised learning of the global policy. These local policies are also updated to avoid drifting away from the global policy.

III. Implementation of GPS-BADMM on a Franka-Panda Robot

We train a collaborative robot, Franka-Panda Emika, to perform reach-to-handover motions with the GPS-BADMM algorithm. The Panda robot, shown in Fig. 3, is a 7 degrees of freedom robot arm with torque sensors at each joint, allowing adjustable stiffness/compliance and advanced torque control. We use OptiTrack motion tracking system to track the positions of the human's hand and the robot end effector. Since it is not practical to have a human trainer/tester perform exactly the same handover motion in all training/testing iterations, we use recorded human hand motions during the training process. We build on the GPS-BADMM implementation of White [28], which was done for a Kuka robot in Gazebo simulation environment. White's code itself was built on the GPS implementation of Finn [29] for a PR2 robot. We use the distributed computing capabilities of Robot Operating Systems (ROS) and run different ROS packages/nodes on different machines as shown in Fig. 2. One PC ("Opti-Track Computer") runs the OptiTrack Motive software and streams the motion tracking data on a local network via ROS topics *mocap_optitrack/franka_gripper/pose* and *mocap_optitrack/human_hand/pose*. The *mocap_optitrack* node runs on the second PC (*GPS Computer*) and converts the motion tracking data to ROS *tf* coordinate frames. These coordinate frames are converted to our `RELATIVE` state representation by the *optitrack_publisher* node and sent to the *agent_ros_control_arm* node via *mocap_optitrack_data_topic*. The *agent_ros_control_arm* node communicates with the *franka_ros* node running on the third PC (*Franka Computer* which is connected to the Franka-Panda robot.

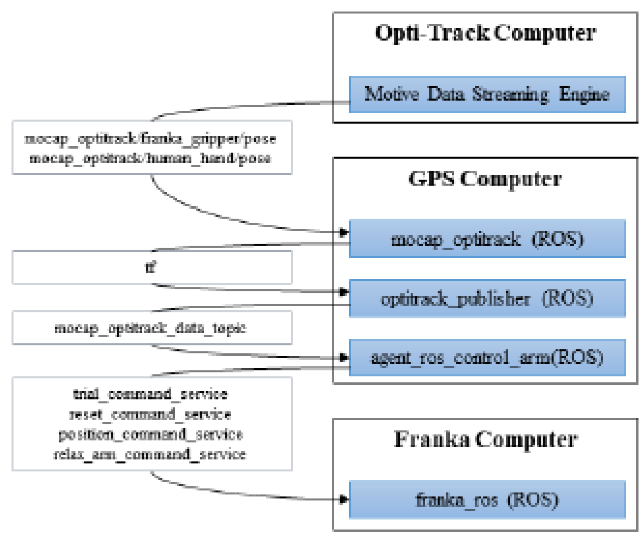


Fig. 2: Distribution of ROS nodes across different computers and connections between them using ROS topics.The nodes are shown in the colored boxes and topics/services are shown in dotted boxes.

Our modifications to the implementation of White [28] are summarized below.

1) Tune PID parameters: A PID joint position controller is used to reset the arm before beginning the GPS trial/test. The controller commands the robot to move to a predefined position, defined in terms of the joint angles. The "proportional" part of the controller applies control input proportional to the error between the current position and the target position. The "integral" part of the controller applies control input depending on the integration of error between the current position and the target position. The "derivative" part of the controller applies control input depending on the difference between the derivatives of the current position and the target position. The default parameters of this PID

TABLE I: Comparison of PID parameters of position controller for resetting the LWR and Panda robots.

Joint	LWR Values [28]				Panda Values			
	P	I	D	I_{clamp}	P	I	D	I_{clamp}
1	2400	0	18	4	6	3	3	1
2	1200	0	20	4	6	3	3	1
3	1000	0	6	4	6	3	3	1
4	700	0	4	4	6	3	3	1
5	300	0	6	2	2.5	1	1	1
6	300	0	4	2	2.5	1	1	1
7	300	0	2	2	2.5	1	1	1

TABLE II: Initial controller values of PR-2, LWR, and Panda robots.

Parameter	PR-2 [29]	LWR [28]	Panda
Joint 1 Gain	3.09	24	0.1
Joint 2 Gain	1.08	12	0.1
Joint 3 Gain	0.393	10	0.1
Joint 4 Gain	0.674	7	0.1
Joint 5 Gain	0.111	3	0.001
Joint 6 Gain	0.152	3	0.001
Joint 7 Gain	0.098	6	0.001
Initial Variance	1	30	0.5
Stiffness	0.5	60	1.0
Stiffness Velocity	0.25	0.25	0.5

controller did not work with the real Panda robot. With the default parameters, the robot's joints did not move at all, only the tip joint would barely turn. We tuned the PID parameters to work with the Panda robot. Table II shows the values used by White [28] for LWR robot in Gazebo simulation, and the values we used for a real Franka-Panda robot. We found that high values of proportional gain resulted in the robot crashing or abruptly halting as it exceeded the joint velocity limits. For more details about the PID parameters we refer the reader to [30].

2) Tune the initial local controllers: The initial local controllers used in the GPS training process are linear gaussian controllers which try to hold the robot's initial position. The initial controller gains are computed with LQR, defined by the parameters described below. It is important to initialize these parameters to ensure that the robot starts the learning process while maintaining stability. The default parameters used in Finn's code for PR-2 or Jack White's code for LWR did not work with the Panda robot. The robot did not move at all with those parameters. We obtained the initial controller parameters for the Panda robot by trial-and-error. The first parameter is *Robot Joint Gains*, which is a vector of scalar values, one for each torque/joint of the robot. These are used to guess the initial dynamics of the robot by LQR. The initial local controllers are extremely sensitive to these gains; a too high gain leads to robot exceeding the joint limits, whereas a too low gain prevents the joint from moving at all. The second parameter is the *initial variance*, which affects the state-space explored by the robot in the initial training step. A higher initial variance results in a larger explored state-space but with higher control inputs, which might exceed the joint limits in some cases causing the robot to halt. A lower initial variance results in smaller control inputs, but also a smaller explored state-space causing the robot to not learn the task.

3) Robot HW Interface: We used Franka HW interface provided by the *franka_ros* ROS package, instead of the Kuka LWR HW interface used in White's work.

4) Motion Tracking Feedback: We used the *rostopicsensor* abstraction of Finn's code to the position of the human's hand and the robot gripper obtained from OptiTrack motion tracking system to the GPS controller.

5) Torque Limits: We found that the control input i.e. joint torques, generated by the GPS controller resulted in joint position/velocity violation on the Franka-Panda robot. We had to restrict the control input to a range of $[-3\text{N.m}, 3\text{N.m}]$ in the *trialcontroller* class of Finn's code.

6) Trial Report Publisher: We encountered communication failures with the robot during the training phase because the robot did not receive the published result of a finished training iteration. To address this issue, we had to replace the real-time trial report publisher of Finn's GPS implementation with a non-real time trial report publisher.

IV. Results

A. Sim-to-Real Evaluation

In our previous work [27], we had evaluated GPS in a simulation environment *MuJoCo* (Multi-Joint dynamics with Contact) [31] as shown in Fig. 4. We had trained a Panda robot for the handovers task with a pseudo-robot arm with a mass rigidly attached to its end-effector, substituting the human operator. In the first experiment of the present work, we are interested to check the feasibility of sim-to-real transfer of the learnt policy.

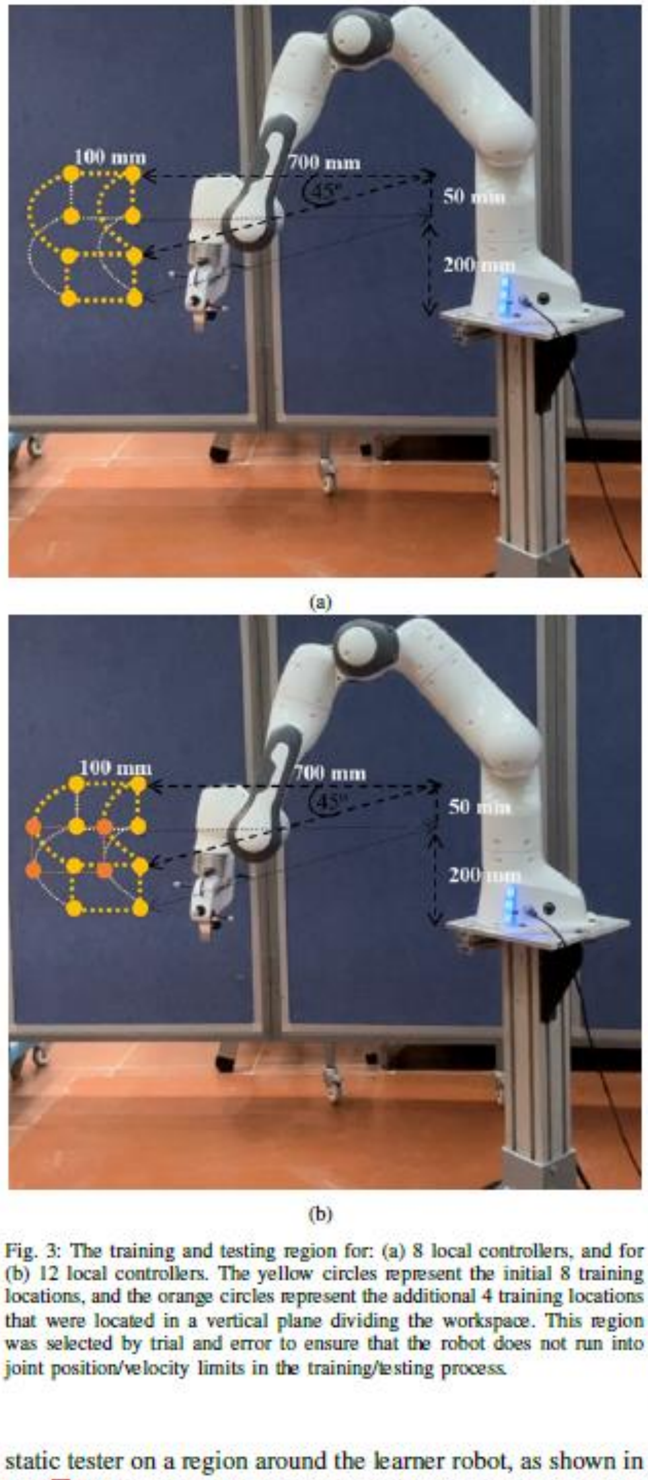
We find that the policy trained in the simulation does not work on the physical physical Franka-Panda robot. This robot has limited acceptable ranges of joint positions, velocities and torques[1]. The torques generated by the GPS policy learned in the simulation are beyond these limits, and thus the policy does not work on the real robot.

B. Real-to-Real Evaluation

We train a real Panda robot to perform reaching motions towards a human's hand over repeated trials with GPS, and test the learnt global policy for large variations in target locations and moving targets. In the remaining text, we denote the Panda robot the "learner", and the human is denoted as the "trainer" for the training phase or the "tester" for the testing phase. Since it is not practical to have a human trainer/tester perform exactly the same reaching motion in all training iterations, we use recorded human hand motions and feed them to the robot via the *rostopicsensor* interface as described in Section III .

The first research question examined in our study is the spatial generalizability of the learned global policy, i.e., how does the global policy perform for large spatial differences between training and testing locations. To answer this question, we test the learnt global policy at different locations of a

[1]Franka-Panda Specifications: https://frankaemika.github.io/docs/control_parameters.html



(b)

Fig. 3: The training and testing region for: (a) 8 local controllers, and for (b) 12 local controllers. The yellow circles represent the initial 8 training locations, and the orange circles represent the additional 4 training locations that were located in a vertical plane dividing the workspace. This region was selected by trial and error to ensure that the robot does not run into joint position/velocity limits in the training/testing process.

static tester on a region around the learner robot, as shown in Fig. 3. For each angle in 5 deg increments, we test on a grid of 3×3 targets, resulting in 90 test locations. We compare two scenarios of local controllers: one with 8 local controllers and another with 12 local controllers. The global policy is trained with these local controllers for 11 iterations. Both the learner and the trainer/tester commence their movement in each trial simultaneously. The learner's movement lasts for 5 seconds, while the trainer's/tester's movement lasts for 1 second. The test performance is measured as the mean error between the learner's gripper position and the tester's hand

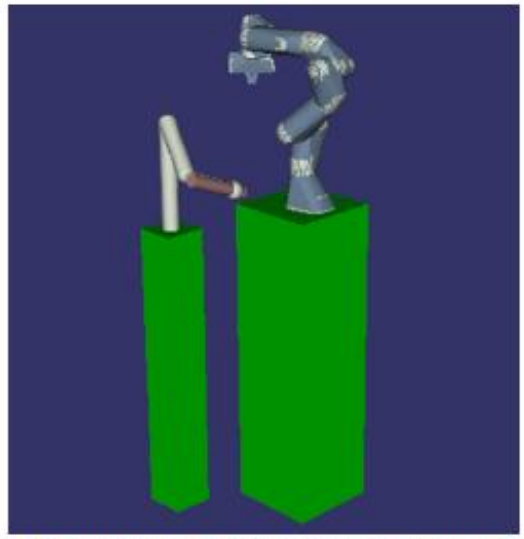


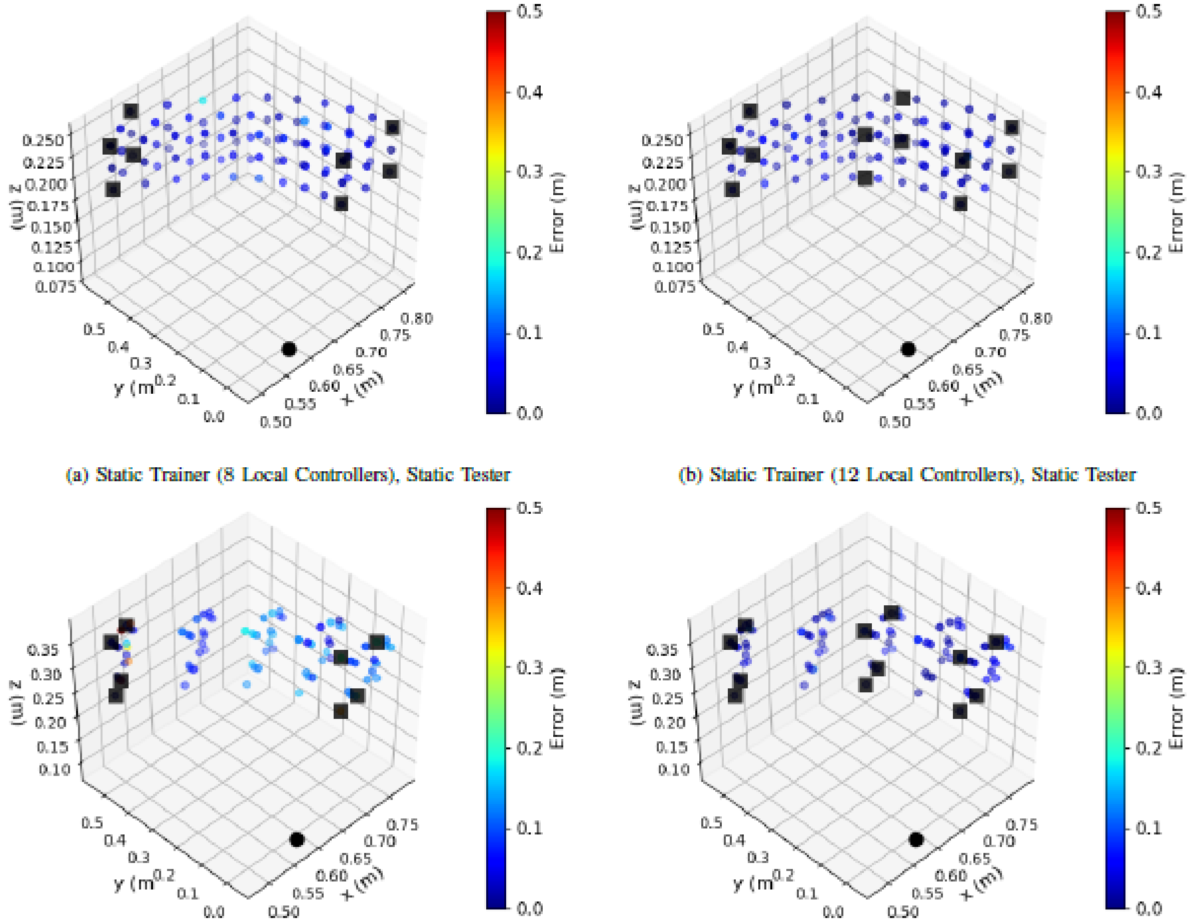
Fig. 4: MuJoCo (Multi-Joint dynamics with Contact) simulation environment for human handover tasks. A Panda robot (right) was trained in simulation on reaching movements in a human-to-robot handover task. The human operator is represented by a pseudo-robot (left).

position at the last time step of each trial.

The performance of the learned global policy is shown in Fig. 5a. The black circle represents the learner's gripper's initial position, and the black squares represent the training locations. Mean error, range, and standard deviation are presented in Fig. 6 (left). The mean testing error (41.71mm) is about twice as large as the mean training error (22.67mm). To replicate our previous findings [27] that the test error can be reduced by adding more local controllers in high error regions, we add 4 additional local controllers in a vertical plane dividing the workspace (Fig. 5b). We find that the mean and standard deviation of the testing error of the global policy, trained with 12 local controllers, is reduced to 29 ± 2mm.

Next, we investigate how GPS performs when the target is moving. First, we use the same global policy shown in Fig. 5a (static training), but instead of a static tester, we use a moving target i.e. a recorded human reaching motion. The final position of the motion is in a region similar to the one shown in Fig. 3. We find that the robot generates highly inefficient trajectories, and sometimes does not even execute these trajectories due to joint/cartesian limits violations. A possible way to address this issue, as found in our previous study [27], is to train the controller with a moving target. We train the robot with recorded human reaching motions, and test the policy on another set of recorded human reaching motions. Some samples of these reaching motions are shown in the video attachment. Fig. 5c and Fig. 5d show the performance of the global policy for various final positions of the tester's gripper, defined as in previous trials. Fig. 6 (right) shows error distributions.

For the global policy trained with a moving trainer and 8 local controllers (Fig. 5c), the mean testing error is 124.28mm. Although the test errors are high as compared to the static tester scenario, the robot stays within the joint



(a) Static Trainer (8 Local Controllers), Static Tester

(b) Static Trainer (12 Local Controllers), Static Tester

(c) Moving Trainer (8 Local Controllers), Moving Tester

(d) Moving Trainer (12 Local Controllers), Moving Tester

Fig. 5: Global policy evaluation for different types of trainers and testers. The black circle represents the learner's gripper's initial position, and the black squares represent the training locations. In the 'static' case, the trainer/tester stays in a fixed configuration. In the 'moving' case, the trainer/tester moves with a human-like trajectory (that were recorded in advanced) and reaches the locations given by colored dots. Thus, each point corresponds to the final position of the tester's gripper in a trial. Error between the learner's gripper position and the tester's gripper position is calculated over the last time step of each trial.

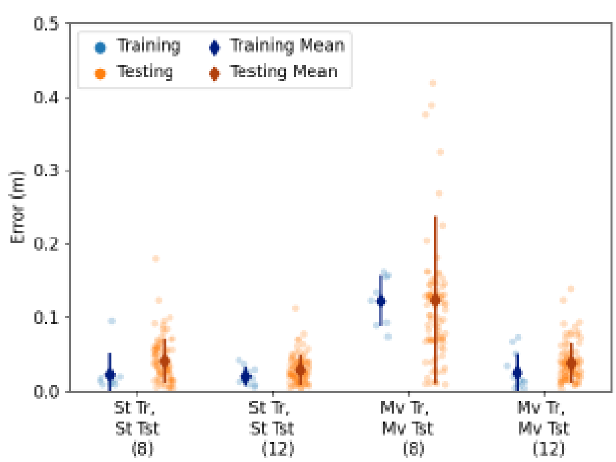


Fig. 6: Distributions of training and testing performance for each target scenario. Each point is the mean error between the learner's gripper position and the tester's hand position over the last time step of a trial. Error bars show one standard deviation around the mean of each distribution

and cartesian limits. Moreover, the variance over target location is high, and the worst-case error is 791.11mm, 442% higher than the maximum error for static tester condition (179mm). Surprisingly, the tester's motion for the worst-case error is close to one of the training motions. This can be attributed to the highly non-linear nature of the global policy. Interestingly, GPS does not converge to a low training error for the moving trainer scenario, 123.23mm which is 544% higher than the training error for a static trainer 22.67mm. Training the global policy with a moving trainer and 12 local controllers (Fig. 5d), reduces the mean testing error to 37.93mm, and the worst-case error also improves to 138.71 mm. Fig. 6 shows the distributions of training and testing performance for each target scenario.

V. Discussion and Conclusion

Our work evaluates the feasibility of GPS as a learning method for generating robot reaching motions in a real-world environment for large variations in target locations and for moving targets. We find some open challenges both in transferring the learning from simulation to the physical robot and directly training the physical robot.

We find that the robot runs into joint position/velocity/torque limit violations, when the policy is learnt in a simulation environment and then deployed to the real robot. This can be attributed to the differences in the simulation model's dynamics and the real robot's dynamics. However, tuning the simulation model dynamics parameters to match the real robot's parameters is not a feasible solution owing to the large number of possibilities. GPS has been shown to be robust to changes in the robot's dynamics within a certain range [27], but our findings suggest that GPS is not robust enough to directly transfer learning from simulation to the real robot.

When GPS is used to directly train the physical robot, we again find that the robot runs into joint position/velocity/torque limit violations during the training phase. We have to reduce the robot's workspace, by trial-and-error, to avoid these violations. In this reduced workspace (Fig. 3), we find that when the robot is trained to reach only static target locations, the global policy performance can be slightly improved by adding local controllers in regions with highest test errors (in the middle of the working plane) (Fig. 5a compared to Fig. 5b). Previously, similar results were found in a simulation environment [27].

When evaluating the global policy trained with static targets on a moving test target, the robot generates highly inefficient trajectories, sometimes resulting in halts due to joint limit violations. To overcome this issue, we train the global policy with moving targets. Nevertheless, this solution is not free of drawbacks. It successfully reduces the mean error and results in more efficient and low-torque trajectories, but also results in a high-variance (unreliable) global policy with significantly larger worst-case errors. This issue can be addressed by adding local controllers to the training phase, improving the global policy performance (Fig. 5d). These finding also support previous findings in a simulation environment [27].

Our study contributes to the understanding of the challenges and applicability of GPS in a real-world context. We use a physical Franka-Panda Emika robot in our evaluation. The low-level controller of this robot has inbuilt safety stops that halt the robot whenever it exceeds any joint or cartesian position/velocity/torque limits. It is not possible to override these limits, since this is an important feature for human-safe operation of the robot. This feature also prevents any damage to the robot's hardware. Thus there is a need to develop GPS algorithms that will train local controllers and global policy while obeying these limits.

References

[1] V. Micelli, K. Strabala, and S. Srinivasa, "Perception and control challenges for effective human-robot handoffs," in *Robotics: Science and Systems (RSS) Workshop on RGB-D Cameras*, 2011.

[2] M. Bdiwi, A. Kolker, J. Suchý, and A. Winkler, "Automated assistance robot system for transferring model-free objects from/to human hand using vision/force control," in *International Conference on Social Robotics*, 2013, pp. 40–53.

[3] M. Pan, E. Croft, and G. Niemeyer, "Exploration of geometry and forces occurring within human-to-robot handovers," in *IEEE Haptics Symposium*, 2018, pp. 327–333.

[4] W. He, D. Sidobre, and R. Zhao, "A Reactive Trajectory Controller for Object Manipulation in Human Robot Interaction," in *International Conference on Informatics in Control, Automation and Robotics*, 2013.

[5] A. Fishman, C. Paxton, W. Yang, N. Ratliff, and D. Fox, "Trajectory optimization for coordinated human-robot collaboration," *arXiv preprint arXiv:1910.04339*, 2019.

[6] M. Pan, E. Knoop, M. Bächer, and G. Niemeyer, "Fast handovers with a robot character: Small sensorimotor delays improve perceived qualities," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 6735–6741.

[7] L. Scimmi, M. Melchiorre, S. Mauro, and S. Pastorelli, "Experimental real-time setup for vision driven hand-over with a collaborative robot," in *International Conference on Control, Automation and Diagnosis (ICCAD)*, 2019, pp. 1–5.

[8] A. Kshirsagar, H. Kress-Gazit, and G. Hoffman, "Specifying and synthesizing human-robot handovers," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 5930–5936.

[9] M. Prada, A. Remazeilles, A. Koene, and S. Endo, "Implementation and experimental validation of Dynamic Movement Primitives for object handover," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014, pp. 2146–2153.

[10] G. Maeda, M. Ewerton, R. Lioutikov, H. Amor, J. Peters, and G. Neumann, "Learning interaction for collaborative tasks with probabilistic movement primitives," in *IEEE-RAS International Conference on Humanoid Robots*, 2014, pp. 527–534.

[11] D. Vogt, S. Stepputtis, B. Jung, and H. Amor, "One-shot learning of human–robot handovers with triadic interaction meshes," *Autonomous Robots*, vol. 42, no. 5, pp. 1053–1065, 2018.

[12] A. Kupcsik, D. Hsu, and W. Lee, "Learning dynamic robot-to-human object handover from human feedback," *Robotics Research*, vol. 1, pp. 161–176, 2017.

[13] F. Riccio, R. Capobianco, and D. Nardi, "Learning human-robot handovers through π-STAM: Policy improvement with spatio-temporal affordance maps," in *IEEE-RAS International Conference on Humanoid Robots*, 2016, pp. 857–863.

[14] J. Medina, F. Duvallet, M. Karnam, and A. Billard, "A human-inspired controller for fluid human-robot handovers," in *IEEE-RAS International Conference on Humanoid Robots*, 2016, pp. 324–331.

[15] K. Yamane, M. Revfi, and T. Asfour, "Synthesizing object receiving motions of humanoid robots with human motion database," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013, pp. 1629–1636.

[16] X. Zhao, S. Chumkamon, S. Duan, J. Rojas, and J. Pan, "Collaborative human-robot motion generation using LSTM-RNN," in *IEEE-RAS International Conference on Humanoid Robots*, 2018, pp. 1–9.

[17] W. Yang, C. Paxton, M. Cakmak, and D. Fox, "Human grasp classification for reactive human-to-robot handovers," *arXiv preprint arXiv:2003.06000*, 2020.

[18] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1334–1373, 2016.

[19] S. Levine and V. Koltun, "Guided policy search," in *International Conference on Machine Learning*, 2013, pp. 1–9.

[20] ——, "Variational policy search via trajectory optimization," in *Advances in neural information processing systems*, 2013, pp. 207–215.

[21] ——, "Learning complex neural network policies with trajectory optimization," in *International Conference on Machine Learning*, 2014, p. II–829–II–837.

[22] J. Du, J. Fu, and C. Li, "Guided policy search methods: A review," *Journal of Physics: Conference Series*, vol. 1748, no. 2, p. 022039, jan 2021. [Online]. Available: https://doi.org/10.1088/1742-6596/1748/2/022039

[23] Y. Chebotar, M. Kalakrishnan, A. Yahya, A. Li, S. Schaal, and S. Levine, "Path integral guided policy search," in *IEEE international conference on robotics and automation (ICRA)*, 2017, pp. 3381–3388.

[24] S. Levine and P. Abbeel, "Learning neural network policies with guided policy search under unknown dynamics," in *Advances in Neural Information Processing Systems*, 2014, pp. 1071–1079.

[25] S. Levine, N. Wagener, and P. Abbeel, "Learning contact-rich manipulation skills with guided policy search," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 26–30.

[26] T. Zhang, G. Kahn, S. Levine, and P. Abbeel, "Learning deep control policies for autonomous aerial vehicles with mpc-guided policy search," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 528–535.

[27] A. Kshirsagar, G. Hoffman, and A. Biess, "Evaluating guided policy search for human-robot handovers," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3933–3940, 2021.

[28] J. White, "Guided policy search for a lightweight industrial robot arm," Master's thesis, Luleå University of Technology and Aalto University/Erasmus+, 2018.

[29] C. Finn, M. Zhang, J. Fu, W. Montgomery, X. Yu Tan, Z. McCarthy, B. Stadie, E. Scharff, and S. Levine, "Guided policy search code implementation," Software available from rll.berkeley.edu/gps (2020/06/19).

[30] T. Faibish, "Human-robot handovers: Human preferences and robot learning," Master's thesis, Department of Industrial Engineering and Management, Ben-Gurion university of the Negev, Beer Sheva, Israel, 2022.

[31] E. Todorov, T. Erez, and Y. Tassa, "MuJoCo: A physics engine for model-based control," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 5026–5033.

תקציר

תזה זו עוסקת במחקר המשימה השיתופית של מסירת אובייקטים מאדם לרובוט. מסירות הן יכולת חיונית לרובוטים שיתופיים. התמקדנו בשני נושאים חיוניים להטמעת מאפיינים דמויי אדם ברובוטים. ראשית, בחנו את השפעת התקשורת הבלתי מילולית של הרובוט על החוויה של האדם, ועל השטף של מסירות מאדם לרובוט. שנית, אנו מפתחים ומעריכים בקר רובוט המבוסס על למידה באמצעות חיזוקים לביצוע מסירה רציפה וטבעית יותר.

החלק הראשון במחקר חוקר את העדפת האדם בנוגע למבט הרובוט במהלך מסירות אדם-רובוט. אומנם קיימת ספרות בנושא מבט הרובוט בהעברות בין רובוט-לאדם, אך קיים מחסור בספרות בדבר מבט הרובוט בהעברות אדם-לרובוט. מחקר קודם שחקר את התנהגות מבט הרובוט בהעברות אדם-לרובוט בחן את דפוסי המבט של מקבל האובייקט בשלב ה"הגעה" בלבד, והשתמש באובייקט מסוים אחד בתנוחה אחת בלבד. בהתבסס על עבודה זו, במחקר הנוכחי חקרנו דפוסי מבט עבור כל שלושת השלבים של תהליך המסירה: הגעה, העברה ונסיגה, הן בווידאו והן במחקרים פרונטליים. כחלק מהמחקר נבדקו גם האם גודלו ושבריריותו של האובייקט או תנוחת נותן האובייקט משפיעים על העדפת האדם למבט הרובוט במונחים של החיבה הנתפסת, האנשת הרובוט ותזמון התקשורת של המסירה.

מערך נתונים ציבורי של סרטוני מסירות נותח פריים אחר פריים כדי לקבוע את התנהגויות המבט השכיחות ביותר בהעברות אובייקטים בין אדם-לאדם. התנהגויות המבט השכיחות ביותר אשר נמצאו היו: התבוננות בידו של הנותן ולאחר מכן בפניו של הנותן (מבט יד-פנים), מבט תחילה בפניו של הנותן, לאחר מכן ביד הנותן ולאחר מכן חזרה להסתכלות בפניו של הנותן (מבט פנים-יד-פנים), והסתכלות רציפה על ידו של הנותן (מבט יד).

רובוט שיתופי Sawyer בעל 7 דרגות חופש תוכנת לבצע את משימת המסירה ולהציג התנהגויות מבט אלו. אובייקטים שונים עם סוגים שונים של תנוחות נותן-מקבל נותחו בשני מחקרים - מחקר וידאו ומחקר פרונטלי. במחקר הווידאו, 72 משתתפים צפו בסרטונים של העברות אובייקטים בין רובוט-לאדם, המדגימים את שלושת התנהגויות המבט, והשוו ביניהם. במחקר הפרונטלי, קבוצה אחרת של משתתפים ביצעה פיזית מסירות אובייקטים אל הרובוט והעריכה את תפיסת המסירות בנוגע להתנהגויות המבט השונות של הרובוט. התוצאות הראו שבשני המחקרים, כאשר הרובוט מביט תחילה בפניו של הנותן, לאחר מכן בידו של הנותן ולאחר מכן בחזרה בפניו של הנותן (מבט פנים-יד-פנים), המשתתפים החשיבו את המסירה כחביבה, אנושית ומתוזמנת יותר($p<0.005$). עם זאת, לא מצאנו עדויות להשפעת גודל החפץ, שבריריותו או תנוחת הנותן על העדפת המבט.

בחלק השני של המחקר, הערכנו את הפוטנציאל של אלגוריתם "חיפוש מדיניות מודרך", שזוהי שיטת למידה מבוססת מודל, של למידה באמצעות חיזוקים, לאימון בקר-רובוט להעברות אובייקטים בין אדם-לרובוט. חיפוש מדיניות מודרך היא מערכת חסכונית בנתונים שאינה מחייבת ידע מוקדם בנוגע לנתוני הדינמיקה של הרובוט והסביבה, ומספקת גישה מבטיחה למשימות העברה. עם זאת, על אף הדגמת חיפוש מדיניות מודרך במשימות ניווט שונות ומשימות מניפולציה אוטונומית, לא דווח על בחינת אלגוריתם חיפוש מדיניות מודרך במשימה פיזית של שיתוף פעולה בין אדם לרובוט.

במחקר זה, שלב ההגעה של המסירה מנוסח כבעיית למידה באמצעות חיזוקים, ולאחר מכן התבצע אימון של זרוע הרובוט השיתופי Panda עם 7 דרגות חופש הן בסביבת סימולציה והן ישירות על הרובוט הפיזי.

התוצאות שלנו מצביעות על כך שבחינת המדיניות שנלמדת בסביבת הסימולציה על הרובוט האמיתי, היא פתרון בלתי אפשרי ליישום בעולם האמיתי. בהערכת יעדים סטטיים בלבד, מצאנו שהביצועים של המדיניות הגלובלית שנלמד על ידי חיפוש מדיניות מודרך ניתנים להכללה טובה יחסית. עם זאת, ביצועי המדיניות הגלובלית השתפרו מעט על ידי הוספת בקרים מקומיים באזורים בעלי שגיאות הבחינה הגבוהות ביותר. בעת הערכת המדיניות הגלובלית שאומנה עם מטרות סטטיות על מטרה נעה, הרובוט יצר מסלולים מאוד לא יעילים והגיע לאזורים מחוץ לגבולות המיקום הקרטזיאניים שלו. אימון על מטרות נעות שיפר את המסלולים, אך הוביל לשגיאות "המקרה הגרוע ביותר" גדולות יותר. עם זאת, ניתן לטפל בבעיה זו על ידי הוספת בקרים מקומיים לשלב האימון, ובכך לשפר את ביצועי המדיניות הגלובלית.

מילות מפתח: העברות אדם-רובוט, שטף, אינטראקציית אדם-רובוט, אינטראקציית אדם-רובוט פיזית, מבט הרובוט, תקשורת בלתי-מילולית, תכנון מניפולציות, למידה באמצעות חיזוקים.

אוניברסיטת בן-גוריון בנגב
הפקולטה למדעי ההנדסה
המחלקה להנדסת תעשיה וניהול

העברות בין אדם לרובוט: העדפות האדם ולמידת רובוטים

חיבור זה מהווה חלק מהדרישות לקבלת תואר מגיסטר במדעי ההנדסה

מאת: תאיר פייביש
בהנחיית: פרופ' יעל אידן וד"ר ארמין ביז'

חתימת המחבר:............................ תאריך:02.01.2022
אישור המנחה:............................ תאריך:02.01.2022
אישור המנחה:............................ תאריך:02.01.2022
אישור יו"ר ועדת תואר שני מחלקתי:........................ תאריך:10/2/2022

טבת, תשפ"ב
ינואר, 2022

אוניברסיטת בן-גוריון בנגב
הפקולטה למדעי ההנדסה
המחלקה להנדסת תעשיה וניהול

העברות בין אדם לרובוט: העדפות האדם ולמידת רובוטים

חיבור זה מהווה חלק מהדרישות לקבלת תואר מגיסטר במדעי ההנדסה

מאת: תאיר פייביש
בהנחיית: פרופ' יעל אידן וד"ר ארמין ביז'

טבת, תשפ"ב
ינואר, 2022