Author:
Stolzenwald, S

Title:
Interacting with a Handheld Robot

Prediction of User Intention and Assisted Remote Collaboration

Author:
Stolzenwald, S

Title:
Interacting with a Handheld Robot: Prediction of User Intention and Assisted Remote Collaboration

University of Bristol

Department of Computer Science

Interacting with a Handheld Robot: Prediction of User Intention and Assisted Remote Collaboration

Schachar Janis Immanuel Stolzenwald

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Engineering.

Wednesday 7th April, 2021

Word Count: 44,120

Abstract

Handheld robots are intelligent tools that can process task and context information to help the user in a manual task. This concept bridges the gap between two traditional categories of robots, i.e. fully autonomous independent robots and highly controlled and dependent wearable devices. They enable novice users to complete a task through the provided task knowledge and accuracy, while the user can effortlessly navigate through uncontrolled environments. Recent work shows that the robot's autonomy can improve performance, however, this presents new challenges as the occurrence of mismatches between the robot's plans and the user's intention leads to frustration in users. To overcome this obstacle, we explore interaction concepts that could suit the requirements for handheld robots.

The first part of this work, concerns ways for the system's perception of the user and their intention. We use a tool-mounted gaze tracking system, which we use as a proxy for estimating the user's attention. This information is then used for cooperation with users in a generic reaching task, where we test various degrees of robot autonomy. Our results measure performance and subjective metrics and show how the attention model benefits the interaction and preference of users.

In the second part, we then use the attention profile over time to make predictions about the user's decisions. Using a support vector machine and the mounted eye tracker, the model derives users' intention from perceived gaze patterns. It yields real-time capabilities and reliable accuracy up to 1.5 s prior to predicted actions being executed. That way, the robot predicts one step ahead in the task and can align its plans accordingly. We assess the model in an assisted pick and place task and show how the robot's intention obedience or rebellion affects the cooperation with the robot.

In the third part, we go one step further in the dimension of human interaction and propose a system that involves three collaborating parties: a local worker, a remote helper and the handheld robot, carried by the local worker. The system enables a remote user to assist the local user through diagnosis, guidance and physical interaction through telemanipulation, with the robot completing subtasks autonomously. We show that the handheld robot can mediate the helpers remote instructions and actions, while the robots semi-autonomous features improve task performance by 24%, reduce the workload for the remote user and decrease required communication bandwidth between both users.

In this work, we explored new ways to interact with a handheld robot. We demonstrate that a tool that makes task decisions can collaborate more effectively when taking into account user intention during real-time task planning. Moreover, this study is a first attempt to evaluate how this new type of collaborative robot works in a remote assistance scenario, a setup that we believe is important to leverage current robot constraints and existing communication technologies.

Acknowledgements

First of all, I want to thank Walterio for all his help and support throughout my PhD and job as an associate teacher. He helped me to keep the big picture in mind with many inspiring ideas and discussions. I learnt a big deal about leading from him and I value most that for him, an individual's wellbeing has number one priority. This was particularly important to me during the times when my life outside the lab felt like being turned upside down. Thank you, Walterio.

Thanks to Austin, for making the handheld robot in the preceding generation of my PhD. His exciting work inspired me to start a PhD.

Thanks to my funding sources, namely the *Studienstiftung des Deutschen Volkes* and the *EPSRC*, keeping me fed and watered. The Studienstiftung connected me to incredible people and boosted both my personal and professional development.

Also thanks to my PGR colleagues, specifically Yannick, Louis, Michael, Hazel, Will, the other Will, Abel, Miguel and Faegheh and everyone associated with the VILab and 1CS. Thanks to everyone participating in my pilot trials and experimental studies.

Thanks to the great support by the members of *Bristol SU* and special thanks to everyone I met at Magic Society, especially George and William. It's been a very exciting time and I'm honoured that I could participate in the impact we made as a student group.

I also want to thank all my friends who supported me. Special thanks to Djen, without him I would never have ended up in Bristol. We had very inspiring fruitful discussions both about work and life in general. Thanks to Tobi and Zay for proofreading pretty much everything I produced during my PhD time and fruitful exchange of thoughts. Thanks to Ellie for her patient support when teaching me all the statistic methods you can find in this thesis, making it a rock solid piece of work. Thanks to Zottty, my oldest friend, for great times we had despite the distance that results from moving to another country. Your visits in Bristol are now in my little treasure box of memories. Thanks to Maurice and Amelia for being great friends and for providing me shelter during the pandemic.

Presumably, I would have completed my PhD much sooner without meeting Maurice and yet, it would have felt much longer, he's a great pal. Thanks to Luke for his support during challenging times in Bristol.

Thanks to all the house mates I've been living with, making my time in Bristol a really nice experience. Special thanks to Sara, who made me smile even during tougher times and also to Andy, Chiara, Mimi, Pi, Niharika and Laura. You all made me feel safe and home in the little flat we shared.

Great thanks to my family members. My sister Miriam was always there when I needed support, which took the edge off some difficult periods. I'm also grateful for the support from the other siblings too. Thanks to Oma and Opa for being the best grandparents I could ever ask for. May you rest in peace. Thanks to Mom and Dad for their encouragement and support too.

Finally, I want to thank Bristol and all the people I met during my journey there. It truly is a very special place on this planet. The friendly environment really made me feel home and enjoying my time. These years were by far the most interesting and intense ones I've had so far and they had a very positive impact on how I'm seeing the world now.

Outcomes

This research has the following publications and outcomes. The associated abstracts can be seen in Appendix A.

Publications

- Janis Stolzenwald and Walterio Mayol-Cuevas. I Can See Your Aim: Estimating User Attention From Gaze For Handheld Robot Collaboration. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3897–3904, October 2018.
- Janis Stolzenwald and Walterio W Mayol-Cuevas. Rebellion and Obedience: The Effects of Intention Prediction in Cooperative Handheld Robots. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3012–3019, Macau, China, November 2019.
- Janis Stolzenwald and Walterio W Mayol-Cuevas. Reach Out and Help: Assisted Remote Collaboration through a Handheld Robot . *arXiv preprint*, pages 1–8, 2020. (In submission for publication)

Contributions to the Following Workshops

- IEEE/RSJ **IROS 2019**: Human Robot Collaboration (HRC): Biomechanical Limits, Modeling and Testing to Support Safe Robot Contacts with Humans.
- **RSS 2020**: AI and Its Alternatives in Assistive and Collaborative Robotics: Decoding Intent.
- IEEE/ASME **AIM 2020**: Workshop on Supernumerary Robotic Devices.

In the Press

- **Tech HQ**: Computer scientists have made a robot that rebels against its user. 24 October 2019.
- **New Scientist**: Robot arm lets you remotely lend friends a helping hand with repairs. 6 November 2019.

- **Digital Trends**: Robot overlords? More like co-verlords. The future is human-robot collaboration. 13 November 2019.
- **Popular Mechanics**: This Robotic Arm Can Lend a Helping Hand With Repairs. 17 November 2019.

Supplementary Demo Videos

Video 1: Estimating User Attention From Gaze For Handheld Robot Collaboration. A summary of Chapter 3.

youtu.be/lsQ4k71NLTA

Video 2: The Effects of Intention Prediction in Cooperative Handheld Robots. A summary of Chapter 4.

youtu.be/H245WdJpNpE

Video 3: Assisted Remote Collaboration through a Handheld Robot. A summary of Chapter 5.

youtu.be/cTJ8tNJJXV0

Declaration

This dissertation is submitted to the University of Bristol in accordance with the requirements of the degree of PhD in the Faculty of Engineering. It has not been submitted for any other degree or diploma of any examining body. Except where specifically acknowledged, it is all the work of the Author.

[redacted]

Schachar Janis Immanuel Stolzenwald, Wednesday 7$^{\text{th}}$ April, 2021

Contents

List of Figures

List of Tables

List of Abbreviations

ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ANS	Autonomic Nervous System
API	Application Programming Interface
AR	Augmented Reality
AI	Artificial Intelligence
BCI	Brain-Computer Interface
BPM	Beats Per Minute
CAD	Computer-Aided Design
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CV	Computer Vision
DoF	Degrees of Freedom
ECG	Electrocardiography
EMG	Electromyography
EEG	Electroencephalography
GSR	Galvanic Skin Response
HF	High Frequency
HMD	Head Mounted Display
HMM	Hidden Markov Model

HRI Human-Robot Interaction

HR Heart Rate

HRV Heart Rate Variability

IK Inverse Kinematics

IPA India Pale Ale

IR Infra Red

LED Light-Emitting Diode

LF Low Frequency

LF/HF Low Frequency over High Frequency

LSTM Long-Short Term Memory

LCD Liquid Crystal Display

MMTS Magnetic Motion Tracking System

pHRI physical Human-Robot Interaction

PNS Parasympathetic Nervous System

RGB Red Green Blue

RGB-D Red Green Blue Depth

RNN Recurrent Neural Network

SD Standard Deviation

SDK Software Development Kit

SLAM Simultaneous Localisation and Mapping

SRL Supernumerary Robotic Limbs

SUS System Usability Scale

SNS Sympathetic Nervous System

SVM Support Vector Machine

TV Television

TLX Task Load Index

VAP Visual Attention Profile

VR Virtual Reality

2D 2-dimensional

3D 3-dimensional

Chapter 1

Introduction

1.1 Background and Motivation

Handheld tools reach far back in human history with the earliest archaeological evidence rooting back to the early Stone Age - 1.75 million years ago [119]. Since then, the use of tools has been a major evolutionary advantage and their design has been developed significantly towards devices that are characterised by high complexity and specialisation. This allows humans to complete tasks that they could not achieve without these tools or only with less efficiency. In other words, they enhance humans physical capabilities.

While the physical design of tools is at an advanced stage, there is a new aspect that has been explored in recent years. That is the tool's understanding of what it is used for and its knowledge about the context and environment it is being used in. The evolution of handheld tools concerning their degree of complexity and autonomy is depicted in Figure 1.1. In this thesis, we refer to *intelligent tools* as handheld devices that can assist in a task through the control of *some* degrees of the tool, e.g. controlling the amount of power used or the tool-tip position in a corrective manner. Our definition of *handheld robots* extends the concept of intelligent tools as they are able to control *multiple* degrees of the device and make high-level decisions about the task, i.e. choosing which object to interact with and which action to perform with or on it. As such, they are robots, that can be hold in hands.

Over the recent decades, a new taxonomy of intelligent tools has been established with groundbreaking innovations in the field of medical applications [92, 128] and fabrication [249, 251]. They can process context information, which allows for assistance concerning accuracy and the suppression of unintended movement. However, these tools are limited to a low number of DoF, small workspaces and high specialisation.

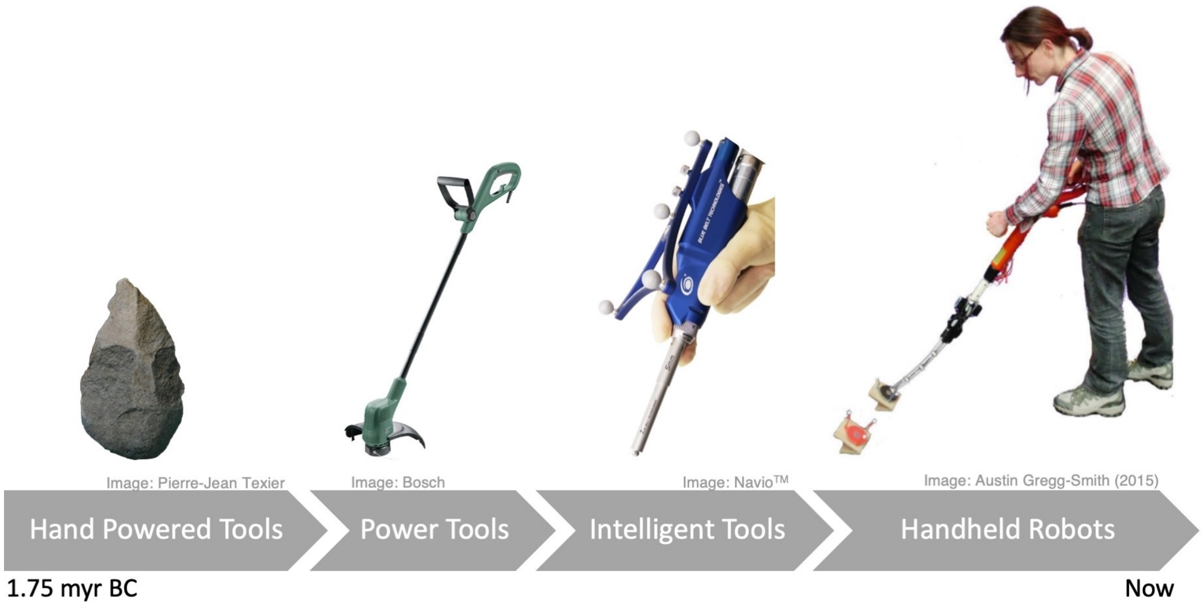


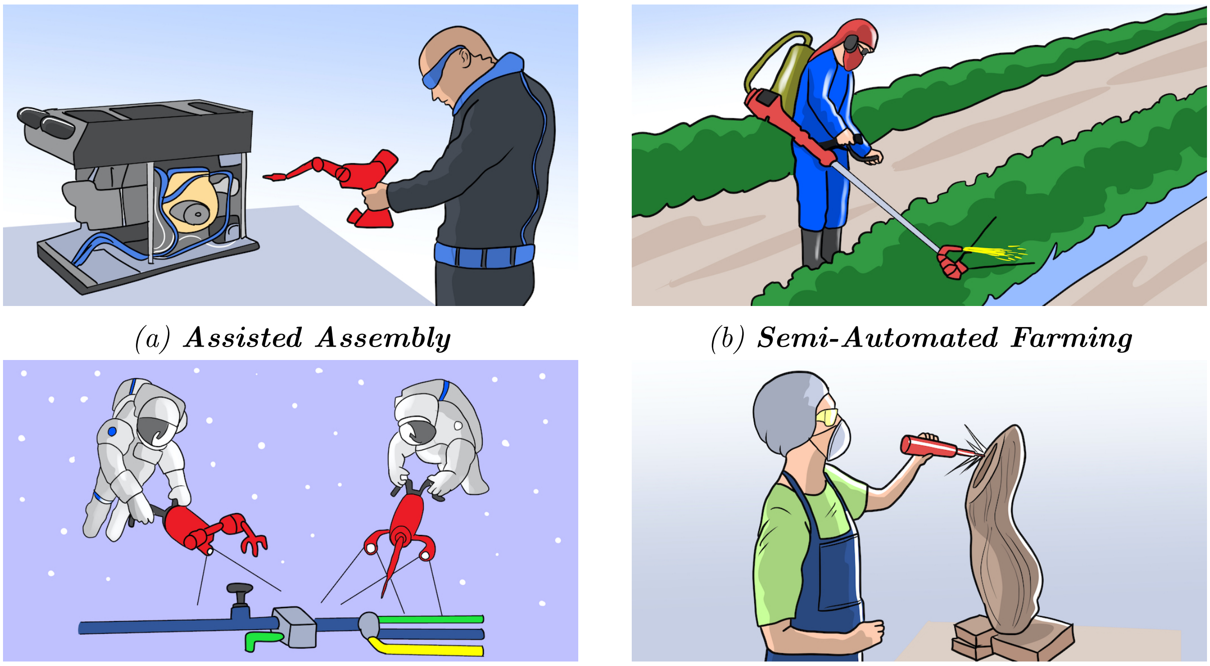
Figure 1.1: ***Evolution of Handheld Tools****. This shows the development of handheld tools. The earliest known today are from the early Stone Age [119]. Modern hand tools are externally powered (e.g. lawn trimmer), incorporate intelligent control based on task knowledge (e.g. sculpting in surgery) and make high level decisions about the task (e.g. handheld robots).*

In recent years, research in this domain [48, 49, 58–60] has led to a further development from assisting tools to collaborative handheld robots that take into account task knowledge such as progress, required steps and their sequence as well as knowledge about relevant objects, including their identity and location. To date, person-oriented robots have been developed mostly in the extrema of fully external autonomous devices such as robot delivery systems or intimately linked to the body of users such as in exoskeletons [56, 120, 195] or Supernumerary Robotic Limbs (SRL) [168, 175, 226]. Handheld robots have the potential to bridge the gap between these extremes and combine the benefits of an autonomous helper and the physical proximity to users that allows for intuitive and immediate control.

The purpose of handheld robots goes beyond assistance that would be limited to some aspects of the task. Their autonomy allows for a collaboration between the robot and the user concerning decision making and progress tracking. *Moravec's Paradox* [150], which states that tasks that are easy to complete for humans provide difficult challenges for machines and vice versa, is turned into a profit through the new paradigm of handheld robots: The robot can complete tasks that overlap with machine capabilities, e.g. mechanical accuracy and speed, whilst the human can focus on the parts of the task that they can naturally complete with little effort, e.g. overall task planning, and the negotiation of obstacles.

This synergy has been observed in recent work on handheld robot prototypes. Notably, tasks could be completed more accurately and with less task load for the users with an increase of the robot's autonomy, i.e. its share of made decisions. Analogous to hand tools mechanically enhancing humans, the robot can assist on a cognitive level, i.e. novice users could complete tasks with the robot, which they could not achieve without it, e.g. due to a lack of expertise.

The developments away from specialised hand tools and towards general purposed handheld robots open up new application fields. For example, in the future such devices could be applied to assist workers in the manufacturing industries, e.g. through assisted assembly or welding. Another conceivable application would be in agriculture, e.g. the robot could pull unwanted weeds or selectively irrigate crop plants while the user carries it through uncontrolled environments and uneven terrain. Even the arts industries could benefit from the progress in handheld robots, e.g. in assisted free-hand sculpting. An overview of envisioned handheld robot applications is presented in Figure 1.2.



(a) ***Assisted Assembly*** *(b)* ***Semi-Automated Farming***

(c) ***Maintenance*** *(d)* ***Tool-Aided Sculpturing***

Figure 1.2: ***Possible Future Handheld Robot Applications****. Handheld robots have the potential to transform the way we use tools as they participate in task decisions, e.g. in assembly (a), farming (b), maintenance (c) and sculpting (d).*

Beyond an improvement in productivity, handheld robots might also help with a more positive association to task outcomes. With the human being involved in some of the steps of an otherwise completely automated production chain, this participation could lead to an increased valuation of the products because they could be considered self-made - a phenomenon that is also known as the *IKEA Effect*[138, 157].

Early work on handheld robot focused on the mechanical design of first prototypes [58, 59] and explored the introduction of robot autonomy to a handheld tool. Furthermore, work in this domain explored various ways to communicate the robot's high-level plans to the user [60]. While these contributions paved the way towards the vision of handheld robots, there are still obstacles that need to be overcome to make it a reality. New challenges arise from the fact that the tool itself makes decisions about the task and participates in task planning. In their studies on cooperative handheld robots, Gregg-Smith and Mayol-Cuevas [58, 60] observed that in some instances the robot's plans did not match users' expectations as they conflicted with their intention. This effect is a general problem in the collaboration with machines and known as *automation surprise* [188].

The introduction of robot-human communication strategies, e.g. rudimentary tooltip gestures [58, 60] helped to leverage the collaboration but also led to the robot's actions dominating user decisions, resulting in irritation and frustration. Therefore, Gregg-Smith and Mayol-Cuevas [58, 60] suggested that handheld robots require more sophisticated strategies for communication in the opposite direction, i.e. human-robot. This raises the question of how robots can gain an understanding about the user's plans, i.e. what they want to interact with next and what sequence of steps they choose to carry out in their underlying solution strategy. I argue that a system for the prediction of user intention can leverage smooth cooperation between these two agents and can help users in regaining their position of control over the task. This thesis looks at explicit and implicit human-robot communication cues and explores how handheld robots can interpret and process them for cooperative task solving.

Another interesting aspect of intelligent tools is that, in contrast to traditional tools, they can be controlled by more than one person. Shared control concepts of this kind have been explored in collaboration studies with multiple users controlling the same robot, e.g. for the training of surgeons [10, 162]. Other work explored the overlap of the workspaces of a human and a remotely controlled robot [193]. Often, the purpose of such arrangements is to consult the help of a remotely located user. While these setups yield promising results concerning remote collaboration, there is a gap in the existing research with regards to the possibility of a hand tool being controlled remotely, e.g. for remote assistance. Remote assistance is a research field that has gained importance due to the continuous growth of complexity in machinery and plants [21], since more expertise is required for tasks such as maintenance or when an unexpected problem occurs. Consulting an expert, who might be located oversea, is expensive and relying on manuals for technicians is inefficient [231]. For this reason, there is a high demand of solutions for remote guidance systems (e.g. [53, 54, 106]) that allow the efficient instruction of novice users by a remote helper.

An alternative to guidance for remote helping is to use teleoperation, i.e. to carry out inspection and maintenance tasks through a robot [16, 17, 195, 204]. Current solutions for this are expensive because they require locomotion systems. I suggest that a handheld robot can combine aspects of remote guidance and teleoperation. Its tactical motion could be performed by a host carrying the robot, whilst the remote expert could have physical access to the worksite through the remote control of the robot's tooltip. In other words, the handheld robot could mediate instructions and tooltip motion to combine the competencies of the two collaborating parties.

Work in the domain of remote collaboration involves robotic devices that are controlled in a strict master-slave manner (see Section 2.8). However, we believe that the handheld robot itself could be considered as an agent that is part of the collaborating team. That way, some parts of the tasks could be delegated to the robot's autonomy to be completed on a smaller local level, whilst humans could focus on task planning and navigation. We hope that an introduction of automation of some parts of the robot's control would benefit the collaborating humans analogue to the already explored single-user setups.

1.2 Aims and Objectives

This research is guided by two main goals. The first one concerns the interaction between the handheld robot and its user in a single-user setup. The second one addresses the challenge of bringing together three collaborating agents, i.e. the robot, its host and a remote user in a remote assistance setup.

With this in mind, our first aim is to explore new ways of human-robot communication for handheld robots with a focus on predicting a user's intentions. We use gaze data as the main predictor for the intention model, employ it to bias the robot's decisions towards a user's plans and assess its effect on cooperation quality through experimental studies. This can be broken down into the following objectives:

- Incorporating gaze tracking and evaluating eye gaze with respect to the poses of the handheld robot and task objects, i.e. creating a 3-dimensional (3D) model that allows for a proxy of a user's visual attention. Subsequently, this requires testing the gaze model in a feasibility study through user experiments.

- Creating a representative example task that covers common human-robot interaction challenges such as coordinated navigation and subtask decision timing. This can then serve as a basis for the collection of training data. The main requirement for the data set is to contain information about a user's aims and associated preceding gazing patterns as a basis to train an intention model.

- Selection and training of a machine learning model to anticipate user intent. Incorporate predictions in the robot's online task planning and assess it in action through experimental studies.

Secondly, we aim to better understand how handheld robots can mediate remote collaboration and do so through experimental validation of both performance and measures of usability and collaboration in our proposed human-robot-human telepresence setup. Consequently, we derive the following objectives:

- The development of an overall concept for human-robot-human remote collaboration with the handheld robot as a mediator between a remote helper and a local hosting user. This requires a definition of the respective roles involved, i.e. the remote user, the local user and the robot.
- Adaptation of the existing handheld robot setup. This includes the integration of sensors for remote perception of the environment, e.g. cameras for visual feedback. Moreover, the setup requires a work station for the remote expert that allows for 5-DoF control inputs and incorporates a display of sensor data.
- Testing the system requires a representative example task that imposes common remote collaboration challenges such as remote inspection, diagnosis, guidance and instruction.
- Developing a concept for mixed-initiative interaction to merge remote and autonomous control and define a subset of tasks that can be delegated to the robot. Then assessing the task in a feasibility study, which are based on experiments and usability questionnaires to assess the performance of the robot as a mediator and its effect on human collaborators concerning task load and usability scores.

1.3 Contributions

This research contributes to an understanding of how humans can interact and collaborate with a handheld robot system with the main subdomains being single-user and multi-user interaction.

1.3.1 Single-User Interaction with the Handheld Robot

Estimation of Visual Attention

- We deliver a method to combine an existing gaze tracking solution with motion capturing to convert in-plane gaze-information to a ray in 3D space. This allows for an estimation of where users are looking at in the robot's environment. The experimental validation delivers the limits of the tracking setup and informs the boundaries of applications for the handheld robot.

- We introduce gaze-based user attention estimation and demonstrate how this information can be used to parametrise the robot's cooperative behaviour.
- To assess the attention system, we propose a reaching task as basis for an experimental setup that allows measuring collaborative performance for validation of the model. Parts of the tasks are simulated in a TV screen, which makes it easy to adapt to future experiment proposals.

Intention Modelling

In contrast to the more reactive attention model, an intention model allows for predictions of user action in the proximate future based on the recent history of the user's attention. Concerning intention modelling, this research makes the following contributions:

- We introduce an intention model for handheld robots with real-time capabilities. It predicts users' intentions to interact with objects in the work scene based on eye gaze and task states. We show how context information can increase the model's accuracy, which we suggest generalises to applications beyond handheld robots.
- We propose a block copy task as a generic example, which is used for data collection and model validation. The task demands timely decisions from the user concerning accuracy and sequence of task steps. Therefore, the resulting completion times serve as a proxy for collaborative performance.
- As a new way of model validation, we introduce obedience and rebellion as intention-based anticipatory behaviour modes. In the absence of universally accepted physiological metrics, this serves as a proxy to evaluate the intention model in action via the differences in robot-induced user frustration levels.

1.3.2 Multi-User Interaction in Handheld Robot Collaboration

The second set of contributions concerns the interaction between two users via the handheld robot.

- We introduce a paradigm of remote collaboration between two humans, with the handheld robot at its core mediating instructions and movement commands between a remotely located helper and a novice user at the local worksite.
- The proposed remote maintenance task imposes common challenges in remote assistance, i.e. problem diagnosis, guidance and collaborative problem-solving. We suggest that this can serve as a benchmark task to compare remote assistance systems.

- We demonstrate that the cognitive load can be transferred from the remote operator to the robot through the delegation of parts of the task to the robot, which in turn completes sub-tasks on a local scale autonomously.
- Our qualitative analysis reveals new collaboration behaviours that emerge between the three cooperating agents, i.e. the remote expert, the robot and the local user. We observed a common task solving pattern with a repeated sequence of exploration, guidance, local task solving and retraction. This information can be used as a guideline for future designs of collaboration systems.

1.4 Thesis Outline

This section gives an overview of the subjects addressed in the remainder of this thesis which is organised as follows:

- **Chapter 2 — Background: A Review of Handheld Robots, Intention Prediction and Remote Assistance** presents an overview of related work that is required for an understanding of the technical chapters. In particular, it reviews the current state-of-the-art of handheld robots and compares their capabilities with existing specialised intelligent handheld tools for medical applications and fabrication purposes. Furthermore, this chapter outlines the overlap between the field of handheld robotics with wearable devices and its implications for challenges concerning mixed-initiative interaction between the tool and its user. Finally, a summary of methods for intention prediction and concepts for remote guidance and telemanipulation form the background basis that informs decisions made concerning single and multi-user setups for collaborative interaction with the handheld robot.
- **Chapter 3 — I Can See Your Aim: Estimating User Attention From Gaze For Handheld Robot Collaboration** explores user-gaze as a new basis for attention-driven human-robot interaction for handheld robots. We integrate a remote eye gaze tracker with the existing robot hardware. The eye gaze data that is delivered with respect to the frame of the tracking device is then coupled with a motion capturing system to construct a 3D gaze ray, which is used as a proxy for the user's attention. The limits of this gaze tracking setup are assessed through experimental studies. In a second step, the gaze model is used to implement an attention-driven behaviour model to assist in a reaching task. The setup is tested for a range of temporal demands in the task and the robot's behaviour assessed with varying levels of autonomy and gaze-awareness. We show that the teamwork performance between a fully autonomous robot versus an attention-driven robot is similar. However, we found that less task load was perceived with attention bias

enabled, particularly when task completion required fast reactions.

- **Chapter 4 — Rebellion and Obedience: The Effects of Intention Prediction in Cooperative Handheld Robots** takes the interaction concepts that were developed in the previous chapter one step further. The gaze model and the user's gaze pattern is used to make predictions about the user's intention with a generic block copy task as an example. The intention model predicts which object the user wants to pick up next and where they want to place it. The chapter consists of two main parts. The first part describes the procedure of data collection through experiments. This data is then used to train an SVM, which forms the basis of the prediction model. Data modelling and validation is completed offline to determine the model's accuracy based on the generated training and cross-validation data. In the second part of the chapter, the intention model is used to bias the robot's task decisions during assistance to validate the intention model in action. Employing the counter-intuitive strategy of using the attention model to make the robot perform the opposite of the predicted user intention, we introduce a new method of intention validation via frustration levels. The idea is that the robot can only frustrate users if the rebellion was based on correct predictions. Thereby, we show that the intention system can make reliable online predictions one step ahead during task completion.

- **Chapter 5 — Reach Out and Help: Assisted Remote Collaboration through a Handheld Robot** extends the challenge of efficient interaction with the handheld robot through the introduction of another user. We explore a remote assistance setup with the handheld robot at its core. It involves a local novice user holding the robot, who receives assistance from a remote expert with the handheld robot serving as a basis for remote guidance and manipulation. In contrast to traditional remote guidance setups, the handheld robot allows for physical access to the local work scene. Furthermore, the robot is considered as a semi-autonomous agent that can assist through decision making and compensation of unintended motion as main assistive features. In a first step, the robot is equipped with cameras for visual feedback to the remote operator and a remote workstation designed that allows them to access the robot and to control it. The system is tested through experimental studies with a remote maintenance task as a generic example for remote collaborative task solving. Our results show how handheld robots can bring together the competences of both the local user and the remote expert and show how the assistive features of the robot can help in higher performance and better usability scores.

- **Chapter 6 — Conclusion and Further Work**. This chapter summarises the research presented in this thesis, outlines and discusses our main findings and puts them into perspective concerning their implications for the robotics research community. We close with an outlook for future work, which could serve as a starting point to accelerate innovation and research in this domain.

Chapter 2

Background: A Review of Handheld Robots, Intention Prediction and Remote Assistance

This chapter covers the related work on which our motivation introduced and described in Chapter 1 is based on. Furthermore, this chapter introduces the reader to background material that informs the methodical decisions made in Chapter 3, 4 and 5.

2.1 Chapter Overview

The research presented in this thesis explores the design scope of handheld robot interaction with a focus on the collaboration between the robot and the humans involved in the respective setups. This chapter reviews related hardware and existing interaction concepts. Each of the technical chapters contains a summary of related works as part of the respective introductions. In that way, the chapters are self-contained and the interested reader is referred to the following sections for details about relevant literature.

Firstly, we introduce the reader to recent literature on handheld robots (Section 2.2). This research mainly builds on this literature with the hardware presented in Section 2.2.2 at its core.

Handheld robots, in a more general sense, can be understood as intelligent tools that make autonomous task-related decisions. An overview of this type of intelligent devices is presented in Section 2.3. These are mostly from the fields of surgery and fabrication.

Wearable robots are close relatives of handheld robots as they are characterised by high physical proximity to the user. Some of the interaction problems that result from this property can be found in this domain as well. Therefore, Supernumerary Robotic Limbs (SRLs) and initial designs of wearable robotic arms are covered in Section 2.4.

All studies presented in this thesis concern Human-Robot Interaction (HRI) within handheld robotics and particularly the aspect of shared control. For this reason, a summary of existing strategies for the implementation of mixed-initiative interaction can be found in Section 2.5.

A key strategy of aligning the robot's actions with the user's plans is to bias the robot's decisions based on the predictions of an intention system. This subject is introduced in Section 2.6, where we discuss different approaches based on various modalities and cues for intention prediction and examples for their application in robotics.

Finally, we are interested in exploring shared control of the handheld robot with the introduction of an external remote helper. The principal elements of this part of our work are remote guidance and teleoperation, which are introduced in Section 2.7 and 2.8, respectively. The chapter closes with a summary that links the background sections to the technical chapters.

2.2 State-of-the-Art Handheld Robots

This section presents the latest developments in handheld robotics. Here, the focus is on work by Gregg-Smith and Mayol-Cuevas [58, 59, 60], as the research of this thesis is based on the hardware of the 6-DoF Handheld Robot (see Section 2.2.1), which is the successor of a 4-DoF design (see Section 2.2.2). Furthermore, related designs and their applications are summarised in Section 2.2.3.

2.2.1 4-DoF Handheld Robot

The notion of non-medical, generic handheld robotics was proposed by Gregg-Smith and Mayol-Cuevas [58]. They introduced a trunk-shaped lightweight tool, which is tactically moved by the user while being able to carry out small-scale motion with its 4-DoF tip. The tool itself is aware of the task and its progress and can use this knowledge to augment users. For instance, it can provide task-related guidance using its 4-DoF end effector to point towards the goal or execute local manipulation. An overview of this concept can be seen in Figure 2.1.

Within a feasibility study, the authors investigated the effect of the robot's autonomy on task performance, using a pick and place task and virtual paint spraying as examples. In both tasks, the experiment was repeated for three different levels of the robot's autonomy:

- **Manual:** the robot's tip would not be actuated at all and remain in its initial position so that the user is completely in charge of moving the tooltip.

- **Semi-Autonomous:** the robot's tip is actuated to guide the user to the next goal of the task and to support spacial and angular aiming.

- **Fully Autonomous:** the robot would not only control the tip's orientation but would also override the input of the trigger. Therefore, it could refuse to colour wrong pixels within the painting task or to pick up an unnecessary item (see Figure 2.1 b and c).

The results show that increasing the level of autonomy improves critical aspects of efficiency such as time-to-complete and perceived workload. At the same time, increased autonomy led to frustration in some participants which was expressed through statements like: *The tool won't go where I want it to*[58]. The authors assumed that a high level of the robot's autonomy would increase the demand for effective human-robot cooperation. Therefore, they suggested improving the communication of the robot's plans and predicting user intention to address this problem in future research.

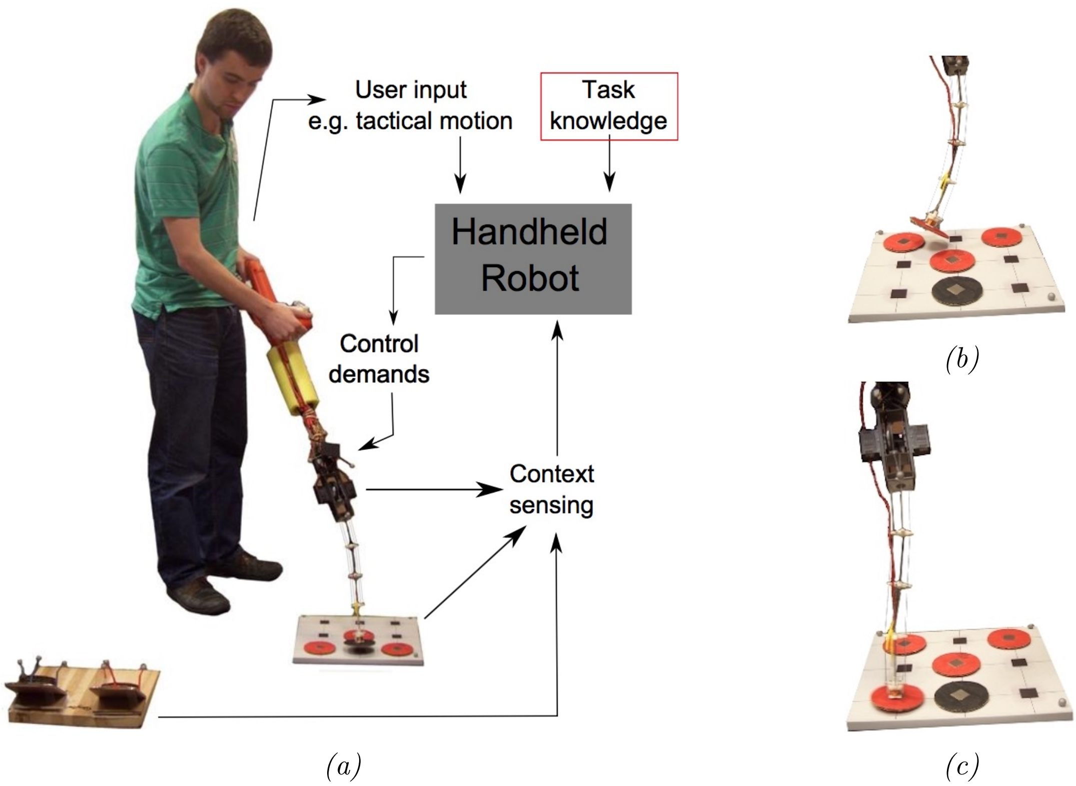


Figure 2.1: ***The Concept of a Handheld Robot.*** *Overview of the cooperative handheld robot and its main components during a tiling task (a). In autonomous mode, the robot refuses to place the tile at a wrong position (b) but assists when the user aims for the correct one (c) [58].*

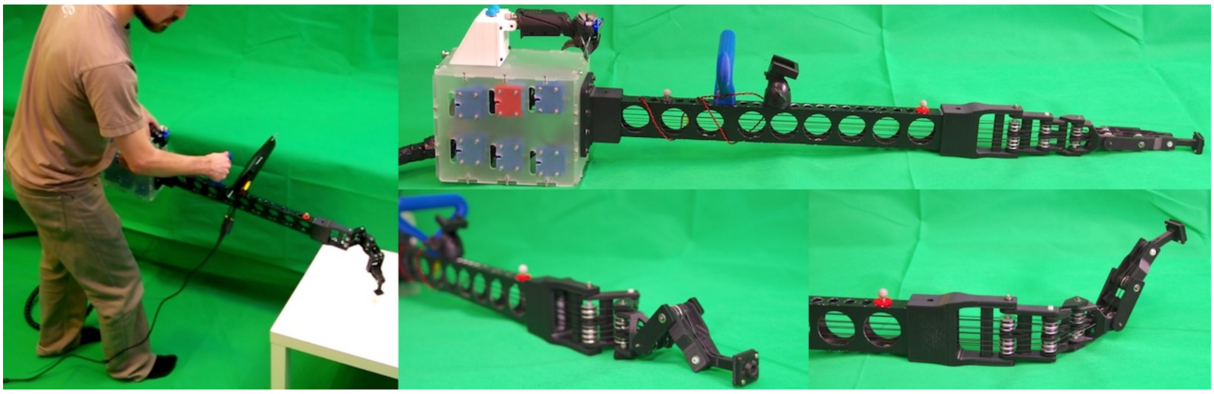


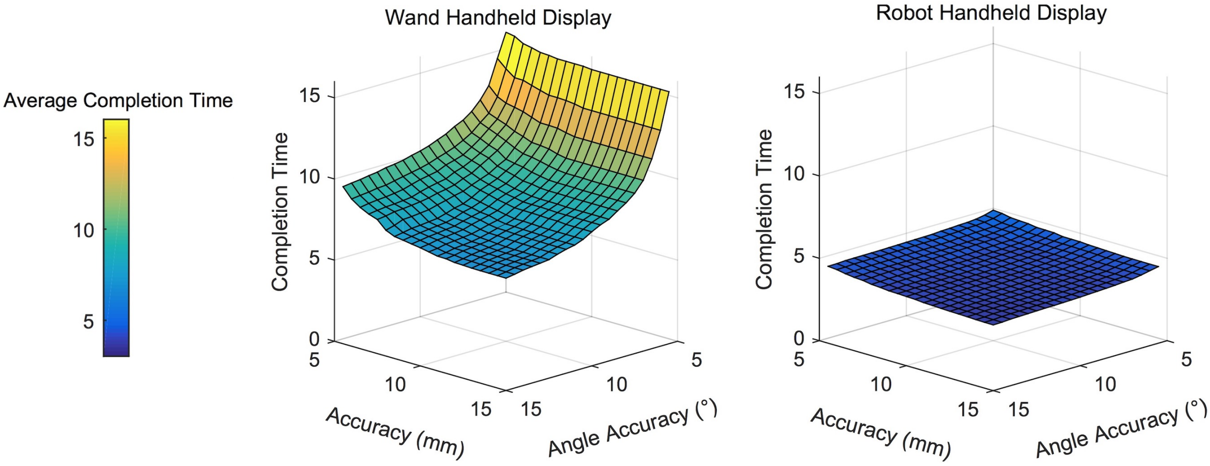
Figure 2.2: Updated design of the handheld robot with increased acceleration and dexterity of the 6-DoF tip [59].

Figure 2.3: 6-DoF handheld robot pointing its tip towards the point it needs to be positioned in order to indicate goal direction. In this example, an LCD display is used for visual feedback [59].

2.2.2 6-DoF Handheld Robot

The 4-DoF design of the handheld robot was later updated to a new version that features 6 DoF in the joint space and 5 DoF in the workspace [59]. The joint redundancy allows for local obstacle avoidance and the new cable-driven design enables motion with high acceleration, speed and dexterity, realised in a lightweight design that can be carried rather comfortably. Figure 2.2 shows the handheld robot, which is also available as an open-hardware project [1]. Another contribution of the aforementioned work is the implementation of the path planner which is based on their derived Inverse Kinematics (IK). Their method allows for solving the IK for a given point outside the robot's workspace. In that case, the robot points towards this point, which is an essential feature for user guidance through tooltip gestures.

Within experimental studies, Gregg-Smith and Mayol-Cuevas [60] investigated the impact of visual feedback on user performance in a 5-DoF reaching task. In addition to the robot's rudimentary tip gestures, an arm-mounted LCD screen, see-through AR or a VR headset were used to communicate the robot's plans as shown in Figure 2.3. The experiment participants were asked to fulfil the job as quick as possible while the level of difficulty was adapted by modifying the desired angular accuracy. For a reference, the task was repeated using a handheld wand for reaching instead of the robot's tooltip. Rather than



(a) (b)

Figure 2.4: ***Decoupling Execution Speed from Accuracy Demands.*** *This diagram visualises the results of the feasibility study concerning the relationship between accuracy and completion time for the condition when the wand is used (a) and when the robot is used (b) to complete a reaching task. Note that the completion time remains constant over accuracy demands when the robot is used, whereas more time is needed to complete the task for higher accuracy when the task is done manually [59].*

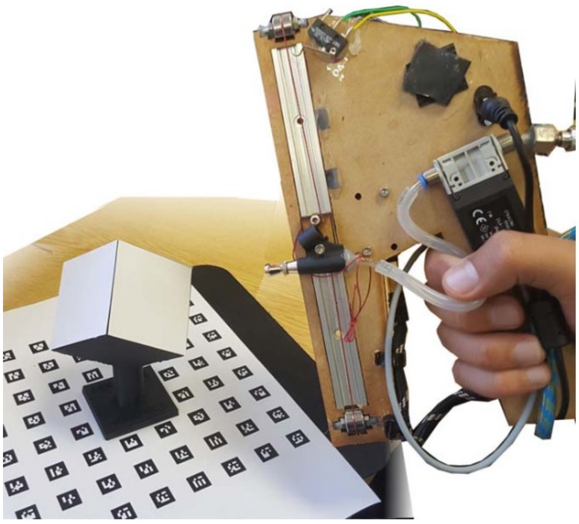
containing any robotic features, the wand was used for tracking purpose only to be able to determine task completion.

The results of the experiments show the participants performing much better when using the robot compared to using the handheld wand, regardless of which means of visual feedback was used. At the same time, participants benefit from increased accuracy while perceiving less workload. Another interesting aspect of the results is the relationship between accuracy and completion time (see Figure 2.4). According to [51, 207], human motor performance is limited to the combination of accuracy and speed so that accuracy performance would suffer from speed increase at a certain level and vice versa. This theory about 2-dimensional (2D) hand pointing is well known as Fitt's Law and has recently been extended by [31] to 3D reaching with similar results. This property is reflected in the wand-use experiment result as accuracy decreases with the decrease of completion time and thus a higher level of speed (see 2.4b). Interestingly, this is not true for cases where the robot was used as results yield a constant level of completion time independent of required accuracy. Gregg-Smith and Mayol-Cuevas [59] suggest that this indicates that the use of robots has the potential to decouple human's accuracy from performance speed which is an essential efficiency criterion for a wide range of manual tasks. At the same time, the research question arises how to determine user intention to enhance cooperation.

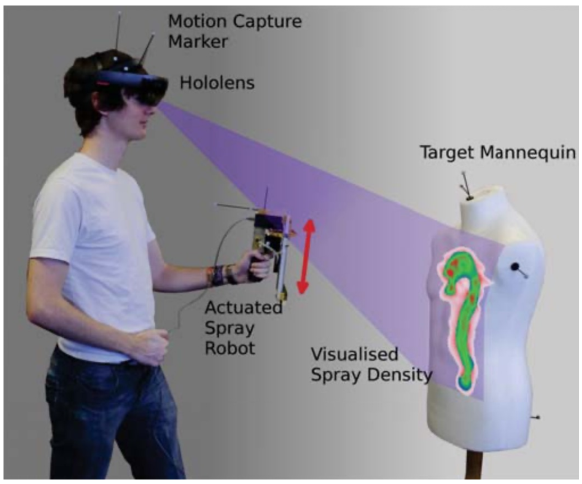
We note that the above works convey useful information in terms of the design of handheld robots and their impact on cooperative task solving. At the same time, the results raise new research questions such as how an effective human-robot communication could look like, what other means of guidance beyond visual feedback could be used and how the robot could retrieve the task knowledge, which its behaviour is based on.

2.2.3 Assisted Painting Using a Handheld Robot

Another contribution in the field of handheld robots is the work by Elsdon and Demiris [48, 49], who introduce an intelligent paintbrush for the application of skin medication on human bodies. Whilst their concept follows the one of Gregg-Smith and Mayol-Cuevas [58] concerning the robot's levels of autonomy, their design is adapted to the specific application of paint spraying. The device houses a trigger to control the valve to spray paint through a nozzle, which can move with 1 DoF along a gentry (see Figure 2.5a). The intelligent paintbrush enables the possibility of leveraging the user's ability to move around the environment while the robot performs the final actuation and keeps track of the task progress. This is fed back to the user through AR in-scene highlighting of the 3D object to be painted to indicate over-spray, under-spray or correct amount. The set up is cooperative in the way that no entity can fulfil the task without the other one. A system overview can be seen in Figure 2.5b.



(a)



(b)

*Figure 2.5: **Intelligent Paint Spraying Gun**. 1-DoF handheld robot with actuated nozzle (a) proposed by Elsdon and Demiris [48, 49]. The robot is aware about the task goals and progress, can feedback the task states through a head-mounted display and provides the final actuation (b).*

Analogous to Gregg-Smith and Mayol-Cuevas [60], the authors found that the use of the robot and the provided feedback helps increasing the quality of the work outcome in terms of accuracy. At the same time, some users found that the fully automated mode spoiled their plan or they found that they were fighting with the system. We note that these studies demonstrate the necessity of a form of human-robot or robot-human communication to overcome differences of plans when dealing with tools that make decisions about the task at hand.

2.3 Intelligent Handheld Tools

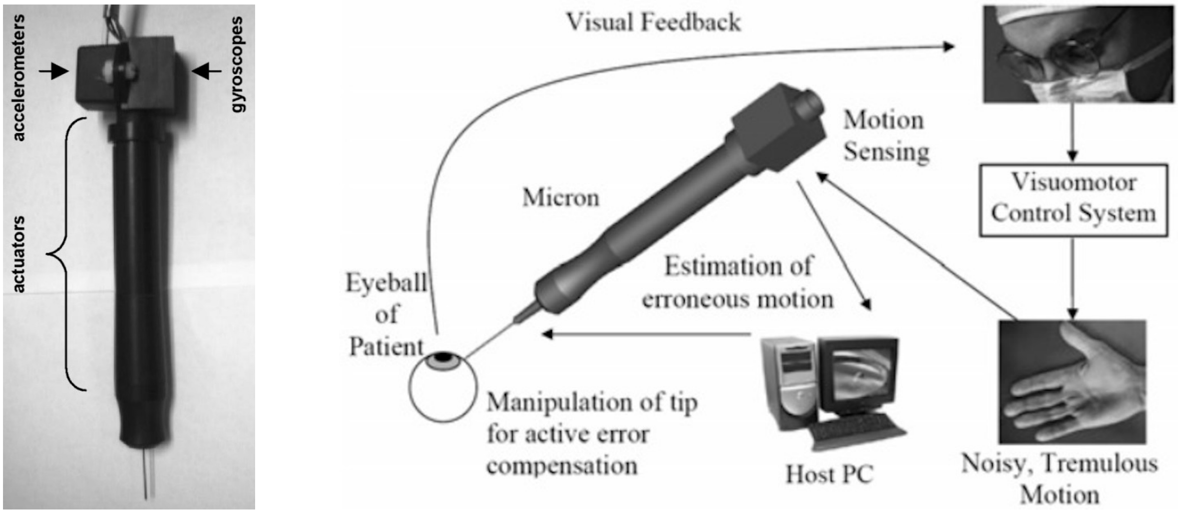
Intelligent tools can assist their users through enhanced accuracy and the correction of motion. In contrast to more general handheld robots, they are purposed for a specific task. They can process task knowledge such as a work part's location and spacial relationship to the tool. Commonly, this information is used to stabilise the tooltip or to start/stop its main function as the user approaches the boundaries of an underlying target model, e.g. the contour of a painting [202] or the surface of delicate tissue in surgery [92, 128]. In contrast to more complex robots, these devices assist but do not make decisions about a task, e.g. the order of steps to complete it. The following presents examples of intelligent tools the majority of which are purposed for medical applications (Section 2.3.1/2.3.2) and fabrication (Section 2.3.3).

2.3.1 Tremor Suppression for Medical Devices

Physiological tremor becomes a problem when its magnitude exceeds the size of the object that is handled. This is particularly true for operations on delicate tissues in microsurgery [184]. One solution for this problem would be to use a stationary robot that downscales the surgeon's motion such as implemented in the *da Vinci* surgical system [11, 23]. Such systems offer many features of the human hand, however, they are expensive and take up substantial of space. One could argue, that the required dexterity was already provided through the surgeon's hand and making their movements more precise could be a rather efficient solution while they retain their naturalness of feel [248]. For these reasons, research groups started investigating handheld solutions concerning tremor suppression.

A majority of contributions in this field concern the development of *Micron*, an instrument for microsurgery, which was introduced by Riviere et al. [185] and can be seen in Figure 2.6a. It features a 3-DoF actuated tip, which is driven through a parallel sequence of piezoelectric actuators. This design was later updated with a flexure-based manipulator which increased the size of the workspace [33]. The device enables active compensation of tremor with a magnitude of 50 μm at a frequency of up to 12 Hz. An overview of the Micron concept can be seen in Figure 2.6b.

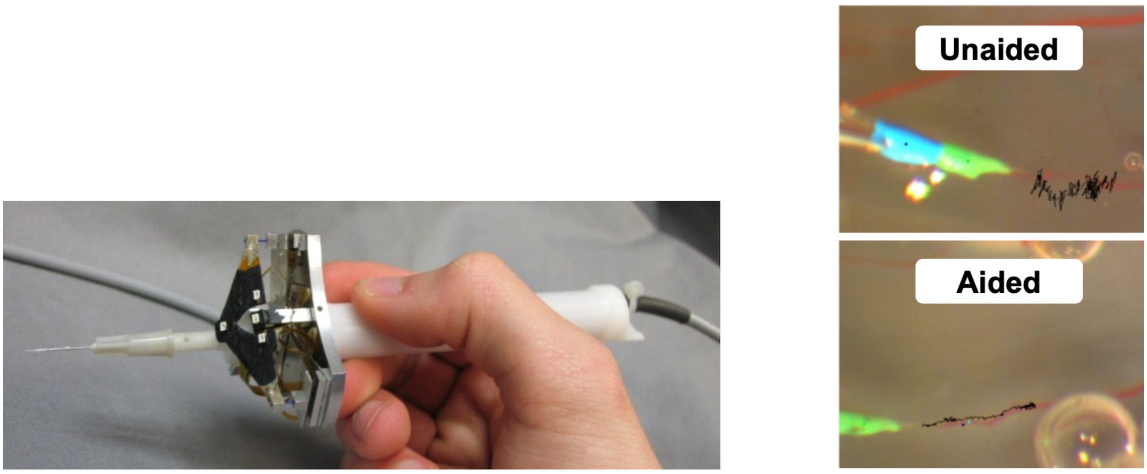
Becker et al. [14] then further improved the system in terms of its estimation of erroneous motion through vision-based control and optical tracking, which was then tested



(a) (b)

Figure 2.6: ***Micron, an Intelligent Surgical Tool.*** *(a) First prototype of the device as introduced by Riviere et al. [185]. (b) Overview of the Micron concept and its interactions with a human in the loop [33].*

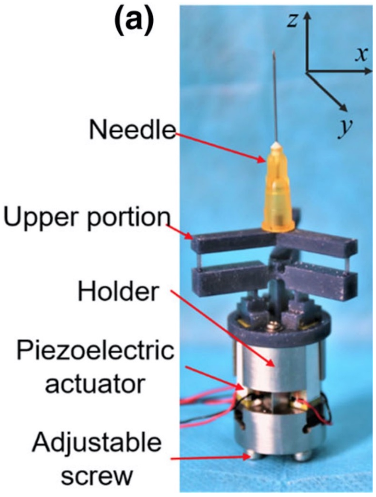
for operations on retinal vessels [14], photocoagulation [13], retinal membrane peeling [15] and intraocular laser surgery [241]. Figure 2.7a shows the device and the differences in performance depending on the tremor suppression being enabled or not.



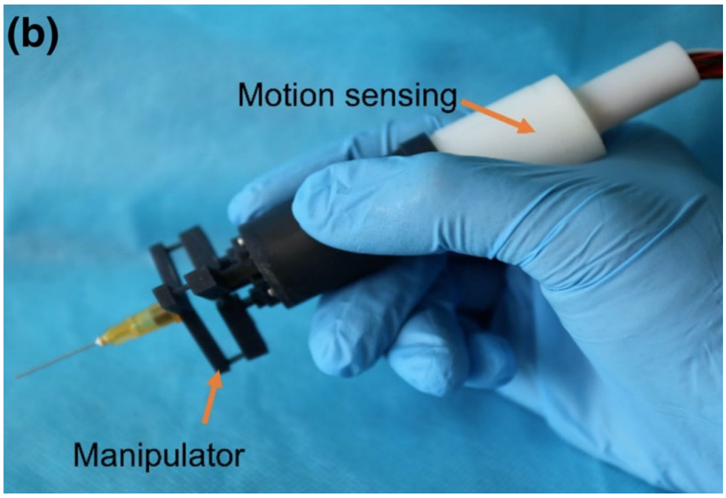
(a) (b)

Figure 2.7: ***Micron Application.*** *(a) Updated version of Micron with a novel delta kinematic for eye surgery and (b) it's performance of tremor suppression during cannulation [14, 15].*

Important for safe vitreoretinal surgery is not only the ability of precise motion but also the force applied to delicate tissue. Therefore, Gonenc et al. [55] added a control mechanism that limits the tool-to-tissue interaction forces with real-time capabilities that helps to prevent irreversible tissue damage. Subsequent studies [130–132], however, found that in some cases the device was difficult to use for ophthalmic surgery due to the limits of the sensing workspace, which resulted from the fact that it relies on external measuring systems which require an unobstructed sight.



(a)

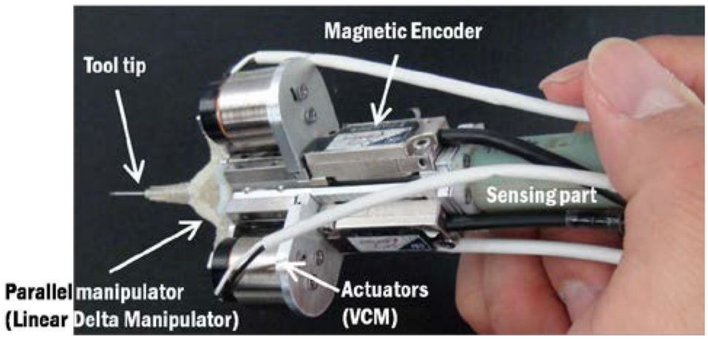


(b)

Figure 2.8: Microsurgical instrument with high accuracy actuated tip and integrated motion sensing for tremor suppression [248].

ITrem [115] is another microsurgical instrument with incorporated active tremor suppression. The compact sensing design is based on accelerometers in the handle so that external tracking is no longer required, as shown in Figure 2.9a. Recently, Chang et al. [32] and Zhang et al. [248] combined an improved version of the delta kinematic from the Micron system with the idea of a handle-integrated sensing, which is based on a high-performance inertia measurement unit (see Figure 2.9b and 2.8). The result is a high precision instrument for tremor suppression in a compact design.

(a) ***ITrem*** *by Latt et al. [115], a slimmed version of Micron.*

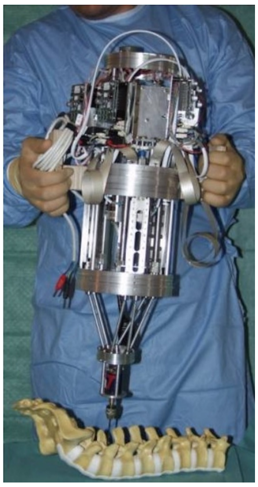


(b) Surgical instrument proposed by Chang et al. [32].

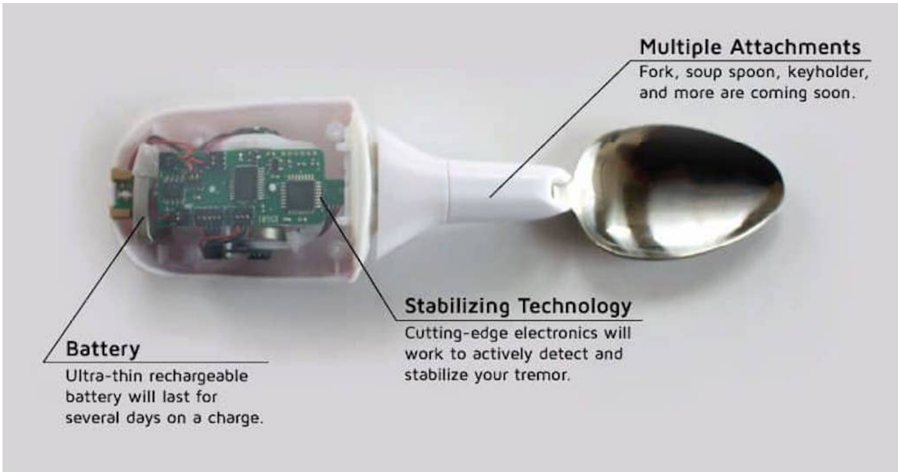
Figure 2.9: ***Tremor cancelling surgical instrument*** *with built in position sensing to overcome weaknesses of navigation solutions based on external tracking systems.*

Works on the compensation of unintended motion are not restricted to microsurgery. Wagner and colleagues [172, 197, 229] present a handheld tool for orthopaedic surgery. It is operated two-handed and the trajectory is stabilised using an optical tracking system for sensing and a 6-DoF parallel kinematics (shown in Figure 2.10a) for the positioning of the tooltip. The fast and accurate position corrections help to control the relative position between the drill and the bone. Note that this method goes beyond the suppression of larger-scale tremor as the system holds knowledge about a pre-planned trajectory which enables active tooltip control to follow that path.

Liftware [146, 148, 189] is an example for a tremor suppression device outside surgery (see Figure 2.10b). A set of eating utensils can be snapped into a mechatronic handle that measures and cancels high frequent hand motion. This allows patients with Parkinson's disease or essential tremor to overcome their otherwise severely limited ability to eat while keeping track of the development of their symptoms.



(a)



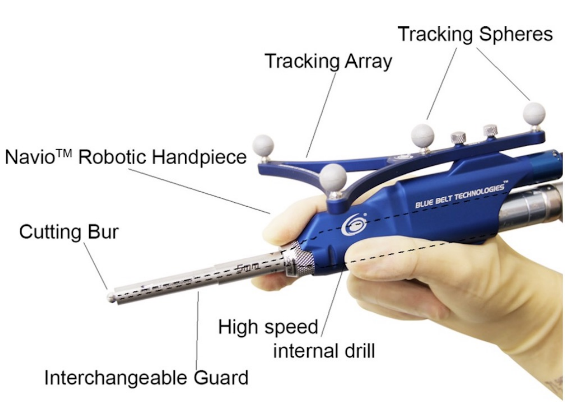
(b)

Figure 2.10: ***Devices for the Suppression of Unintended Motion.*** *(a) A handheld robot for orthopedic surgery [197]. (b) Liftware, a feeding utensil to compensate Parkinson-related or Essential Tremor.*

2.3.2 Guidance Active Avoidance in Surgery

In surgery, there is a strong demand for high accuracy movement since small deviations from an ideal trajectory can result in irreversible damage of tissue. Therefore, intelligent handheld tools were investigated to assist the surgeon through the planning phase, follow a pre-defined trajectory or to avoid specific features.

Navio™ [30, 128, 241], shown in Figure 2.11a, is one of the most advanced devices of its kind today. It is a precision freehand sculpturing tool which allows surgeons to detect landmarks of bones and tissue to construct 3D models which are then used to plan and execute a trajectory for substance removal. A similar device [25] was introduced before, however, it lagged the feature of online mapping and instead required insertions of markers in the bone. This is also true for the *Craniostar* [92], a cutting device that helps surgeons follow a pre-planned trajectory for a cutting procedure of the skull surface as demonstrated

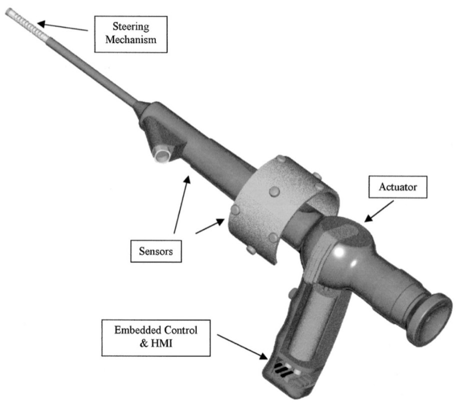


(a) ***Navio***^{TM} *sculpting tool, Lonner et al. [128]*

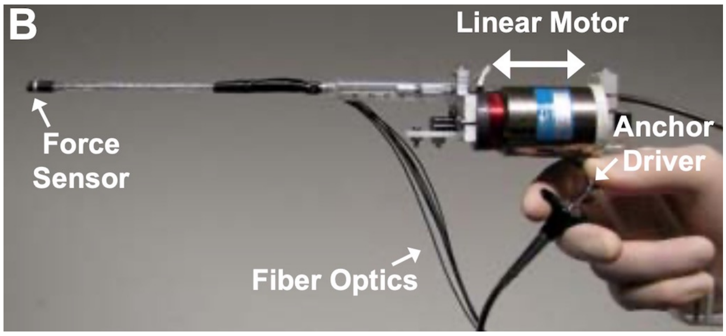
(b) ***Craniostar*** *device for skull cutting operations, Kane et al. [92]*

Figure 2.11: ***Medical intelligent Sculpting Tools.*** *This shows modern surgical instruments for cutting along predefined 3D profiles based on optical motion tracking and multi-DoF semi-autonomous tool tip control.*

in Figure 2.11b. All of those devices share the property of active guidance, based on an infrared optical tracking system for localisation and navigation.



(a) Instrument for arthroscopy, Dario et al. [35].



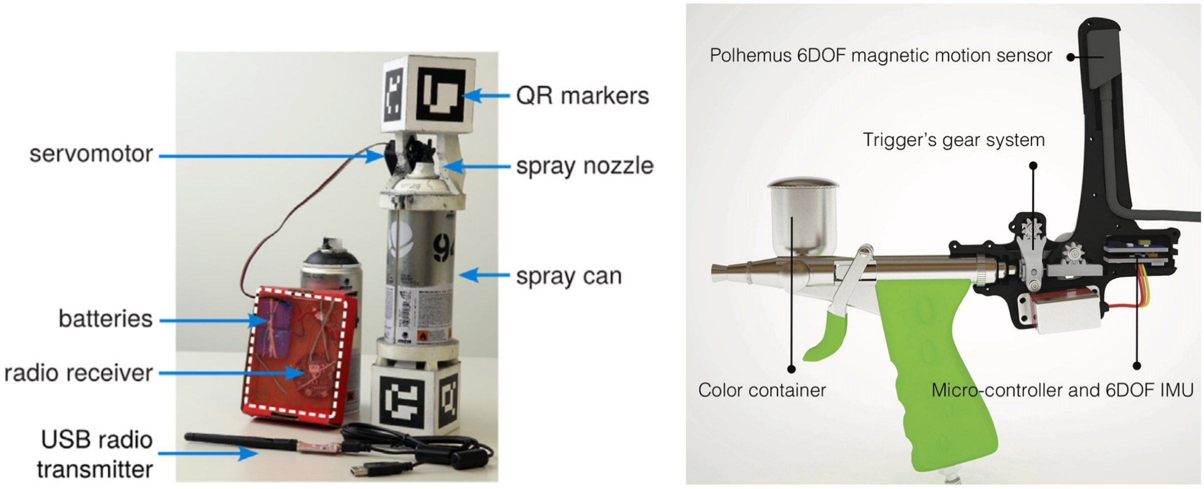
(b) Surgical tool for beating heart operations, Yuen et al. [244].

Figure 2.12: Mechatronic surgical instruments with integrated force limitation to prevent damage of tissue.

Other instruments provide active avoidance instead of guidance. They provide the surgeon with a free workspace that is constrained by safety-relevant features such as force limits. For example, Dario et al. [35][36] introduce a minimally invasive handheld tool, which limits the contact forces between the tooltip using a strain gauge as a sensor and a Hall effect angle sensor for the 1-DoF tip control. Similarly, Yuen et al. [244, 245] present a tool with 1-DoF actuation that maintains a constant distance to tissue during beating heart surgery, which helps reducing contact forces by 75% compared to using a manual tool. The designs of both tools are displayed in Figure 2.12.

2.3.3 Non-Medical Intelligent Handheld Tools

While extensive work on intelligent instruments was carried out in the field of surgery, the application of handheld smart tools outside the scope of medicine is rather new. Apart from the handheld robots mentioned in Section 2.2, there are only a few examples, which are mostly from the area of intelligent crafting and augmented arts.



(a) Air Brush by Prévost et al. [176] *(b) Air Brush by Shilkrot et al. [202]*

Figure 2.13: ***Intelligent Air Brushes*** *with localisation and pose-dependent valve control for assisted large scale painting.*

(a) Painting result, Prévost et al. [176] *(b) Target image [176, 202]* *(c) Painting result, Shilkrot et al. [202]*

Figure 2.14: ***Performance Comparison of Painting Devices.*** *This shows example results of intelligent painting experiments.*

Prévost et al. [176] present an example of an intelligent handheld airbrush (see Figure 2.13). The combination of optical motion tracking and active control of the airbrush's valve allows control of paint application as a function of the position of the tool. This allows novice users to produce large scale wall paintings from a pre-defined 2D image after a short time of training. The system is aware of the task progress and can feedback under-painted areas to the user through a separate work station. Similar to the evolution of surgical tools, Shilkrot et al. [202] introduced another intelligent airbrush that has on-board localisation through a Magnetic Motion Tracking System (MMTS), as shown in Figure 2.13b. This allows for higher accuracy in paint application and eliminates common problems of external tracking systems such as occlusion and the necessity of a large physical setup and calibration. Figure 2.14 shows a comparison of painting results using the two different airbrushes based on a given target image.

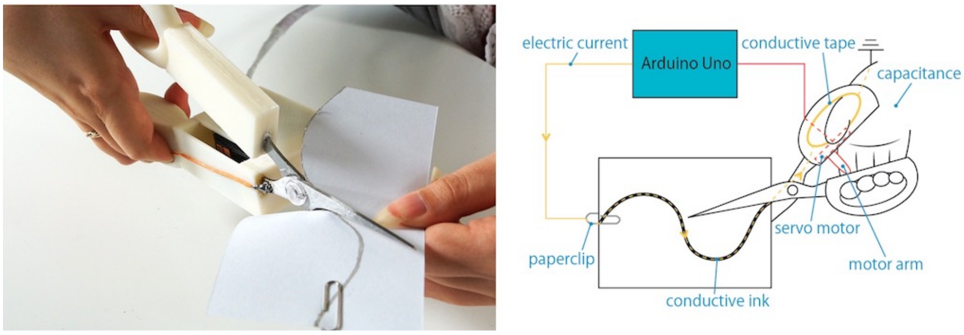


*Figure 2.15: **Enhanced Scissors** (left) only cuts in pre-defined areas. The system overview (right) shows the design principle, which based on conductive ink [240].*

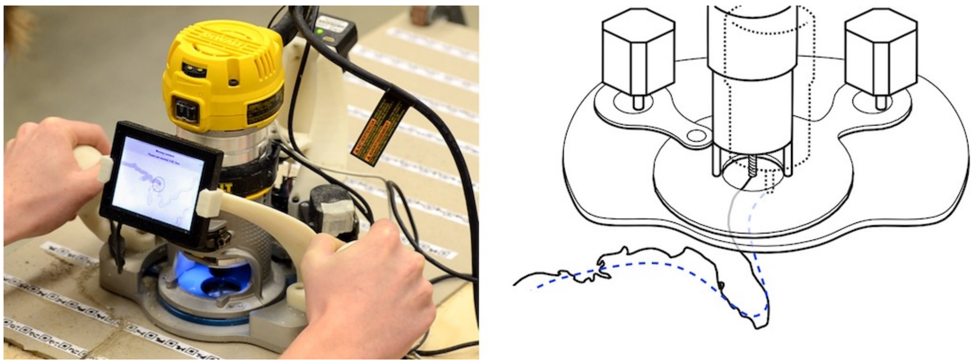
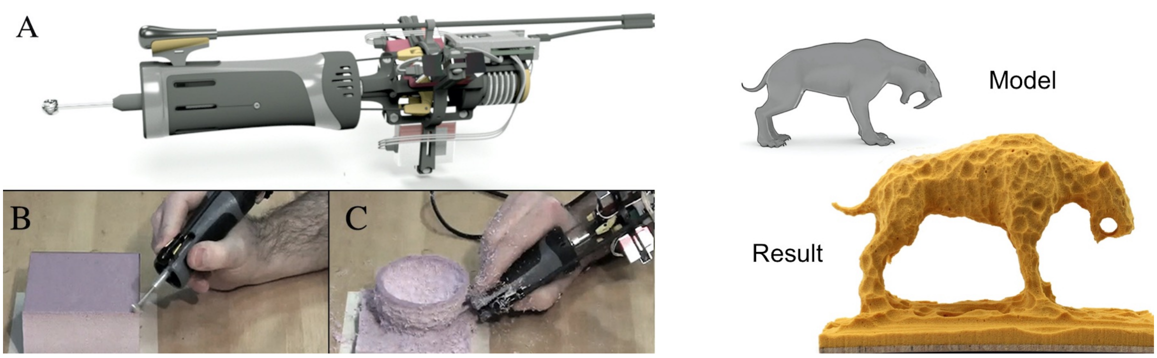


*Figure 2.16: **Position Correction 2D Milling Tool** (left) with integrated actuation to follow a complex profile while the user leads along an approximation of the path (right) [183].*

The concept of enabling or disabling a tool depending on its pose and position can also be found in the area of crafting. Yamashita et al. [240] introduced a pair of scissors that only cut in an area that was predefined with conductive ink. An overview of the system can be seen in Figure 2.15. In the field of fabrication, there are milling tools that help to align the cutting head through an active motion control to conform with a trajectory to the shape of an underlying Computer-Aided Design (CAD) model. For example Rivers et al. [183] present a 2D milling device that corrects a user's motion to match details of a cutting profile using an onboard camera for localisation (see Figure 2.16).

Zoran and Paradiso [249] transferred this concept to 3D sculpting. Their semi-autonomous handheld *FreeD* localises through a MMTS and incorporates a 3-DoF tooltip control, which allows for freehand sculpting as shown in Figure 2.17. In [250, 251], the authors extend the system's interactive capabilities and add a feature, which allows for online customisation of the underlying mesh model.



(a) *(b)*

Figure 2.17: ***FreeD Sculpting Device****. (a) Overview and example of usage [249]. (b) Mesh model and sculpting result, Zoran et al. [250].*

While the above-mentioned tools mainly assist the user through actuation, the *Intelligent Welding Gun* [47] for vehicle assembly is an example of a non-actuated device that assists users through visual feedback, which is shown on a display that is mounted on the welding gun. Attached markers allow outside-in tracking to determine the relative position of the tool to the work part. The device holds knowledge, which is used to lead welders to the right location and to improve their positioning of the stud as demonstrated in Figure 2.18. On-display animations such as *compass* and *notch and bead* serve as navigation guidance. The authors report a faster and more precise task completion of welders using the intelligent welding gun compared to a regular one. This example shows the importance of guidance features for effective collaboration and communication of the systems task plan. This subject will be covered more in-depth in Section 2.7.

Figure 2.18: ***Intelligent Welding Gun*** *in operation. Reflective markers enable localisation and a mounted display delivers feedback to the welder Echtler et al. [47].*

2.4 Wearables

Wearable robots are closely related to handheld robots as they share the characteristics of proximity and physical dependency to the user. Gregg-Smith [57] suggests that the handheld taxonomy fills the gap between external robots and wearables as illustrated in Figure 2.19.

External robots are characterised by their user-independent interaction with their environment and a high level of autonomy. They might be able to interact with humans as well but can be left without supervision for some tasks. Robots of this kind are for example used in search and rescue applications [126] and service robotics [40].

Opposed to that, wearable robots such as exoskeletons [155] and soft exosuits [39] enhance humans' body functions, which requires synchronous movement and thus makes both agents highly co-dependent. Typically, such devices are fully controlled by the user as they follow their lead and thus possess a limited level of autonomy.

This review focuses on wearables that are attached to the human body but have a local independent workspace, which excludes other technologies such as exo-suits and actuated prosthetics. The interested reader is referred to recent extensive reviews on powered prosthetics [114] as well as on upper and lower limb exoskeleton systems [116, 201]. Instead, this section summarises work on SRL and a worn gripper.

(a) (b) (c)

*Figure 2.19: **Handheld Robots Bridging Traditional Robotic Concepts**. A new taxonomy of personal robotic platforms (b) bridging traditional applications of external robots (a) and wearable devices (c) [57].*

2.4.1 Supernumerary Robotic Limbs (SRL)

SRL are a class of wearables that are distinct from traditional technologies such as exoskeletons and prosthetics as instead of substituting or enhancing, they aim to add effectors, for example, extra legs, arms or fingers, to the human body.

A prototype of SRL was first introduced by Parietti and Asada [166] within their work on a set of two robotic arms, which are attached to a human's hips. They help to compensate for involuntary motion and provide bracing in an assembly task. In subsequent work [165, 167], the authors present an updated version of the robotic system, which is assessed through an aircraft assembly task as an example application. The focus of theses works is on optimising the robot's bracing functions. Unlike exoskeletons, SRL can take on a variety of postures to optimise the effect of load-bearing. Similarly, Llorens-Bonilla and Asada [127] introduce a design that is shoulder worn and helps workers supporting assembly parts above their head. Moving from manufacturing to the medical field, Parietti et al. [168] later introduced a set of hip-mounted extra legs, which provide balance augmentation for bipedal walking by reducing the weight load on the human's legs. An overview of torso-mounted SRL can be seen in Figure 2.20.

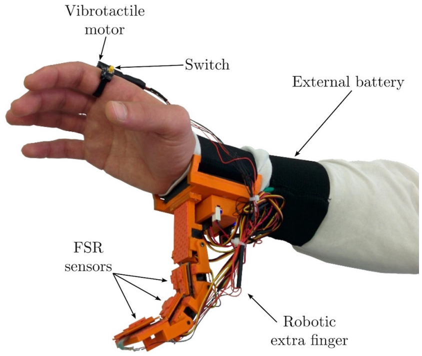
(a) (b) (c) (d)

*Figure 2.20: **Supernumerary Robot Limbs (SRL).** Concepts and prototypes for manufacturing and medical use. (a) Concept of SRL proposed by Parietti et al. [167]. (b) Hip-mounted SRL for assembly, Parietti and Asada [165]. (c) SRL for balance augmentation, Parietti et al. [168]. (d) Shoulder-worn SRL for over-head tasks, Llorens-Bonilla and Asada [127].*

Furthermore, SRL technology was explored for the case of additional robotic fingers that are mounted to the hand wrist. Wu and Asada [236, 237] introduced a prototype of this kind in their work on two extra fingers that extend the human hand (see Figure 2.21a). This enhances users' grasping capabilities through an enlarged workspace [236] and helps in completing bi-manual tasks one-handed, which is useful for people with impaired motor functions. Similarly, Prattichizzo et al. [173, 174, 175] introduced a 4-DoF robotic *Sixth-Finger* that helps users grasping large objects. A prototype can be seen in Figure 2.21b.

Later, Hussain et al. [76, 77, 79] refined the design of the Sixth-Finger and added haptic

(a) Wu and Asada [237]



(b) Hussain et al. [76]

(c) Malvezzi et al. [135]

*Figure 2.21: **Supernumerary Fingers.** An Overview of recent prototypes in the field of robotic extra fingers.*

feedback, which allows users to estimate the magnitude of force load on the device. This improved task performance and users' perceived effectiveness. This technology was then tested with stroke patients who suffer from lower limb motor dysfunction [78]. With the haptic feedback device on the healthy hand, the extra finger enabled subjects to execute bi-manual manipulation tasks, which they could not complete without the robot [78]. Another design iteration led to a cable-driven design of the *Soft-SixthFinger* [80, 83], which can be controlled through an Electromyography (EMG)-based interface [80, 81]. The authors' experiments show that this technology helps to compensate for missing grasping capabilities in chronic stroke patients [82, 84, 85]. Recent research in this field by Malvezzi et al. [135] concerns a multi-finger design, which allows grasping of objects with complex geometries. Figure 2.21 shows the technological evolution of supernumerary robot fingers.

2.4.2 Wearable Robotic Arms

Building on the aforementioned SRL ideas and works, Vatsal and Hoffman [221–223] introduced a wearable supernumerary robotic forearm. Their work is highly related to the handheld robot used for our studies, as their design has a similar workspace and shares physical dependency of the user's arms. Its base is attached to the human's elbow and its 5-DoF design carries a 1-DoF gripper, which makes it similar to a human forearm. The authors suggest that this technology falls in between torso-mounted [167] and wrist-mounted [76] SRL in terms of weight, power and scale.

Their concept aims to enhance wearers' reaching extent and supports work processes through bracing, stabilisation and reduction of cognitive load. Vatsal and Hoffman [224] validated their design in simulations and experimental user studies and found that the

(a) First prototype of the wearable forearm by Vatsal and Hoffman [221]

(b) examples of self-handing, object stabilisation and assisted human-human cooperation using the wearable robot [224]

Figure 2.22: ***Wearable Robotic Forearm****. An elbow-mounted supernumerary arm that cooperates with humans in augmented task completion.*

robot can increase a user's workspace by 246%. Furthermore, they investigated the HRI aspect of the robot and found that the robot's autonomy increases task efficiency in test cases of fetching an object while the human's hands are occupied and during object stabilisation.

These findings go in line with results by Gregg-Smith and Mayol-Cuevas [58] concerning the autonomy of handheld robots. Notably, Vatsal and Hoffman [224] also found that the robot's autonomous control is preferred over voice-commands. However, similar to Gregg-Smith and Mayol-Cuevas [58], they suggest that the interaction between an autonomous robot and a wearer requires further studies.

In their most recent work Vatsal and Hoffman [225] introduced a prediction system that helps to infer users' involuntary motion 77 ms ahead, which allows its suppression for end effector stabilisation resulting in an increase of accuracy of 20.1%. However, the presented approach only works for small movements and not for larger ones such as turning towards a specific object. One could argue that these large-scale movements result from the human's task decisions and that predicting those could help planning the robot's trajectory in augmented task completion. An overview of the prototype and its applications can be seen in Figure 2.22.

Similar to the elbow-mounted design presented above, Veronneau et al. [226] introduced a teleoperated robotic arm that is attached to the user's waist. Its characteristics are a 3-DoF lightweight design with a 3-fingered soft gripper as a manipulator. The device is similar to a human arm concerning shape and mass and driven through magnetorheological

Figure 2.23: ***Waist-Mounted Extra Arm****. This shows a number of example tasks used to test the robot: fruit picking (a,b), painting (c) window cleaning (d), assistive power tool handling (e,f), hands-free badminton [226]*

.

clutches and hydrostatic transmission lines. The current design depends on a tethered external power unit, which is located on the ground. This is due to its weight being too high to be carried comfortably (4.2 kg), which limits the mobility of the system.

The robot has high dynamic characteristics with an end effector top velocity of 3.4 m/s. At the same time, its strength is sufficient to hold industrial hand tools. The authors demonstrate its application for various tasks, such as picking fruits, painting, fetching and holding of hand tools and badminton playing (Figure 2.23).

Notably, the control of the device lies with a remote user who steers it through a joystick. During the execution of the presented experiment tasks, the controlling user is located at the worksite and has full visual access to the scene. As such, remote control by a third person is set in place to demonstrate the hardware rather than to explore interactions between the two humans through the robot. Nonetheless, the setup sets an example of how a user-carried robot can bring together the capabilities of two people working together towards common goals.

The work demonstrates that robot's and humans can benefit from each other's abilities when working together in close proximity, which underlines the importance of further exploration of this field. Similar to Vatsal and Hoffman's [221–223] robot forearm design, the mechanical characteristics of Veronneau's device are at an advanced stage and suffice for requirements of many applications. New challenges mainly concern the design of the interaction between robot and human to bring the system closer to the vision of human-robot synergy. Thus the question remains open, to what extent a robot of this type could help in a task based on autonomous motion and which user cues could be used to inform the robot's decision to conform with the user's intent.

2.5 Mixed-Initiative Interaction

Over the last decades, interaction between humans and machines evolved from rudimentary physical interaction with a mechanical apparatus to sophisticated interaction with robotic tools. These are characterised by increased complexity through higher integration of sensors and sophisticated control systems [129]. This development led to a need for new concepts of humans and machines working together towards a common goal. Hearst [67] argues, that an important aspect of an efficient multiagent collaboration setup was the implementation of a *mixed-initiative interaction* concept, i.e. a flexible interaction strategy that enables each agent to contribute to the task with what it does best. The subdomains of mixed-initiative interaction that matter most in the context of this work on handheld robots are teleoperation, shared control, traded control and teaming. Here is a brief description of each, listed in order of increasing level of autonomy:

- **Teleoperation:** The robot is controlled by a human through a remote interface [105]. Traditionally, robots do not have any autonomy in such setups as the user is fully in charge of the robot's motion.

- **Shared Control:** The human and the robot manage the task simultaneously [163] through a fusion of operator inputs and autonomy inputs [217], which requires arbitration, i.e. a mechanism that decides to what extent each agent has control. For example, the human might be in charge of the main body of a tool while the robot is in control of fine-tuning the motion of the end effector [129].

- **Traded Control:** The human and the robot take turns in controlling a robotic system so that each agent is in charge of a specific part of the task [105]. Usually, the human takes over control to solve a difficult subtask or navigate the robot out of a dangerous situation such as in devices for disaster response [158]. Afterwards, the control is handed back to the robot. For intelligent tools, roles for these tasks are often swapped so that the user navigates to a target where the robot then completes a task.

- **Teaming:** This domain describes the collaboration of two independent agents. Each contributes to the task autonomously, however, they communicate their plans to coordinate their actions [105]. Due to the physical proximity between users and tools, the two agents cannot be considered as completely independent agents. However, they can be independent concerning decision-making, therefore coordinated planning is an important aspect in this context.

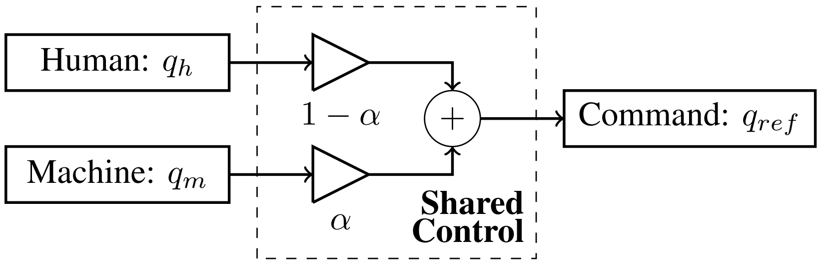


Figure 2.24: ***Arbitration.*** *The control ration between the human input $\boldsymbol{q}_h$ and the machine input $\boldsymbol{q}_m$ is balanced to create a desired output $\boldsymbol{q}_{ref}$ with the robot's autonomy level α as a parameter [163].*

The above domains are not strictly separated and can at times be combined in a single application. For example Nudehi et al. [159] describe a control system that merges teleoperation with shared control. A surgical robot is simultaneously remote-controlled by an expert and a novice and their input is weighted according to their experience level before being passed on to the robot's control unit. Alternatively, a system could switch between the aforementioned domains through arbitration. Owan et al. [163] suggest an approach to merge human input $\boldsymbol{q}_h$ with the input of the machine $\boldsymbol{q}_m$ to create a desired output $\boldsymbol{q}_{ref}$, given the level of autonomy $\alpha \in [0, 1]$ as follows:

$$\boldsymbol{q}_{ref} = \alpha \boldsymbol{q}_m + (1 - \alpha)\boldsymbol{q}_h, \tag{2.1}$$

which is illustrated in Figure 2.24. Note that the parameter α can be used to tune the control ratio in a shared control setup and can be dynamically changed to transition between the modes of mixed-initiative interaction.

In summary, the rise of more complex human-machine interaction scenarios comes with a demand for collaboration concepts. The spectrum between teleoperation and shared control defines the design space which we are exploring in this research.

2.6 Intention Prediction

The demand for prediction of user intention by robots traditionally stems from the field of safe human-robot interaction. However, in recent years technological progress enabled robots to become autonomous partners in collaborative setups [110, 220]. For a robot to assist effectively, a channel to communicate a user's intent is required for anticipating their next move [71, 72]. Traditionally, this has been implemented through explicit commands such as keyboard inputs or voice commands [73]. However, fluent collaboration requires multimodal implicit means of communication as it occurs in human-human collaboration. For example, humans observe each other's gaze to coordinate manual tasks in teamwork settings [230]. In the context of assistive robots, the equivalent of such intuition is a system for intention prediction.

Concerning tasks involving the handheld robot, we define intention as the user's choice which specific object they want to interact with next. With this in mind, we review recent work on intention inference where different methods can be distinguished by the predictor that was used as a basis for the respective prediction systems, mainly body motion, Electroencephalography (EEG) signals and gaze data.

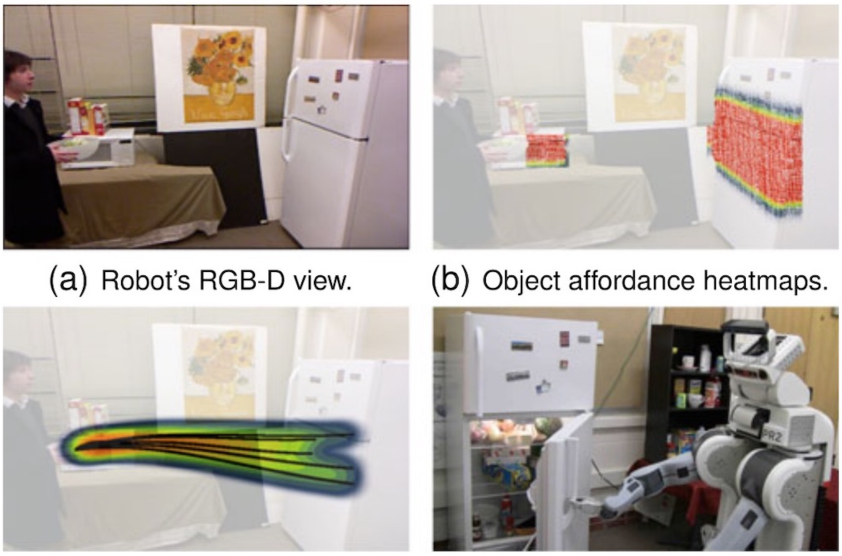
2.6.1 Motion as an Intention Predictor

Early work on human action prediction concerning manual tasks uses marker-based wrist tracking to estimate sub-sequences of an assembly task [118]. The authors used the hands' velocity, acceleration and jerk as predictors to train a Hidden Markov Model (HMM) and achieved 92.26% accuracy in predicting basic classes of movements such as reaching out, retraction and grasping. Similarly, Mainprice et al. [133] use upper limbs pose data from optical motion tracking to predict future reaching movements in one-handed collaborative tasks using inverse optimal control and iterative re-planning.

Ravichandar and Dani [178, 179] investigated intention inference based on human full-body motion. They used a generic reaching task and a Microsoft Kinect for data collection, which was then used to train an ANN to predict future reaching locations. The model allows for accurate predictions within an anticipation time of approximately 0.5 s prior to the hand touching the object. Later, they added human eye gaze tracking to their system and used the additional data for pre-filtering to merge it with an existing motion-based model [180], which helped increase the anticipation time to 0.78 s.

Similarly, Zunino et al. [252, 253] use onset motion, i.e. the beginning of an action to predict an underlying intention using an ANN, which was trained based on 3D motion tracking data from a Vicon system and 2D video data. Remarkably, their classification system discriminates the actions *passing*, *placing*, *pouring* and *drinking* with an accuracy of 80.5% even though they all start with similar onset, i.e. moving a bottle.

Furthermore, Koppula and colleagues [90, 103, 104] use body motion for prediction, which is captured through an RGB-D camera. Their approach takes into account a predicted task's motion-based affordance, which allows for action predictions with 84.1%/74.4% accuracy 1 s/3 s in advance, respectively. Figure 2.25 shows an example of an action prediction and how a service robot might use this information to assist with a task. Building on these contributions, Dutta et al. developed a probabilistic model through



(a) Robot's RGB-D view. (b) Object affordance heatmaps.

(c) Trajectories heatmap. (d) Robot opening the door.

*Figure 2.25: **Human Action Anticipation for Service Robotics**. The approach is based on using an RGB-D as an input (a) to model associated action affordance (b) which serves as a basis for the anticipation of a trajectory probability (c). In consequence, the robot augments anticipated actions (d) [103].*

their recent series of work [43–46] to predict human actions. It is based on a description of object-affordance, that is the spatiotemporal relationship between a human and an object. Together with the human's latent body pose, this allows the prediction of motion goals and the construction of a probability heat map that represents possible future hand trajectories. Their concept was validated through experiments with basic actions such as reaching, moving, pouring and drinking. In their most recent work [46], their motion-based prediction model yields an accuracy of up to 93.02% for a variety of human activities.

2.6.2 Brain Computer Interfaces

Thanks to recent advances in neuroscience, sensor technology and methods for data analysis, a variety of neural methods is now available to measure intent information [149]. The growing field of Brain-Computer Interface (BCI) aims to extract electrophysiological signals from neurons and use them as a basis for a communication link between the human brain and an external device [4]. Due to technological progress concerning real-time temporal resolution, EEG attracted researchers' attention with regards to implementing BCI systems to derive cognitive functions, such as decision making, attention and motion planning [228]. Put short, an EEG setup is an array of electrodes that measures regional brain activity through the detection of fluctuating neural electric potential [20].

Different EEG methods vary with respect to the electrodes' proximity to the neurons. Sarac et al. [192] used surface EEG to detect user intent through electrodes, which were arranged noninvasively on the scalp. The EEG signals are decoded using a linear discriminant analysis to infer patients' motor planning in a rehabilitation robot setup. Within their work about intention inference for prosthetics, McMullen et al. [144] use so-called electrocorticography. This is a form of EEG where the electrodes get placed on the surface of the brain through a surgical procedure. The electrodes' proximity to the brain allows for a higher resolution of the EEG signal. Using this BCI as an additional input for the prosthesis controller, patients were able to perform grasping tasks more accurately.

Note that this kind of intent inference is at an early stage [4]. While existing solutions are useful to derive a proxy of human motor intent, more neuroscientific knowledge is required to develop BCIs that allow for intention prediction on a task planning level, i.e. as envisioned for the handheld robot.

2.6.3 Inferring Attention and Action Intent from Gaze

Literature on using eye and/or head gaze for intention estimation is vast and spans many decades. Standard methods to detect gaze directions commonly involve remote [154] or head-worn [137] eye trackers or determining head orientation for example through 2D cameras [117, 208]. The information about both head orientation and eye gaze has been linked to a person's focus of attention and intention in the past [152, 161]. Here, we distinguish between *attention* as a person's current object of involvement and *intention* as the plan of future interaction.

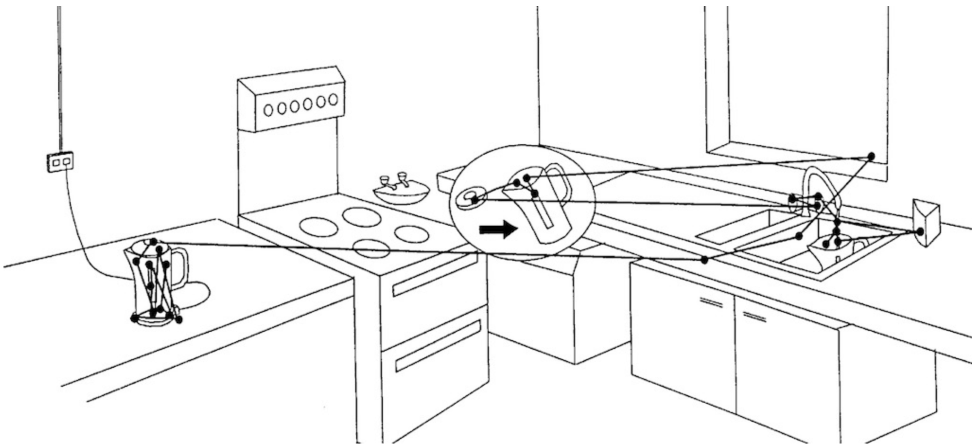


*Figure 2.26: **Fixations During Every-Day Tasks**. This figure shows the scene of a task with an example of an according series of fixations of a participant during task execution. The fixations correspond in time to the actions that are associated with the fixated objects [112].*

Fixations are the short times between saccades, where the gaze rests at an object as a visual scene is explored and usually have a duration of a few hundred milliseconds [187, 191]. Land et al. [112, 113] found that there is a close relationship between gaze fixations and manual actions as the eye gaze precedes interactions. They conducted an experiment where subjects completed every-day tasks such as making a tea as demonstrated in Figure 2.26. The results yield that the fixation sequence had a time shift of around 500 ms relative to the object interaction sequence. The fact that this relationship can be found in *automated* tasks suggests that this phenomenon of unconscious attention might occur in other life situations as well and might thus reveal underlying planning activities.

This goes in line with studies on hand-eye coordination where subjects were asked to prepare a sandwich from a given set of ingredients [64] or to wash their hands [170]. Both indicate that object fixations during task completion do not depend as much on salience as on its functional relevance to the task. Furthermore, they found that while most fixations were associated with the current sub-task, a small fraction of around 5% was dedicated to future objects during task-switching for instance right before another ingredient was picked by subjects during the sandwich-making task. Pelz and Canosa [170] concluded that the planning of movements is done a few seconds prior to actions.

Similarly, Mennie et al. [145] studied human gaze behaviour during an assembly task where they linked object fixations to task planning. The authors suggest that the human eye gathers visual information about a task object trough *look-ahead* fixations right before interaction rather than relying on longer-term visual memory when it comes to task planning.

Beyond the planning of motion, gaze fixations are also linked to the process of making decisions about a task. In experiments with virtual [12] and real [169] block design tasks, participants were asked to copy a given block pattern. One might expect that the subjects would look at the pattern and memorise it before replication. However, their gaze behaviour was characterised by repeated saccades towards the model during the assembly process, which, according to the authors supports a *just-in-time* planning process.

With the development of more accurate systems for head and eye gaze tracking, researchers started using this data to infer human intention. The studies by Yi and Ballard [243] is an example of early work, where eye gaze is used as a feature to recognise human task behaviour in a sandwich-making task using a dynamic Bayes network. However, the authors do not indicate that this model could be used to make predictions of future actions. Furthermore, both head and gaze data were used to predict the behaviour of car drivers in traffic [41]. The authors use a sparse Bayesian learning model [143] trained on gaze data of a time window of 3 s prior to lane changes to predict driver's decisions. While their work demonstrates the importance of gaze cues for intention prediction, presumably, their model does not generalise well for the case of cooperative handheld robots as head-turning movements were identified as the major predictor. Arguably, the occurrence of these movements is rather specific to the domain of car driving and less relevant to interaction with intelligent tools, where users mainly direct their head towards the tooltip.

In recent work, Koochaki and Najafizadeh [100] present a vision-based approach to intent prediction that is solely based on eye gaze data in every-day tasks. Their algorithm first extracts the human's regions of interests in the visual field by identifying clusters of fixations. In a subsequent step, the objects in these image regions are identified using a Convolutional Neural Network (CNN). Amount and temporal properties of the fixations associated with the object are then used as features for an SVM-based task classification. The model performs with 95.68% accuracy. In subsequent work [101], the authors change the model and take into account the sequence of selected objects to predict tasks using a Long-Short Term Memory (LSTM), which yields 82.27% average accuracy across tasks. The authors suggest that the prediction model could be used as an input for motion planners of assistive devices.

While the aforementioned works by Koochaki and Najafizadeh demonstrate the potential of gaze-based early intent prediction, they come with two major limitations. Firstly, the eye gaze data was not recorded during actual task execution. Instead, experiment subjects were asked to look at relevant objects in a given 2D image after being told a task for ground truth. However, it is unclear whether the subjects' gaze behaviour is the same as it would be during task execution. Secondly, the authors do not report to what extent their method allows for predictions in the future, i.e. the model's accuracy as a function of the point in time of task anticipation.

The studies by Huang et al. [75] are of particular interest for our research as their approach aims for intention prediction in collaborative setups rather than focusing on one person only, like in the works mentioned above. Furthermore, their SVM-based prediction model only uses eye gaze data as input. The data is derived from a customer-worker setup where the customer chooses a set of sandwich ingredients through verbal requests while their gaze is tracked using a head-mounted gaze tracker. In contrast to [100, 101], their prediction approach is bottom-up, i.e. the position and identity of the objects in the scene are known through visual tracking. This allows the construction of features such as gazing duration and number of glances for individual objects, which are then fed to the SVM's input. Through this method, the authors achieved the prediction of ingredient selection with an accuracy of 76% and correct predictions around 1.8 s prior to the verbal request.

In subsequent work, Huang and Mutlu [74] used the model as an input to an assistive robot's path planner (Figure 2.27). The predictions were used for *anticipatory control* in a collaborative pick and place task. The robot approached objects with increasing likelihood of their selection through the human, which improved completion time compared to following verbal commands only. While it is unclear whether this model would

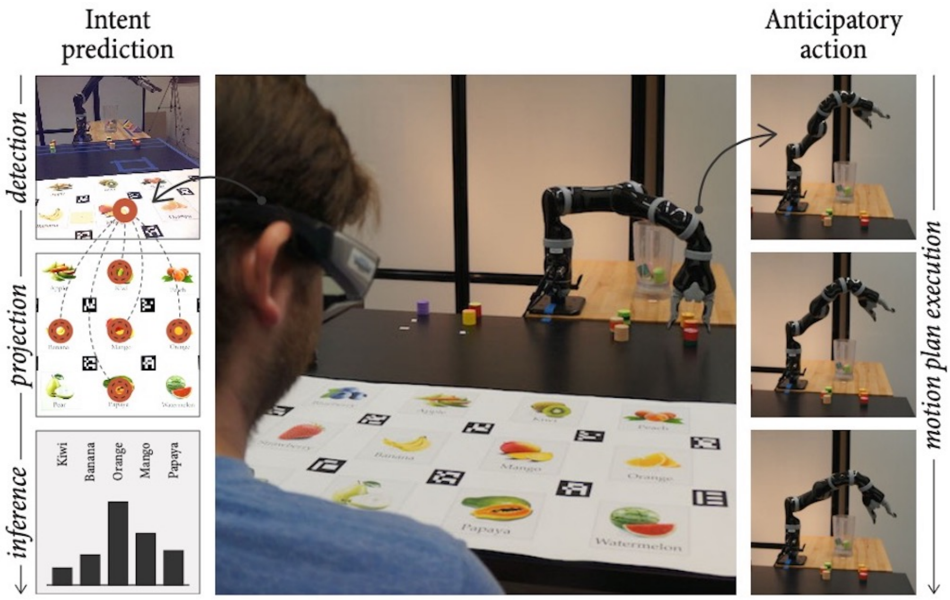


Figure 2.27: ***Anticipatory Robot Control.*** *The robot derives user intention from eye gaze and uses this information to plan proactive actions within a collaborative task Huang and Mutlu [74].*

be feasible in a shared-control setup such as we face with handheld robot applications, it demonstrates that it is suitable for cooperative task solving and enables the robot to accelerate task completion through proactive actions.

2.6.4 Summary of Methods for Intention Prediction

This literature review shows that intention prediction has been studied extensively in recent years with a broad variety of modalities, deriving human intent from motion, gaze or BCIs. At the same time, the use of intention models for the control of anticipatory robots is at an early stage and has not yet been explored for setups with close proximity between the user and the robot, i.e. such as we face in handheld robot setups. We suggest that eye gaze is a promising predictor for the use in our work as it is accessible (easier to detect than signals for BCIs) and not constrained by the setup itself (e.g. body motion is tight to the robot's position). For this reason, the reader shall also be introduced to the mathematical representation of the user's eye gaze in the following section.

2.6.5 Mathematical Representation of the Eye Gaze

The analysis of eye gaze for intention prediction requires a mathematical representation of the gaze. Therefore, the reader shall be exposed to a brief excursion about the definition of a ray and how it can be described with respect to related coordinate frames (e.g. the eye, a robot and the world) using homogeneous transformation. The remainder of this section is based on the mathematical theories described in [24].

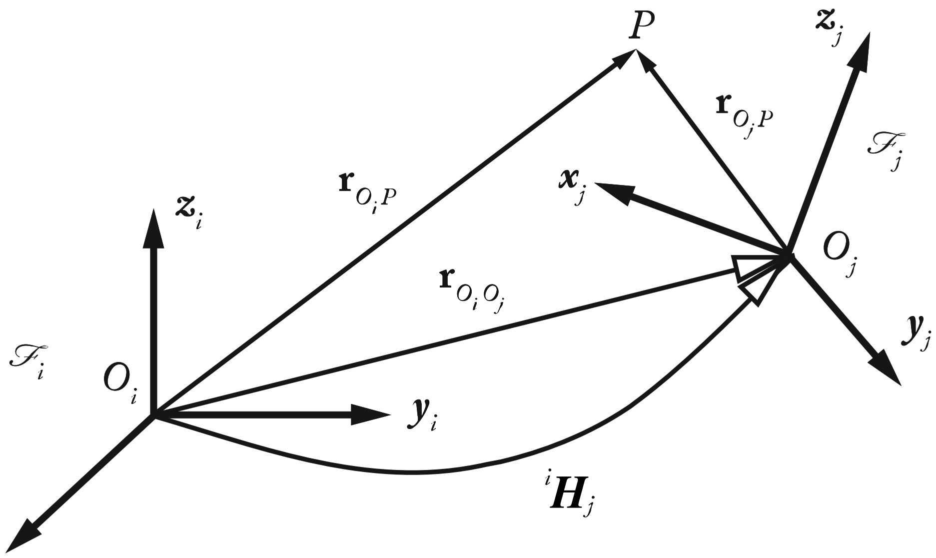


Figure 2.28: ***Homogeneous Transformation.*** *Illustration of the transformation of a vector with respect to two relative frames [24].*

Let P be location and ${}^i\boldsymbol{x}$ a vector that points to P, represented in the coordinate frame $\mathcal{F}_i$. Then P can be represented in coordinate frame $\mathcal{F}_j$ through ${}^j\boldsymbol{x}$ (see Figure 2.28). The relationship between ${}^i\boldsymbol{x}$ and ${}^j\boldsymbol{x}$ can be described using the homogeneous transform matrix ${}^j\boldsymbol{H}_i$:

$$\boldsymbol{x}_i = {}^i\boldsymbol{H}_j\,\boldsymbol{x}_j. \tag{2.2}$$

The transformation matrix $\boldsymbol{H}$ is a 4×4 matrix with the following elements:

$$\boldsymbol{H} = \begin{pmatrix} & \boldsymbol{R} & & \boldsymbol{t} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{2.3}$$

where $\boldsymbol{R}$ is a 3×3 rotation matrix and $\boldsymbol{t}$ the translation vector. $\boldsymbol{R}$ can be constructed through various ways, e.g. roll-pitch-yaw angles or Euler angles.

${}^i\boldsymbol{H}_j$ can be calculated if the transformations of $\mathcal{F}_i$ and $\mathcal{F}_j$ are known with respect to a fixed world frame $\mathcal{F}_w$. The relationship between ${}^w\boldsymbol{H}_i$ and ${}^w\boldsymbol{H}_j$ is:

$${}^w\boldsymbol{H}_j = {}^w\boldsymbol{H}_i\,{}^i\boldsymbol{H}_j, \tag{2.4}$$

hence,

$${}^i\boldsymbol{H}_j = ({}^w\boldsymbol{H}_i)^{-1}\,{}^w\boldsymbol{H}_j. \tag{2.5}$$

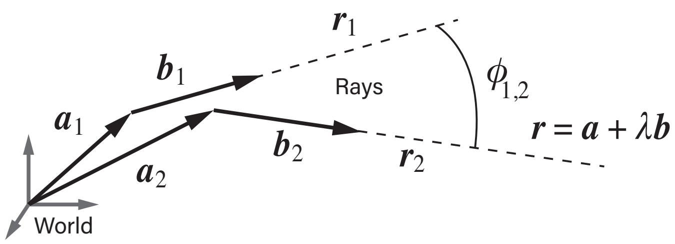


Figure 2.29: ***Comparison of two Rays.*** *Illustration of the mathematical representation of a ray which is defined by its origin* $\boldsymbol{a}$ *and its direction* $\boldsymbol{b}$*. Two rays* $\boldsymbol{r}_1, \boldsymbol{r}_2$ *can be compared with respect to the difference of their directions* $\phi_{1,2}$*.*

An eye gaze can be represented using the definition of a ray $\boldsymbol{r} \in \mathbb{R}^3$ through the equation [26]:

$$\boldsymbol{r}(\lambda) = \boldsymbol{a} + \lambda \boldsymbol{b} \quad \text{with } \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3, \lambda \in \mathbb{R}, \tag{2.6}$$

where $\boldsymbol{a}$ is a vector to the ray's origin and $\boldsymbol{b}$ the orientation. The angle $\phi_{1,2}$ between two rays $\boldsymbol{r}_1, \boldsymbol{r}_2$ (e.g. as illustrated in Figure 2.29) is equal to the angle between the according direction vectors $\boldsymbol{b}_1, \boldsymbol{b}_2$:

$$cos\phi_{1,2} = \frac{\boldsymbol{b}_1 \cdot \boldsymbol{b}_2}{|\boldsymbol{b}_1| \cdot |\boldsymbol{b}_2|}. \tag{2.7}$$

If the gaze ray $\boldsymbol{r}$ is known in a local frame, e.g. of an eye tracking device, then it can be transformed to world coordinates using the aforementioned transformation theory.

2.7 Remote Guidance

A collaborative remote assistance system is usually based on three parties: an unskilled local worker, a skilled remote expert and a device that mediates information in both directions and commands from the expert to the local worker. In this work, we demonstrate how the handheld robot can serve the purpose of a remote collaboration device (Chapter 5). A major novelty is the physical access of the remote user to the worksite through the remote control of the robot's mechanism in a human-robot-human setup. Alongside traditional communication, e.g. voice and vision, this can be used for guidance and manipulation. Therefore, the related literature draws from the field of remote guidance and telemanipulation.

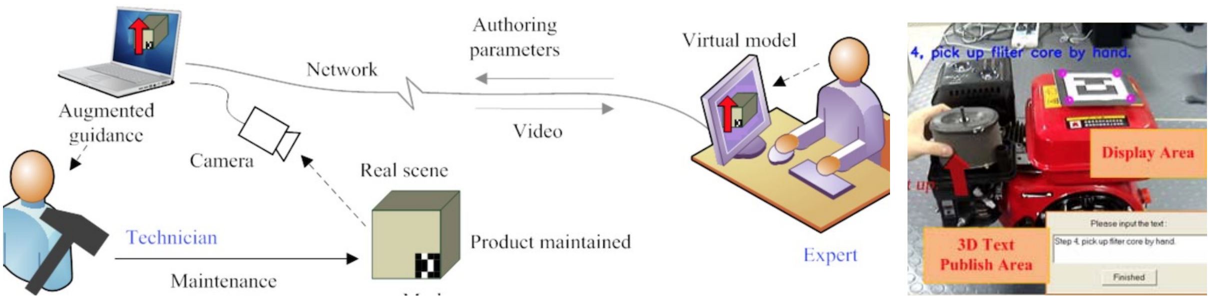
Remote guidance systems allow a remotely located person to assist a local person through instructions and directions. Due to its relevance to industries, e.g. for maintenance and training, extensive research has been carried out in this field in recent years, which aims at overcoming the limitations of traditional consulting methods such as audio or video calls [106]. Current solutions to communicate the local conditions of the task state require a camera at the worker's site, which is either stationary [69, 121, 162, 177, 231], portable (e.g. smartphone or tablet) [21, 53, 54, 140, 206], worn by the local user [22, 69, 186] or operated by the helper [61, 107–109]. These solutions have in common that the video feed of the local workspace is displayed to the remote helper either on a desktop screen, tablet or smartphone alongside audio communication. However, the methods are distinct in their respective communication features for the helper's instructions to the local user, which are outlined in the following.

2.7.1 Instructions Through Video Markups

In early work, Ou et al. [162] introduced video markups in addition to traditional video consulting. They describe a remote assistance system that includes a camera on the worker's side, that is remotely controlled by a consulted specialist. The specialist can use predefined forms and free-hand pen drawings to overlay the video stream with annotations for the worker. These annotations are then transmitted to a screen at the worker's side. Their preliminary test results with physical tasks show that using the annotation system improves the performance over the use of video-only systems. The authors suggest that this could be used to consult a specialist in surgery (Figure 2.30). The problem with video-markups is that annotations become meaningless when the camera is moved. This problem was address in subsequent research involving AR solutions.

Figure 2.30: ***Remote Guidance in Surgery****. Pen drawing video-overlay during specialist consultation [162]*

For example, Wang et al. [231] introduced a system for maintenance, where the remote user can make annotations to the product, which is equipped with a QR-tag. This allows the remote expert to create 3D animations and to label specific parts for detailed instructions. The system is object-specific as the system requires a model of the maintenance object at the remote site as a prerequisite. This is useful for complex machinery, with recurrent problems or maintenance task with user support through the manufacturer, i.e. in the place of a manual for the technician. However, it does not work for objects that are large (e.g. car maintenance) or that are new to the system.



(a) (b)

*Figure 2.31: **Product Support through Telemaintenance.** The expert remotely instructs a technician (a) with annotations and animations using a video overlay tool (b) [231].*

For this reason, recent works [21, 53, 54, 140] focused on remote maintenance systems with AR features for unstructured environments. The local technician is equipped with a handheld screen, e.g. a smartphone or tablet computer (Figure 2.32). It displays the camera image to which a remote expert can add annotations to. These are world-stabilised with respect to the work scene, i.e. annotations stick with objects in the environment even if the camera gets moved. The advantage of these systems is that they are based on widely available consumer hardware and do not require any physical pre-setup at the technician side, e.g. for tracking. However, as opposed to aforementioned solutions with static cameras, the remote user is unable to move the camera and is thus fully dependent on the technician following their instructions in the inspection and diagnosis part of the collaborative maintenance task. In addition, a separate display can potentially distract the technician from the real work site and occupies an otherwise free hand.

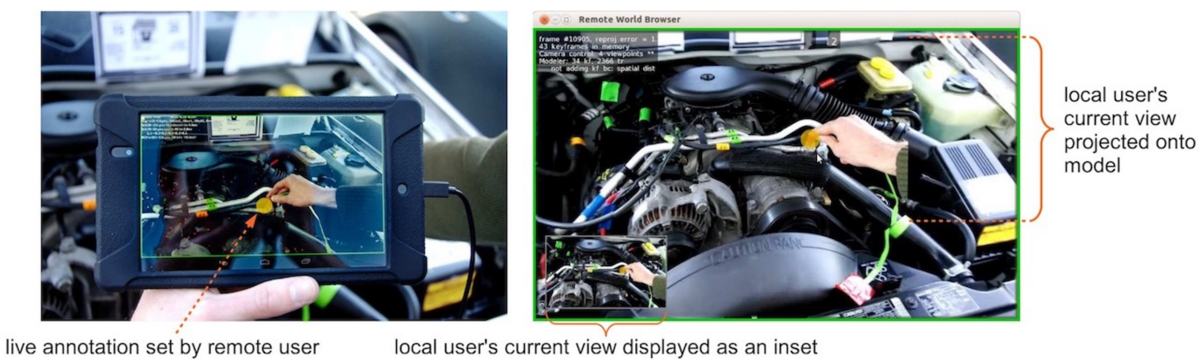


*Figure 2.32: **Handheld AR System for Remote Maintenance.** The device records the environment for a remotely located expert, who assists through world-stabilised annotations [54].*

2.7.2 AR Headsets in Remote Guidance

The idea of free-hands instruction display was implemented in recent works [22, 151], where instructions are displayed through an Head Mounted Display (HMD). For example, in recent work in this field Mourtzis et al. [151] propose a remote maintenance system where a technician creates a malfunction report, which is sent to an expert, who then compiles a list of predefined AR instructions. The technician then works through the instructions using an AR headset. Note that this solution allows for free-hand operation and detailed instruction. However, expert consultation is an offline process. Hence, this is useful for complex maintenance for special cases, i.e. in a customer support scenario.

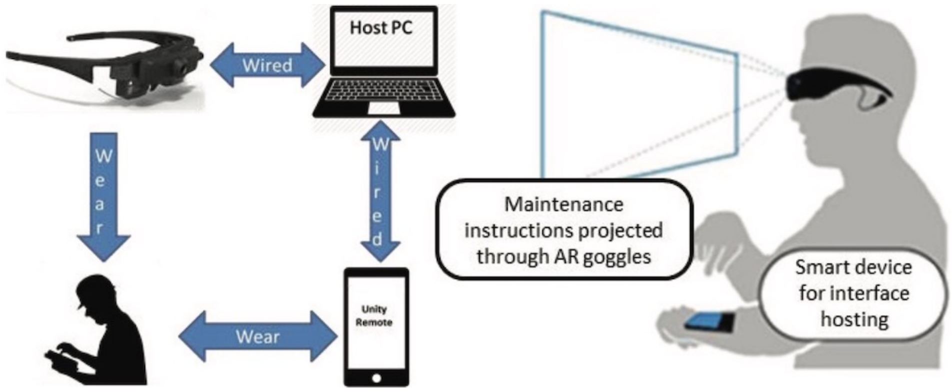


Figure 2.33: ***AR Maintenance Instruction System.*** *The illustration shows the hardware setup for the local technician introduced by [151].*

Tait and Billinghurst [213] experimented with a remote collaboration setup where the remote users' view is independent of the local user's actions. The technician wears an AR headset, which transmits a 3D model of the environment to the remote user and allows an independent scene exploration on a 2D screen. They can place annotations in the 3D model which become visible for the local worker through the AR headset. Within their experiments, the authors tested the setup with varying degrees of independence of the remote user's view to study its effect on collaborative performance. The results show that higher levels of independence increase remote users' confidence while less verbal communication was required to complete the task. Subsequent studies on the effect of view independence on video-mediated remote collaboration supported these results [97].

2.7.3 360° Cameras and VR for Telepresence

Following the idea of giving the remote user independent control, recent works [94, 141, 215] explored a new body-ghost paradigm (Figure 2.34). Cameras with 360° view angle allow to record the environment and enable a remote user to observe it with a perspective of their choice. Matsuda and Rekimoto [141] combined a 360° camera with a mobile platform as a telepresence system. Similarly, Tang et al. [215] describe a system where a 360° camera is mounted on the local user's backpack while the remote user accesses the video stream through a tablet. Notably, Kasahara et al. [94] present a system where the local user wears a 360° recording device around the head which transmits to a remote user's VR environment, this allowing for a natural view control through tracking of their head angle.

Similarly, Linn et al. [121] developed a telepresence solution purposed for remote maintenance of machinery. A 360° camera is placed in the machine casing, which allows a remote technician to perform an inspection and check for defects. Notably, the authors suggest that the system could be improved through mounting the camera on a robot tip to give the remote user a higher degree of control over the view perspective. To the best of our knowledge, the question of how this could be implemented in a collaborative setup, e.g. in a human-robot-human scenario is yet unanswered.

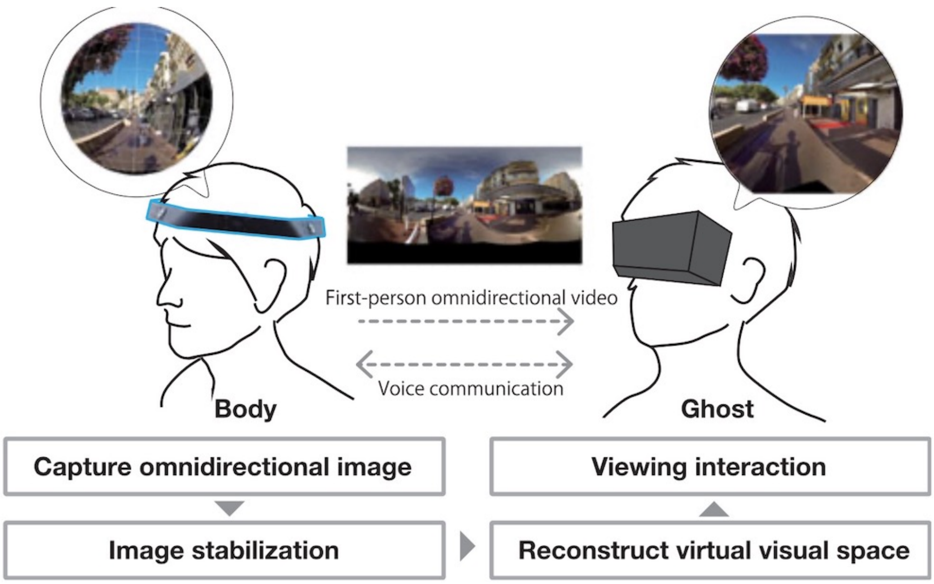


Figure 2.34: ***The Body-Ghost Paradigm.*** *The Local user is equipped with a 360° recording device. The remote user (ghost) can choose the view angle for independent inspection of the environment.*

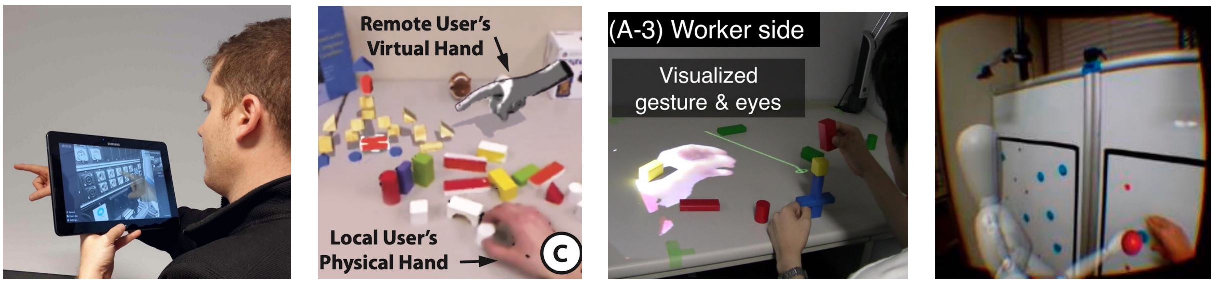
Note that the above mentioned 360° vision solutions do not feature any visual feedback to the local user or require expensive calibration procedures as is the case with Linn et al. [121]. Previously, Gurevich et al. [61] presented *TeleAdvisor*, a remote assistance tool. Its main components are a camera and a pico projector, both mounted at a teleoperated robot arm. This allows a remote user to perceive the worksite environment and to project annotations on to it to assist in physical tasks. Kangas et al. [93] combined this idea with the aforementioned body-ghost paradigm for assisted maintenance. The local worker carries a 360° camera and a projector. The remote expert wears a VR headset and accesses the projector for annotations. Additionally, the projector is aligned with another camera for a detailed view of the local scene. Their user study shows that pointing is a useful feature in combination with verbal instructions. Similar results can be expected when using a handheld robot for remote pointing during remote assistance.

2.7.4 Guidance through Natural Gestures

A more natural approach to remote guidance is taken in works where the helper gives instructions via hand gestures [69, 186, 206], captured full-body motion [239] or through a robotic mechanism [108, 109].

Higuch et al. [69] experimented with a collaboration system where the remote user's hand gestures and gaze fixations are visualised in the worker's scene. The setup was tested for a static setup where a stationary projector was used for the visualisation and a mobile case, which involved a see-through HMD and a head-mounted camera. They found hand gestures and fixations were used in various ways to complement verbal instructions. Notably, in some cases, the local user was able to predict instructions after observing the remote user's fixations.

Robert et al. [186] introduce a remote guidance prototype *MobileHelper*, implemented for a tablet computer. It allows a remote expert to use hand gestures for instructions, to guide a local technician through various physical tasks. The technician can work hands-free as the gestures are visualised using an AR headset.



(a) Higuch et al. [69] *(b) Robert et al. [186]* *(c) Sodhi et al. [206]* *(d) Yamamoto et al. [239]*

*Figure 2.35: **Examples of Remote Guidance with Gesture Use**. Existing solutions range from in-screen overlays to scene projection and AR integration of a helper's pointing gestures.*

Sodhi et al. [206] present *BeThere* in a proof-of-concept study for future mobile technologies. Similar to the aforementioned work, the prototype allows the remote user to provide guidance through hand gestures. However, rather than transmitting a segmented picture of the expert's hand, the authors used a Kinect to capture the hand pose. The local user sees a hand animation in the work scene as an overlay in their phone screen. Their user study demonstrates that the system can be used for various physical collaboration tasks.

The idea of using gestures for interaction was further extended through work by Yamamoto et al. [239]. Their system captures an instructor's full-body motion. Using an AR headset, a 3D animation of the body pose is brought to the worker scene. The authors suggest that such a system has the advantage of covering much bigger workspaces than previous work that focused on hand gestures.

By comparison, less effort was spent on the exploration through embodied guidance, i.e. mediated through a robot. An example is *GestureMan* [108, 109], which is a mobile robot equipped with an actuated camera and pointing stick. However, the pointing mechanism is limited by a few DoF and thus not suitable for manipulation.

The above examples demonstrate that collaboration in remote guidance setups greatly benefits from deictic elements for multimodal instructions. The ability to highlight things in the work scene often replaces wordy instructions. While existing methods offer efficient solutions for the communication of instructions, they do not allow any direct physical interaction within a collaborative setup. This leaves local workers in charge of carrying out any manipulative actions by themselves.

2.8 Telemanipulation

In contrast to remote guidance systems, telemanipulation allows a remote operator to execute physical operations, i.e. through a remote-controlled robot rather than instructing another person. Early examples of such master-slave manipulation systems date back to 1945, when the company Central Research Laboratories[1] developed a remote manipulation device, which was purposed to handle highly radioactive goods (Figure 2.36).

Today's application examples exist in the form of sedentary robots for remote maintenance and inspection of machinery [16, 17], telerobotic surgical systems [62] such as *da Vinci* [11, 233]. Furthermore, remote manipulation is used for safe physical interaction in hazardous environments, e.g. the outer space [195], nuclear [139, 214] or fusion [204] power plants and for the handling of dangerous materials such as nuclear decommissioning [2]. In terms of mobile robots for telemanipulation, the most advanced examples have been explored in the context of disaster response [38, 65, 198] and exploration of the outer space [28, 87] and deep sea [232]. These systems enable a remote user to navigate through unstructured environments and manipulate physical objects.

The remainder of this section highlights the two areas of the field of telemanipulation that are most relevant to this work, i.e. semi-autonomous teleoperation (Section 2.8.1) and remote maintenance robots (Section 2.8.2). Furthermore, Section 2.8.3 describes initial

[1] www.centres.com

developments towards remote collaboration through wearable robots.

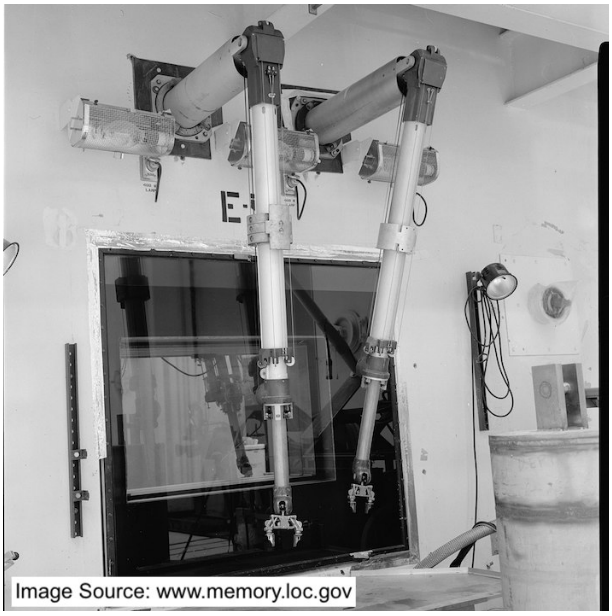
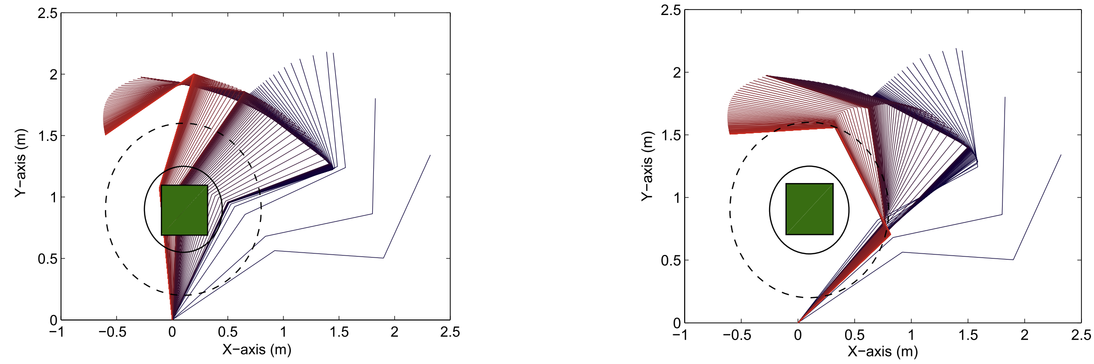


*Figure 2.36: **Master-Slave Manipulator Mk. 8 (MSM-8)**. This shows an early example of a remote manipulator developed by Central Research Laboraties in 1945.*

2.8.1 Semi-Autonomous Slave in Teleoperation

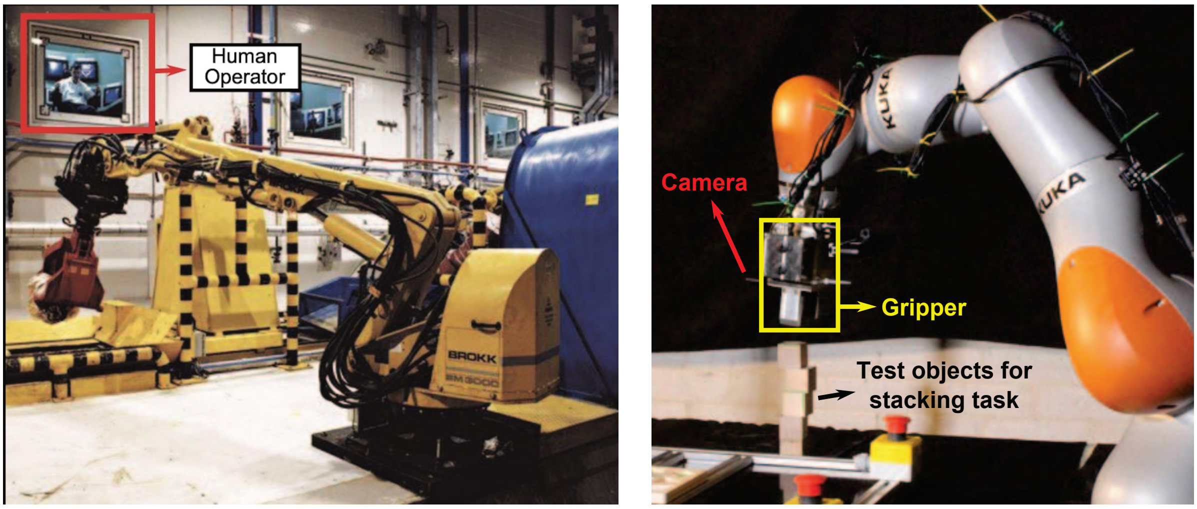
While telepresence has been studied extensively concerning control problems emerging from time delays [95] and the bilateral design of force feedback mechanism [153], there are few examples for semi-autonomous control in this area. Liu and Chopra [124, 125] present a method that allows the user to control a subset of a robot's DoFs while the remaining ones are autonomously controlled to deal with the subtasks of avoiding singularities, joint limits and collisions, which increases the manipulability of the slave-robot (see Figure 2.37). While these features make the control easier and more robust during usage, the system requires continuous supervision of the operator, i.e. the robot is unable to complete parts of the task that require task knowledge for high level decisions.

In contrast, the master-slave telemaintenance system introduced by Marturi et al. [139] demonstrates how operating a robot can be facilitated through the automation of local subtasks. *Brokk* is an industrial robot that was developed for teleoperated sorting and segregation of nuclear material (see Figure 2.38a). The authors experimented with a stacking task setup and a 7-DoF robot as a mockup (Figure 2.38b). Computer vision was used to assist in grasping and releasing of objects with the idea being that this method could later be transferred to the Brokk system. The operator manually navigates to a task object and then delegates the grasping operation through on-screen selection. The comparative user studies show that the semi-autonomous features reduce workload and cut down completion time.



(a) Configuration without sub-task control. *(b) Configuration with sub-task control.*

Figure 2.37: ***Semi-Autonomous Control of Redundant Slave Systems****. Illustration of numeric proof-of-concept study where the slave robot semi-autonomously controls redundant joints to avoids singularities, joint limits or obstacles (shown here) [125].*



(a) *(b)*

Figure 2.38: ***Semi-Autonomous Telemanipulation*** *[139]. (a) shows Brokk, a tele-operated robot for nuclear decommissioning that might benefit from autonomous features. (b) shows the testing setup for semi-autonomous stacking of blocks.*

2.8.2 Robots for Remote Maintenance

Sedentary [16, 216, 242, 247] and mobile [205] maintenance robots allow a remotely located operator to inspect, diagnose and interact with a physical structure at the work site.

In early work, Bellamine et al. [16, 17] introduce a remote maintenance system that allows for planning and executing operations (Figure 2.39). A robot arm is located at the worksite and equipped with a vibration sensor, which is used to detect cracks in the casing. The operator can observe the site through a camera from a third-person perspective. There are two operation modes for the robot. In continuous manual mode, the robot's motion is under control of the operator. In supervisory control mode, low-level tasks are executed autonomously, following the operator's high-level instructions, e.g. contact points at the

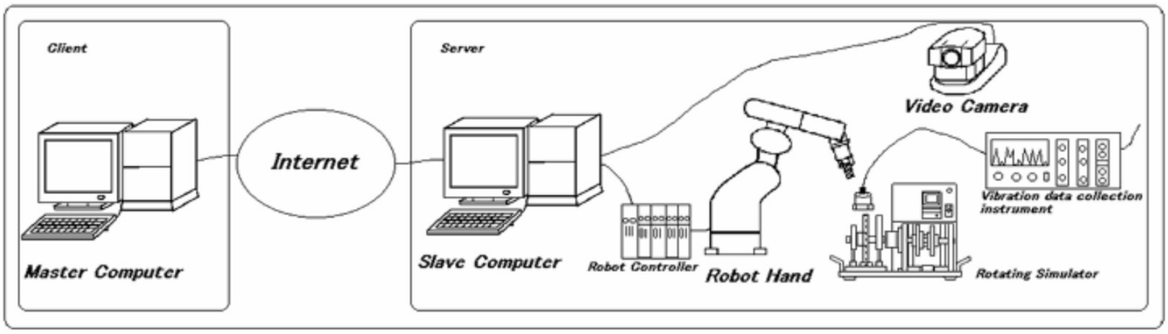


Figure 2.39: ***Master-Slave Setup for Maintenance.*** *Conceptual model of a remote maintenance system proposed in [17].*

machinery to place the sensor. At the time of its reporting, the system is at an early stage of development. The planning process is done offline using a 3D model and AR. The alignment of the 3D model to the real world requires a time-consuming manual matching procedure.

Similarly, Yew et al. [242] present a system for remote maintenance through an industrial robot at the worksite, which is equipped with a camera at the tip. As a novelty, they introduce AR-enhanced interaction between the operator and the robot. The components of the worksite are in part available as a physical copy at the operator's side and part of the AR environment otherwise (see Figure 2.40). The components are aligned with a virtual version of the robot, which is visualised in simulated part of the environments too. This enables more spacial awareness and an intuitive estimate of joint limits during robot control through a hand controller. The work is in its early stages, e.g. it neglects obstacles that could be in the way and that are not part of the AR system. However, using a plumbing maintenance task as an example, the authors demonstrate that novice users could pick up how to control the robot quickly due to the intuitive design of the interface.

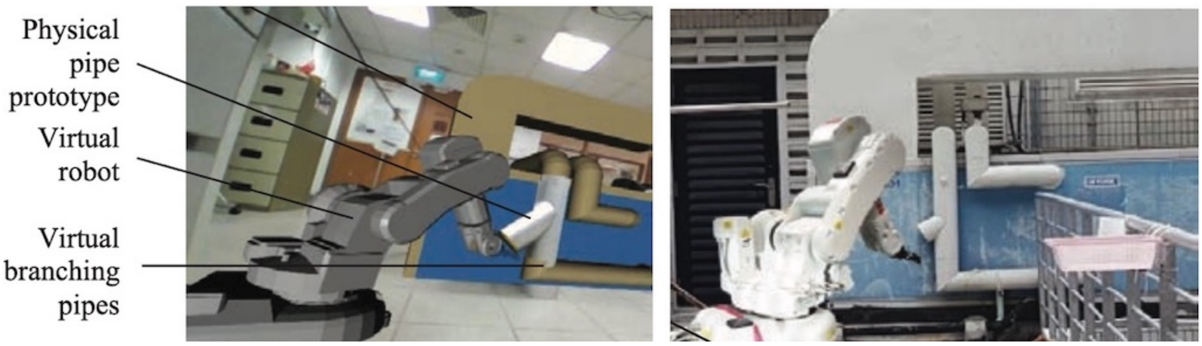
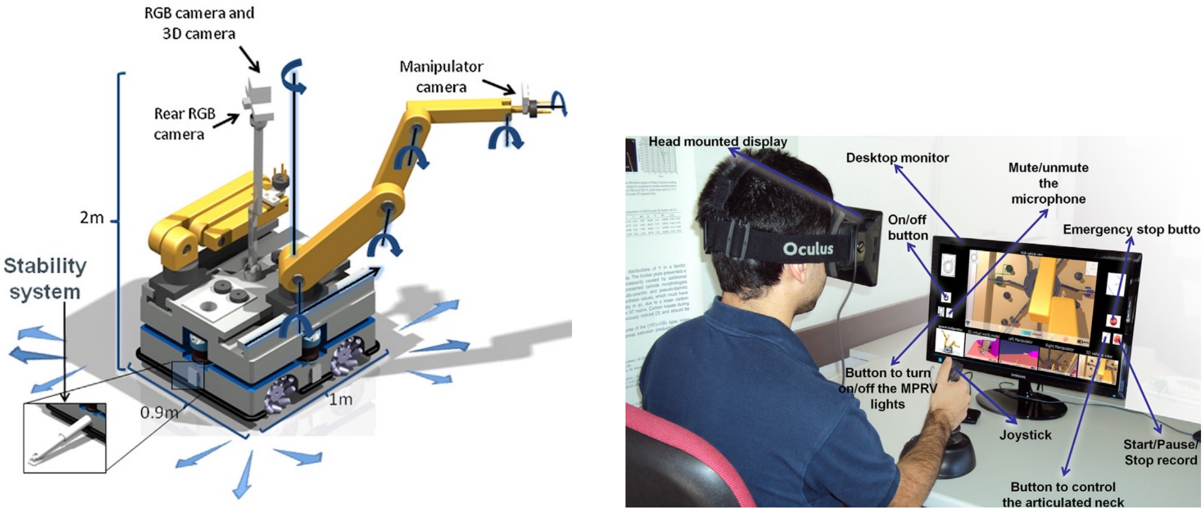


Figure 2.40: ***System for Remote Maintenance.*** *A mix of real and virtual components (left) allows for intuitive operation of the robot at the work site (right) [242].*

Today's advanced maintenance robots are purposed for applications in environments that are inaccessible to humans, such as industrial offshore platforms [216] and fusion reactors

[205, 247]. For example, in the context of the fusion power project *ITER* [52], remote handling of goods in hazardous environments is required as part of some maintenance procedures. Soares et al. [205] present a mobile multi-purpose robot for safe routine maintenance in the vacuum vessel of the power plant. The system has an omnidirectional mobile platform, equipped with several sensors and two robot arms (Figure 2.41a). Each has 6-DoF and is equipped with a camera for a detailed view during manipulation. A central camera allows for a wide perspective around the worksite. The operator accesses the robot through a workstation, which is equipped with a display and a joystick for manual control (Figure 2.41b).



(a) (b)

Figure 2.41: ***Mobile Maintenance Robot System****. CAD model of the robot platform (a) and an overview of the workstation for remote control (b) [205].*

Similarly, Zhang et al. [247] report about a robot that is highly specialised for inspection and maintenance of the fusion vessel of the aforementioned ITER reactor. The snake-like design of the robot has multiple DoF and a lightweight cable-driven mechanism. It is mounted to an industrial robot which positions and inserts the robot through a small opening in the vessel. The multiple DoFs allow the robot to follow complex paths while its body follows the tip. This prevents collision with parts of the power plant during the exploration of the vessel. The setup can be seen in Figure 2.42.

Existing solutions for telemanipulation are useful for environments that are inaccessible for humans and scenarios with highly specialised robots and operators (i.e. surgery). However, the research question of how remote guidance and telemanipulation could be combined in a collaborative setup remains unanswered.

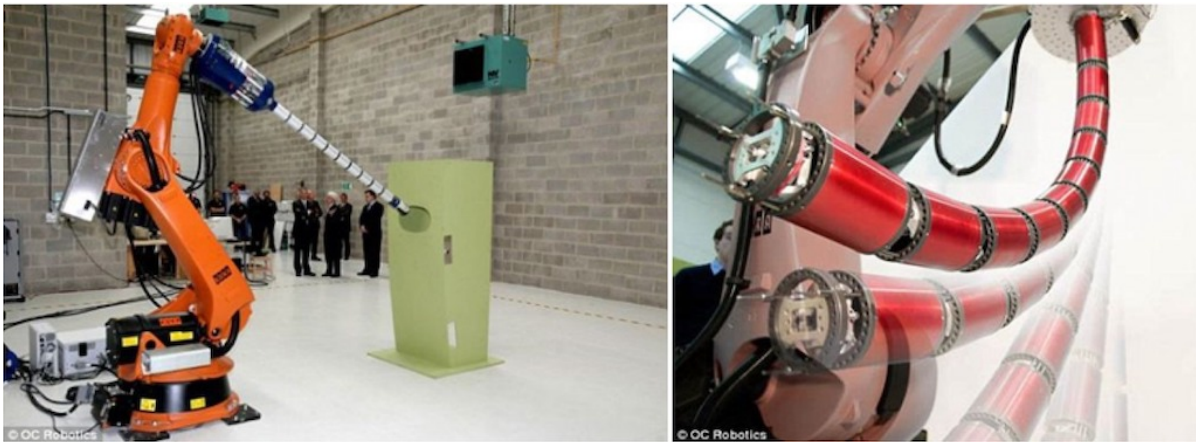


Figure 2.42: ***Snake-Like Robot****. Multi-DoF robot of OC Robotics for inspection and maintenance. [247].*

2.8.3 Teleoperation of Wearable Robots

A new approach in telerobotics is the concept of a teleoperated robot that is carried by another person, e.g. Veronneau et al. [226] present a simple remote control scheme for their waist-mounted robot (introduced in Section 2.4.1). While the remote control aspect of this system is in its early stages, it demonstrates the feasibility of remote collaboration through a robot and with the help of a proximate user as a carrier and helper.

Saraiji et al. [193] present a system with two remote-controlled robotic arms that are attached to a user. The arms can be controlled through one or several remote users with visual feedback provided through a camera system that is attached to the host. The remote control implementation is based on a VR system (Figure 2.43). The system is a first attempt of merging the capabilities and workspaces of several users. However, recent works present challenges in teleoperation, particularly in the domain of remote assistance, which need to be addressed in research that focuses on the teleoperator. For example, in the aforementioned designs, the remote cannot control the viewpoint of the cameras. Most importantly, the possibility of coupling teleoperation with semi-autonomous slave control (as per Section 2.8.1) is yet unexplored for the case where a robot is used in conjunction with a carrying host.

2.9 Summary

To summarise the usage of the background in the remainder of this thesis, Table 2.1 presents an overview of the reviewed material (in order of its presentation in this chapter) and links the specific work sections to the discussed topics.

The presented literature shows that great progress has been made over the past years

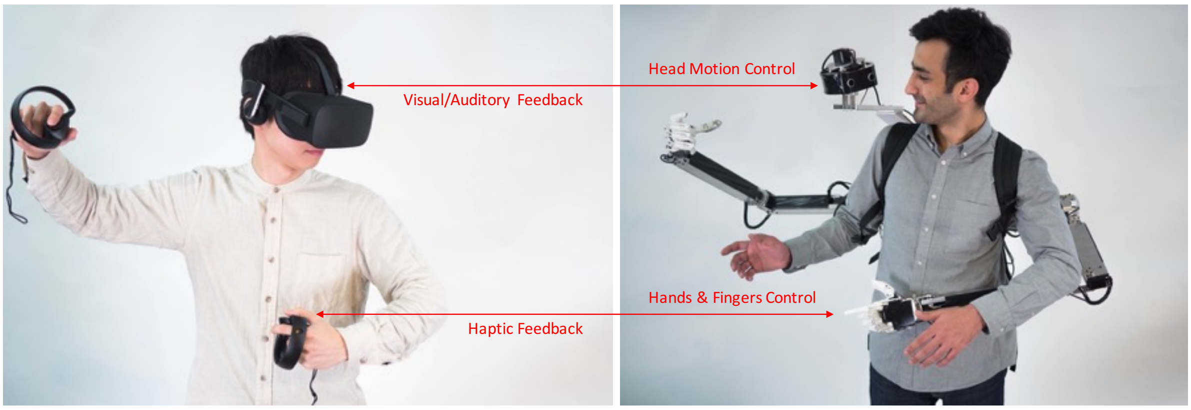


*Figure 2.43: **"Fusion"**. A bi-manual robot that is attached to a host (right) and controlled by a teleoperator (left) for collaborative tasks [193].*

concerning the development of intelligent tools that can use task knowledge for assistance. This is particularly true for highly specialised devices, e.g. surgery. With the recent introduction of handheld robots for a more general-purpose, the tools' assistance went beyond corrective movement and mistake prevention towards a system that participates in task decisions. Work in this field presents new challenges, e.g. we lack mechanisms that allow smart tools to incorporate users' plans in their decision processes. We argue that this needs to be addressed to bring this type of robots one step further: away from being purely assistive towards a collaborating agent that follows the paradigm of *working together*.

New technologies allow the use of a single device through multiple users and existing solutions demonstrate that teleoperation concepts are an effective approach to merge workspaces of humans and robots. There are only a few examples of robots that can be controlled whilst being hosted by another collaborating user, e.g. worn or carried, which allows for new applications due to increased mobility. These examples are based on master-slave concepts with direct control of every movement of the robots. However, we see a great potential regarding the robot as a separate agent that could act as an intelligent mediator between the collaborators. This requires the exploration of possibilities to bring together the abilities of the involved agents and to consider how their respective strengths could be combined in a complementary way.

Our work is a first step towards addressing the above challenges of human-robot interaction and human-robot-human interaction. With the aim for a better understanding of how humans could interact with handheld robots, we hope to make a contribution to the research community towards a future with collaborating tools.

Background Section	Background Topic / Specific Work Topic(s)	Specific Work Section(s)
2.2	**State-of-the-Art Handheld Robots**	
	Handheld Robot Hardware Adaptations and HRI Challenges	3/4/5
	Levels of Autonomy in Collaborative Task Solving	3.5.3/4.6.1/5.2.6
	Metrics for Collaboration Performance and Quality	3.6/4.7/5.4
2.3	**Intelligent Handheld Tools**	
	Compensation of Unintended Motion and Assisted Aiming	3.5/4.6/5.2
2.4	**Wearables**	
	Combining Attention/Intention with Task Knowledge for the Control of Physically Dependent robots	3.5/4.2.2
2.5	**Mixed-Initiative Interaction**	
	Shared Control	3.5/4.6
	Teleoperation and Traded Control	5.2
2.6	**Intention Prediction**	
	Gaze Ray Modelling in 3D Space	3.2
	User Attention from Eye Gaze	3.5.1
	Collection of Gaze Data for the Training Set	
	Gaze-Based Intention Modelling and Algorithm Selection	4.2.2
	Methods for Qualitative and Quantitative Validations of Intention Models	4.3/4.4
	Experiment Design for the Assessment of an Anticipatory Robot	4.6
2.7	**Remote Guidance**	
	Remote Assistance Study Design	5.2.1
	Remote Collaboration Setup	5.2.3
	The Maintenance Experiment Task	5.2.4
	Layout and Design of the Remote Workstation	5.2.5
	Trial Procedure and Data Collection	5.2.7
2.8	**Telemanipulation**	
	Camera Setup for Visual Feedback	5.2.3
	Layout and Design of the Remote Workstation	5.2.5
	Semi-Autonomous Completion of Subtasks	5.2.6

Table 2.1: ***Usage of Background in the Specific Work Sections***

Chapter 3

I Can See Your Aim: Estimating User Attention From Gaze For Handheld Robot Collaboration

In this chapter, we explore the estimation of handheld robot user attention, which is important for overcoming existing obstacles in human-robot interaction as mentioned in the preceding Chapter 2. The subsequent Chapter 4 will built on the attention model presented here as it concerns gaze-based intention prediction.The outcomes of this chapter are summarised in the supplementary video[1] (scan QR code).

The main results of this chapter were presented at the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) [209].

3.1 Introduction

The vision of a handheld robot is one of an intelligent power tool that can lead its carrier through a task by cognitive and mechanical augmentation (see detailed review in Section 2.2). That way, the use of such a robot could enable untrained or disabled people to perform tasks which they would otherwise not be able to achieve without it.

Advanced systems in the area of intelligent handheld tools demonstrate how robots can enhance users' mechanical abilities and help prevent mistakes or suppress unintended motion. For example, in the field of medical applications, some devices assist surgeons

[1]Chapter 3 Summary Video: https://youtu.be/lsQ4k71NLTA

to cut bone substance along pre-defined 3D shapes [92, 128] or suppress hand tremor in microsurgery [32, 248]. Tools in the field of fabrication and assembly can correct users' motion in assisted 2D [183] and 3D [249, 250] milling and an intelligent welding gun [47] can lead its users through a car manufacturing task. More examples of this kind are described and discussed in Section 2.3.

The above works yield little evidence of aiming to understand intention or user attention to enhance the cooperation between robot and user. Furthermore, the explored designs are highly specialised towards specific tasks. Their small workspaces limit their actions to corrections of human motion with few assistive DoFs. Arguably, a device that is a more efficient collaborator requires the ability to make task decisions and carry out more complex subtasks.

The notion of a general-purpose handheld robot with a larger workspace and more freedom concerning the selection of subtasks has been explored over recent years by Gregg-Smith and Mayol-Cuevas [58, 60] (see Section 2.2.1 and 2.2.2). They demonstrate how a robot's task knowledge can be used for augmentation even without explicit feedback to the user, e.g. via rudimentary robot gestures [58], in some instances combined with visual feedback [60]. While informing the user of the robot's intention significantly improved task performance, a remaining challenge was identified: the conflict between user intention and the robot's plans. This led to frustration in users and a negative impact on cooperative work measures. The studies particularly yield that this problem is rooted in unidirectional intention communication, i.e. the robot displays its aim without observing the user. Therefore, the question of how user feedback could be integrated to enhance the task performance remains open.

Furthermore, when the robot operates under idealised full autonomy, its planning and performance tend to have high efficiency and its actions dominate the decisions of the user [58]. This has conflicted with users' perception of cooperative task solving, e.g. sometimes there was confusion about what the robot will do next. This goes in line with results from user studies about intelligent devices (Section 2.4) such as worn robots [221–223].

Another limitation of the current state of research is the assumption of the robot being omniscient concerning knowledge about both the task and the progress state. That way, the robot's plan is efficient and thus performance is higher when the robot's decisions dominate the user's intention. However, an application in an uncertain environment might require the robot to merge its plan with the user's and we lack knowledge about how that could be achieved. While this work still remains dependent on a high level of omniscience, we explore how strategies to achieve the aforementioned merging process could look like.

An important aspect of efficient multi-agent collaboration is a careful selection of an appropriate type of mixed-initiative interaction [67], i.e. deciding to what extent the machine or the robot should have control over task-related actions (see Section 2.5). Throughout this thesis, different types are tested in the context of the handheld robot and this chapter focuses on shared control [163], i.e. the fusion of user inputs and autonomous for simultaneous task management of the two parties.

The motivation for the work presented in this chapter stems from aiming to address the issues above and here we start by looking at incorporating models of user attention so that we can both

i) provide the robot with user information and a proxy for their intention and

ii) allow the evaluation of instances of conflicts between the user's and the robot's plans.

Remote gaze tracking is a natural choice for this due to the extensive body of work linking gaze and action prediction (Section 2.6.3). Land et al. [112] found out that eye gaze is closely related to a person's focus of attention as it precedes the location of actions during every-day tasks. This raises the question of whether eye gaze information could be used to inform a handheld robot about the user's intention or preferences. Therefore, this chapter is guided by the following research questions:

Q1 How can user attention be used to enhance cooperation with handheld robots?

Q2 How does the incorporation of attention affect task performance and the user's perceived task load?

For our study we use the open robotic platform[2] reported in [59], and modify it with the incorporation of a remote gaze tracker as described further down. The gaze information is first modelled and then used to evaluate its utility in a gamified cooperative task.

This chapter consists of two principal parts, one that looks at the gaze tracker modelling and characterization and another concerning the description and evaluation of the cooperative task under different modes of autonomy. The chapter sections are organised as follows: In a first step, a remote eye gaze tracker is repurposed for its use in the handheld robot (Section 3.2). Then, gaze data is combined with motion capturing data to construct a 3D eye gaze in Section 3.3. Furthermore, its limits are tested in user studies and its proximity to user focus compared against the use of head gaze as a predictor instead. Subsequently, an attention model is introduced that is based on both eye gaze information and task knowledge (Section 3.5). Its performance is then tested through user studies and compared against behaviours that are based on task knowledge only, raw eye gaze or non-motion, respectively.

[2] 3D CAD models available from www.handheldrobotics.org

The research contributions of this chapter are summarised as follows:

- A 3D gaze tracking framework is proposed, which is purposed for mobile robots that are working closely together with humans.
- A user attention system is introduced, where eye gaze and task knowledge are used to inform the model for attention estimation.
- A target reaching task is introduced as an experimental setup for data collection and model validation. Due to its main components being semi-simulated, it can be customised swiftly for other testing scenarios.

3.2 Eye Tracking for a Handheld Robot

The aim of the integration of an eye tracking system into the handheld robot is to gain attention-relevant gaze information of the user while the handheld robot is operated. In this section, we describe how a 3D gaze ray is constructed by merging 2D in-plane gaze information detected by a remote eye gaze tracker with support from motion capturing. The objective is to make the 3D gaze model available for the already existing robot simulation environment [57].

3.2.1 Eye Tracker Selection

Recent solutions for eye tracking are purposed to study human behaviour in correlation with their gazing behaviour. Commonly, they involve a camera system to detect the position of the pupil, from which the gaze can be derived [19, 137]. Existing models can be divided into two categories: head-mounted mobile devices and stationary remote trackers. Mobile trackers such as *Tobii Glasses 2*[3] and *Pupil Core* [29] benefit from a continuous high accuracy that is independent of the head pose. Previous research used them to study in-shop customers' gazing behaviour prior to decision-making [98]. Current models reach accuracies of visual angle detection of ∼0.6°. However, these mobile devices are costly with open source eye trackers (e.g. Pupil Core) starting from \$3,000 and commercial products (e.g. Tobii Glasses 2) from \$10,000, which exceeds the budget for this project.

Apart from the cost difference, a remote gaze tracker is preferred with regards to its application for handheld robots as it goes in line with the notion of a self-contained handheld tool. That way, users are not imposed to wear anything to use the tool which gets the concept closer to the handheld robot's aim of *pick up and use*. Remote trackers

[3]Tobii Glasses 2: www.tobiipro.com/product-listing/tobii-pro-glasses-2

are usually attached to a computer screen and deliver the 2D point of a user's current visual focus. Recent research made use of such devices to observe humans' behaviour during online shopping [235] and gaming [86]. Traditionally, remote tracking comes with the constraint that the head has to be fixed, which is unfeasible for many applications. However, new methods allow for remote gaze detection during natural head motion. Low-cost solutions (e.g. *WebGazer* [164]) detect gaze from screen-mounted Red Green Blue (RGB) cameras, e.g. webcams. However, their accuracy is low (∼3°) and they are prone to poor facial illumination. The development of recent models is mainly driven by high performance demands of the gaming industries. Hence there are now consumer products available on the market that are cost-efficient while their accuracy compares well to mobile trackers. This is a result of a combination of Infra Red (IR) illumination and a set of IR cameras on which the measurements are based on. Existing models differ in sampling rates from 90-150 Hz and their pricing ranges from $190 (e.g. *Tobii Eye Tracker 4C*[4]) to $1,995 (e.g. *GP3 HD Eye Tracker 150Hz*[5], hardware only).

Usually, manufacturers provide an Application Programming Interface (API) for their devices. However, only some of them are open source (see Table 3.1). This is a problem for 3D gaze modelling using remote trackers. Commercial APIs provide developers with the 2D gaze intersection point in screen coordinates, however, the underlying ray model is inaccessible. The gaze ray can be reconstructed for trackers that detect the 3D position of the eyes (see Section 3.2.2).

Model Name	Position	Accuracy [°]	Sample Rate [Hz]	Weight [g]	Open Source	Price [$]
Pupil Core	Head Mounted	0.60	200	23	Yes	3,000
Tobii Pro Glasses 2	Head Mounted	0.62	100	312	No	10,000
Tobii Eye Tracker 4C	Remote	~ 0.50	90	95	No	190
GP3 HD Eye Tracker	Remote	0.5 - 1.0	150	155	Yes	1,995
WebGaze	Remote	~ 2.7 - 3.6	30	~ 150	Yes	~ 50

Table 3.1: ***Summary of Recent Eye Gaze Trackers.*** *The accuracy is the error of the visual angle. The Open Source collumn indicates whether there is open software available for the device.*

Figure 3.1: Tobii Eye Tracker 4C [3]*, used for gaze tracking in this work (see application details in Figure 3.2).*

Table 3.1 summarises the properties of recent tracker models. Based on this information, the Tobii Eye Tracker 4C (see Figure 3.1) was chosen for this project. The API requires

[4]www.gaming.tobii.com/tobii-eye-tracker-4c
[5]www.gazept.com/product/gp3hd

the development of a wrapper for the gaze ray construction but the bundle is cost-efficient and accurate enough at tracking distances that suits attention estimation purposes for the handheld robot. Moreover, it is light-weight and so the additional load to be carried by users is small. The chosen tracker processes sensor information on an onboard chip, which reduces the Central Processing Unit (CPU) load of the lab machine, which might otherwise slow down the loop for attention estimation. With 90 Hz, its sampling rate is on the lower end but faster than the update loop of the robot controller, which operates at 75 Hz.

3.2.2 Gaze Ray Construction

We use a Tobii Eye Tracker[6] which delivers gaze information such as the user's current eye position and the 2D intersection between eye gaze and screen surface in screen centre coordinates.

For the following, let $\mathcal{F}_w, \mathcal{F}_b, \mathcal{F}_c$ be the frames of the world, the tracker's base and the centre of a screen plane, respectively. Furthermore, let ${}^i x_{eyes}$, ${}^i x_{gaze}$ and ${}^i x_{screen}$ be vectors describing the eye's position, the gaze direction and the gaze intersection with a screen plane in the associated frames ($i = w, b, c$). Consequently, ${}^i\boldsymbol{H}_j$ represents the homogeneous transformation between two frames $\mathcal{F}_i, \mathcal{F}_j$ [24]. The relationship between the frames can be seen in Figure 3.2.

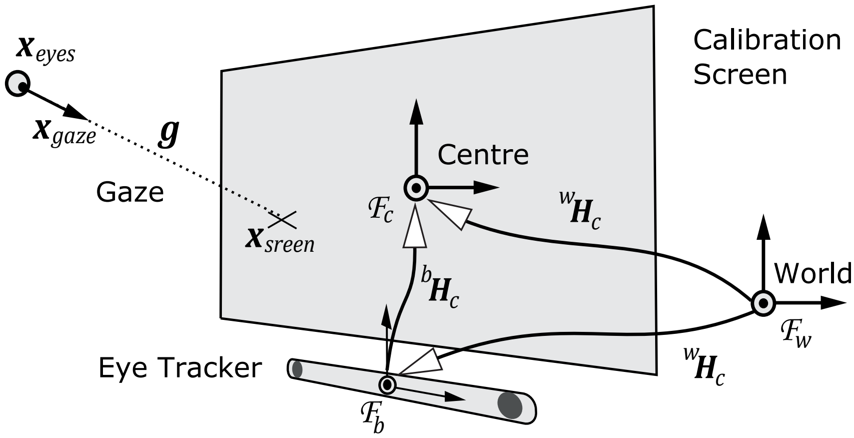


Figure 3.2: ***Eye Tracker Calibration****. This figure illustrates the calibration setup which is used to determine the (white tipped) transformation ${}^b\boldsymbol{H}_c$ which is required for the construction of the user's gaze $\boldsymbol{g}$.*

Both the eye tracker's base and the calibration screen are coupled to the motion tracking system, hence ${}^w\boldsymbol{H}_b$ is known at any time, ${}^w\boldsymbol{H}_c$ is solely known during the offset calibration process while ${}^b\boldsymbol{H}_c$ is initially unknown. As the tracking device delivers the eye-positions

[6] https://help.tobii.com/hc/en-us/articles/213414285-Specifications-for-the-Tobii-Eye-Tracker-4C

and gaze intersection with an associated screen, we can obtain the local gaze ${}^{c}\boldsymbol{g}$ with respect to $\mathcal{F}_c$ as follows[26]:

$$ {}^{c}\boldsymbol{g}(\lambda) = {}^{c}\boldsymbol{x}_{eyes} + \lambda {}^{c}\boldsymbol{x}_{gaze}, \tag{3.1} $$

where ${}^{c}\boldsymbol{x}_{eyes}$ is the mean of the two eye positions and ${}^{c}\boldsymbol{x}_{gaze}$ is the gaze direction which can be derived by

$$ \boldsymbol{x}_{gaze} = \frac{\boldsymbol{x}_{screen} - \boldsymbol{x}_{eyes}}{|\boldsymbol{x}_{screen} - \boldsymbol{x}_{eyes}|}. \tag{3.2} $$

Furthermore, while the eye tracker is attached to the calibration screen, we store the transformation ${}^{b}\boldsymbol{H}_c$ between tracker base and screen which is known through the equation:

$$ {}^{b}\boldsymbol{H}_c = ({}^{w}\boldsymbol{H}_c)^{-1}\, {}^{w}\boldsymbol{H}_b\,. \tag{3.3} $$

Now, we want to use the eye tracker without it being attached to the screen which means we lose information about ${}^{w}\boldsymbol{H}_c$. However, it can be derived from combining the tracker's base with the stored transformation

$$ {}^{w}\boldsymbol{H}_c = {}^{w}\boldsymbol{H}_b\, {}^{b}\boldsymbol{H}_c, \tag{3.4} $$

and finally, the gaze in world coordinates can be calculated in real time

$$ {}^{w}\boldsymbol{g} = {}^{w}\boldsymbol{H}_c\, {}^{c}\boldsymbol{g}. \tag{3.5} $$

Note that ${}^{b}\boldsymbol{H}_c$ remains time-invariant once the system is calibrated, hence, the gaze ray ${}^{w}\boldsymbol{g}$ can be calculated for any times given the user's eye gaze is detected by the eye tracker.

The eye tracker is mounted on to the handheld robot as can be seen in Figure 3.3. The position and orientation of the eye tracker can be adjusted so that the system can be adapted to varying user heights. In its generic configuration, the tracker is aligned such that it has the best accuracy within the robot's local workspace. Figure 3.4 shows a picture of the complete system.

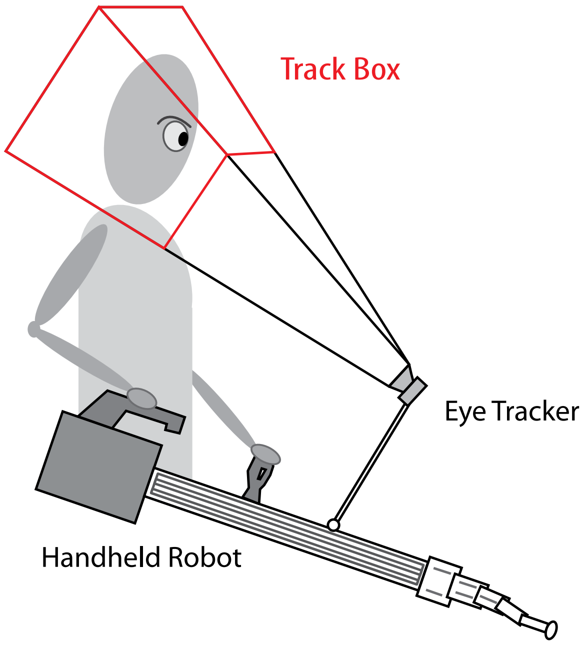


Figure 3.3: Illustration of the handheld robot with the mounted eye tracker. The mount supports 2-DoF adjustment so that the user's head remains in the (red) trackable volume.

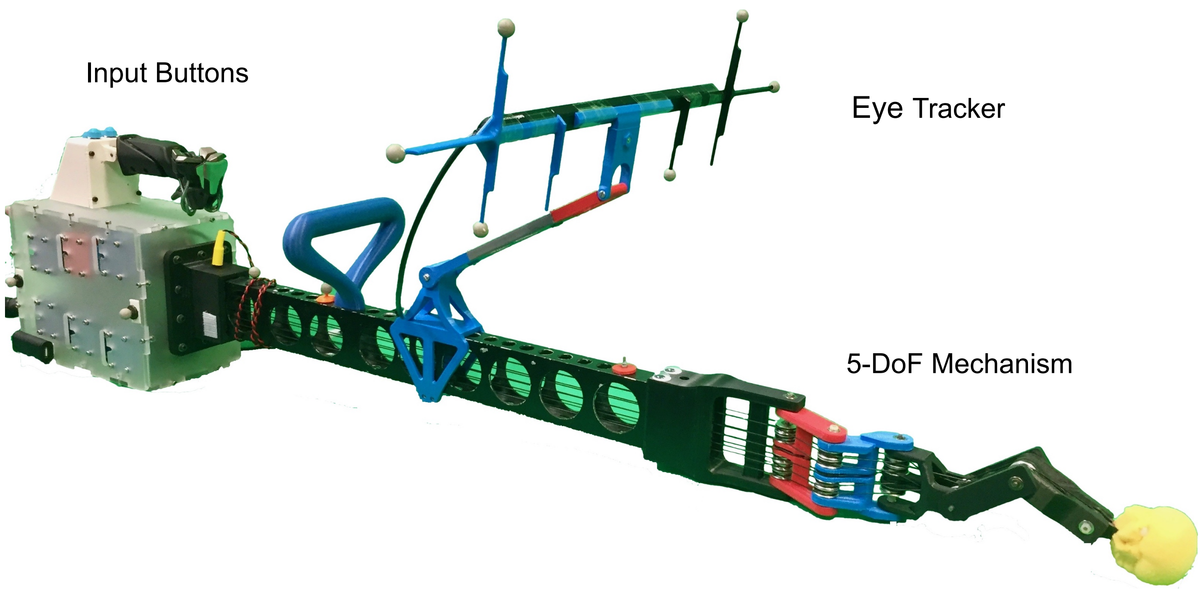


*Figure 3.4: **Handheld Robot** with extended user perception capabilities through a newly integrated eye tracking system.*

3.2.3 Merging Eye Tracking and Motion Capturing

The implementation of the gaze ray construction requires the connection of eye gaze tracking data and the tracker's world localisation. This section describes practical issues and how the equations from Section 3.2.2 are linked together using Tobii Eye X Software Development Kit (SDK), OptiTrack and C#/XNA as the principal software modules. An overview of the system can be seen in the flowchart illustrated in Figure 3.5.

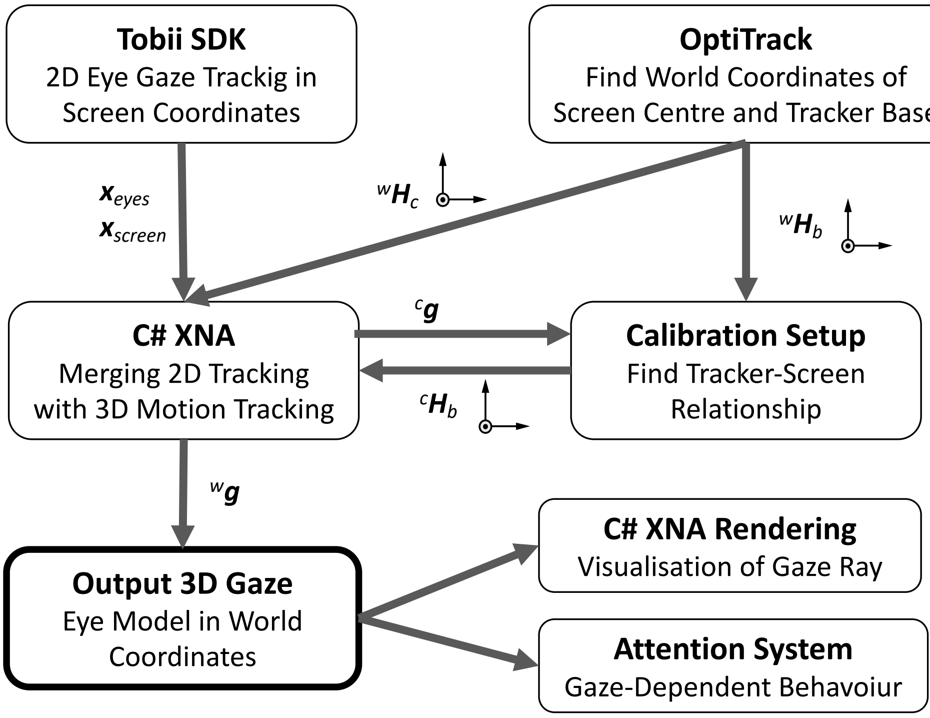


*Figure 3.5: **3D Eye Gaze Construction.** This illustration shows the data flow among the software components used for the gaze construction.*

As mentioned before, the Tobii SDK outputs eye gaze tracking data in 2D (screen) coordinates and the eyes' positions (with respect to the screen centre). This allows for a ray construction in screen coordinates. OptiTrack is used to localise the screen during the calibration step (see Figure 3.2) and the eye tracker base both during calibration and while it is mounted on the robot (see Figure 3.3).

Ideally, a handheld robot should be able to localise itself using onboard sensors, e.g. through Simultaneous Localisation and Mapping (SLAM) [9, 37]. However, external tracking systems are highly accurate and do not require additional sensors on the robot. At the same time, it allows focusing on human-centred research in this work. OptiTrack is an optical motion tracking system with two main components: a set of reflective markers and cameras (model Flex 3). The markers are attached to rigid bodies and reflect IR light coming from the IR Light-Emitting Diode (LED) that surround each camera. OptiTrack then combines the marker images for the localisation of a rigid body. A successful localisation requires at least 3 markers to be visible to at least two cameras. Initially, 4 cameras were used to localise the handheld robot [57].

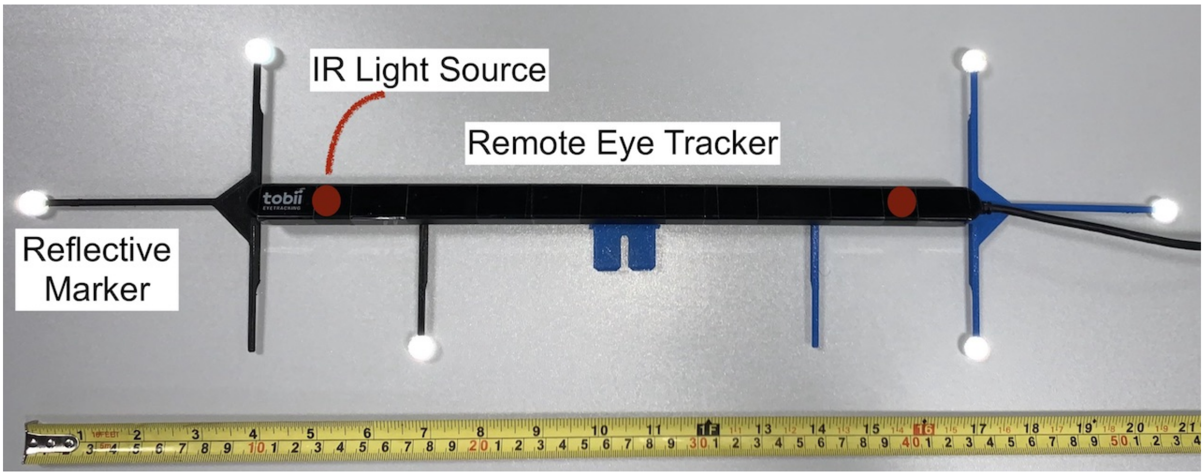


Figure 3.6: ***Gaze Tracker with IR Markers****. This shows the setup of the markers for motion capturing. Note that the markers were placed away from the IR light source to minimise interference. Knowing the pose of the device is required for gaze ray construction (see Figure 3.8).*

A problem with the Tobii tracker is that it emits pulsed IR light to illuminate the user's face. This interferes with the IR cameras of the OptiTrack system and leads to faulty localisation. To solve this problem, the markers were placed on antennas away from the IR LEDs. In addition, the camera set was extended to 7 cameras around the ceiling. That way, the localisation still works even if some cameras are blinded by the IR source. The eye tracker is not disturbed by the IR lighting from the cameras, presumably, because it uses pulsed light for the eye's illumination. Figure 3.6 shows the marker setup for the eye tracker. The arrangement of the cameras in the lab and the associated representation in the motion capturing software are displayed in Figure 3.7.

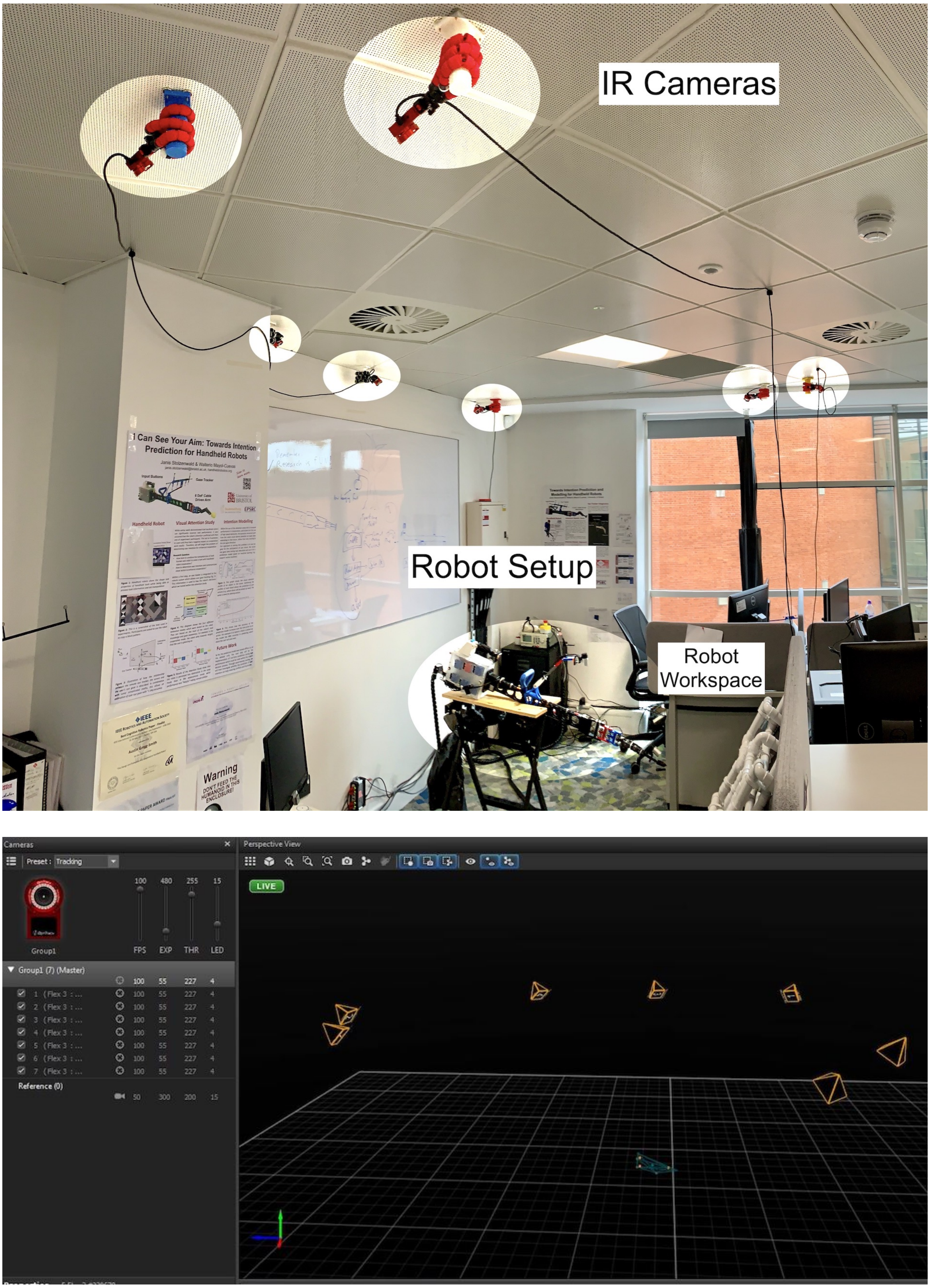
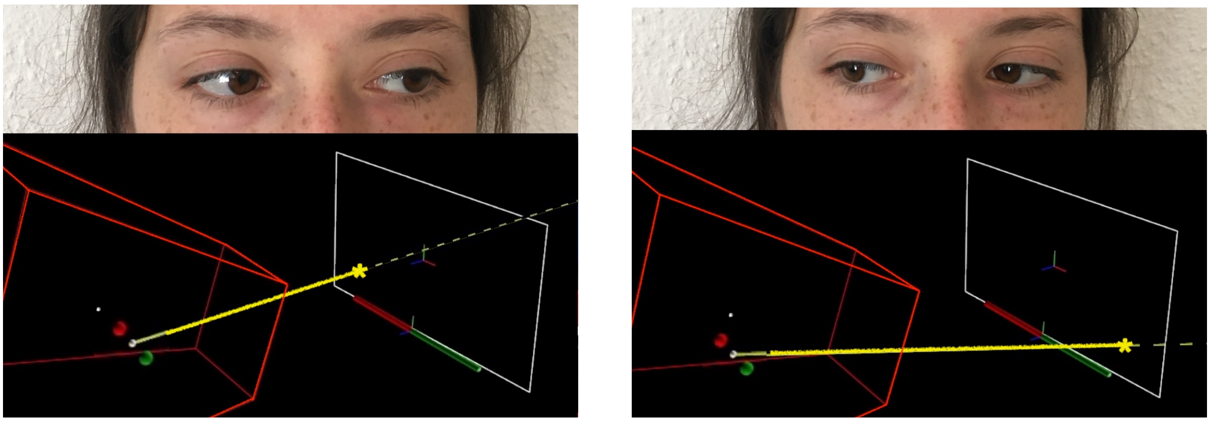


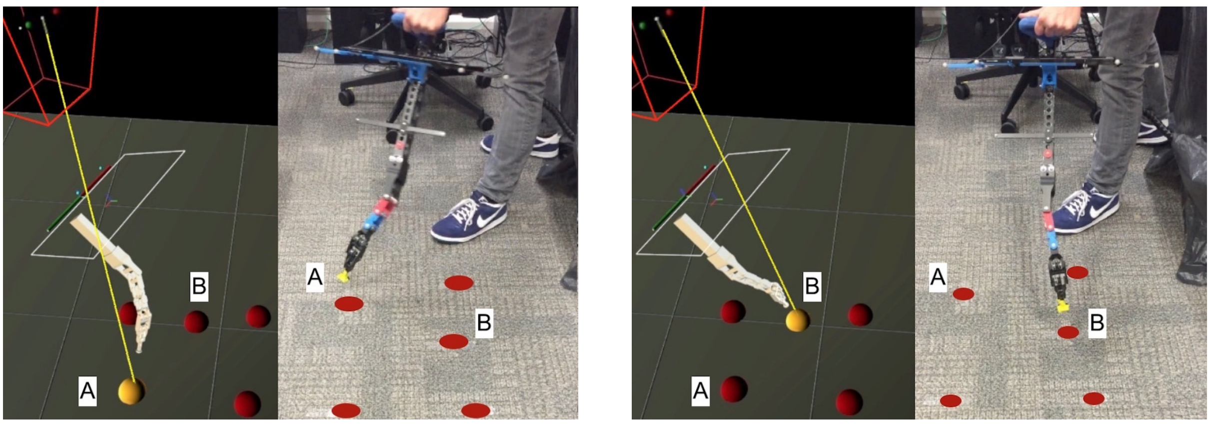
Figure 3.7: ***Camera Setup for Motion Capturing****. This figure shows the lab space with the camera setup for motion capturing (top). The cameras are positioned around the ceiling as indicated. The bottom part of the figure shows a screen shot of the camera poses rendered through Motive (supplementary software for OptiTrack). The setup allows for tracking the pose of the robot and the eye gaze tracker in real time. This is used to construct a 3D eye gaze ray with respect to the robot and its current environment.*

The data from gaze and motion tracking systems is processed using C# library *XNA*. It is a framework for game development, which offers efficient functions for real-time matrix operations and rendering for visualisation. XNA was chosen because the API for the handheld robot [57] is based on its libraries and so extending it with a module for eye tracking was the fastest way of integration. As can be seen in the flowchart in Figure 3.5, the XNA module reads the output of the gaze tracker and merges it with the OptiTrack output to construct the eye gaze in the 3D environment using Equations 3.1 to 3.5. The gaze model can then be used as an input for gaze-dependent robot control and gaze rendering. An example of gaze capturing and visualisation can be seen in Figure 3.8. An example implementation for the robot control with the eye gaze as an input can be seen in Figure 3.9, where the user selects a target for the robot through gazing. Figure 3.10 shows an overview of the complete system.



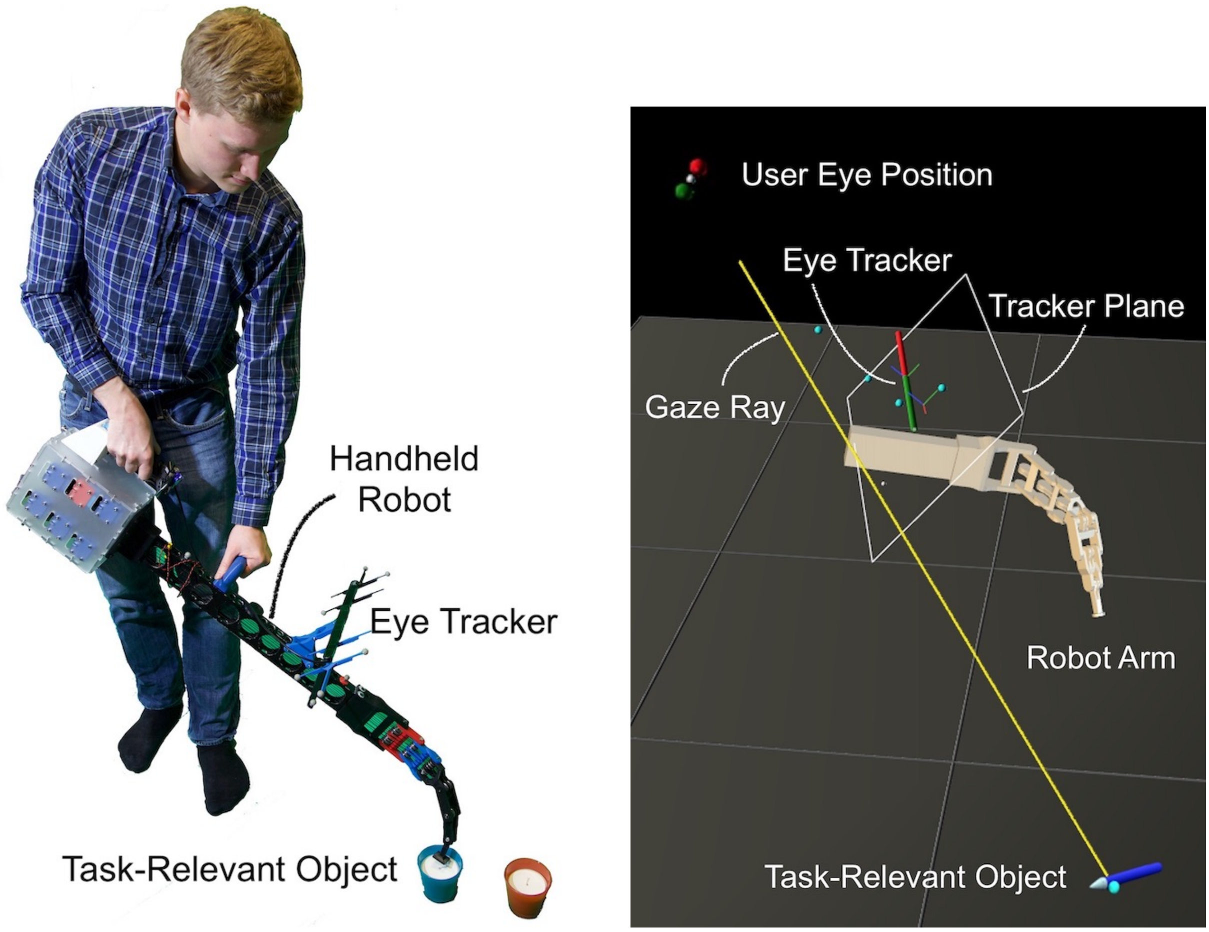
(a) Looking Left *(b) Looking Right*

Figure 3.8: ***Eye Gaze and 3D Model.*** *Observed eye gaze (top) and rendered output of the gaze construction module (bottom) in screen coordinates.*



(a) *(b)*

Figure 3.9: ***Gaze Following.*** *Demonstration of real-time capabilities of the 3D gaze model. The robot follows the user's eye gaze as it shifts from target A to B. Left and right are the rendered gaze model and the response of the robot, respectively. A demonstration video is available at https://youtu.be/D1gk4BxGRPs.*



(a) Real Setup *(b) Rendered Gaze Model*

Figure 3.10: ***Incorporating Gaze Tracking with the Handheld Robot****. This figure shows an overview of the handheld robot system after the integration of the remote eye gaze tracker. To the left is an example pose and to the right the rendered result of the proposed gaze model.*

3.2.4 Gazed and Fixated Objects

Integrating task-related objects in an attention-driven cooperative behaviour for the robot requires a measure of their visual attention. In the first step, gazed objects are identified and gazing times are used to derive fixations. Recall that fixations are the small time intervals that the eye spends on gathering visual information about an object of interest during scene exploration and task planning [187]. With the proposed 3D gaze model in place and the coordinates of an object to interact with, attention can be modelled as gaze-probability:

$$P_{gazed}(t) = \exp(\frac{-d(t)^2}{2\epsilon^2}), \tag{3.6}$$

where d is the angular distance between the measured eye gaze and the eye-to-object direction at a given time t and ϵ is the angular distance at which there is a significant drop of P_{gazed}, i.e. the expected angular error of the gaze sample. P_{gazed} can then be used to identify fixations on objects, i.e. when P_{gazed} exceeds a threshold for more than 200 ms [191].

With this in mind, gaze-aware objects were introduced to the robot API that have the following properties:

- `isGazed`: indicates whether the user is currently looking at this object, i.e. the gaze intersects with the object's boundaries
- `t_gazed`: duration of time the user has been looking at this object.
- `isFixated`: indicates whether the user is just passing the gaze through the object or is looking at it. `isFixated` is `true` if and only if `isGauzed` has been true continuously for at least 100 ms.

These form the basis for the robot's abstract behaviours such as following gazed or fixated objects.

Initially, the estimation of ϵ (see Equation 3.6) was unknown for the new 3D eye gaze model and needed to be identified through user studies. Furthermore, it is of interest to investigate to what extent a user's head direction indicates the visual attendance of an object, i.e. whether the head gaze error ϵ_h is small enough to distinguish between gazed objects. For these reasons, the following sections describe user studies concerning eye and head gaze accuracies, in which we explore to what extent each serves as a proxy for visual attention.

3.3 Eye Tracking Accuracy Study

The purpose of the accuracy study is to identify the limits of the introduced remote eye tracking system. Within the context of handheld robot application, we are particularly interested in constraints of the trackable area relative to the robot's workspace; that is how far the user can look away from the robot's tip for an accurate gaze capturing. Furthermore, we assess the accuracy of the measured gaze.

3.3.1 Eye Gaze Data Collection

The general approach is to keep track of eye gaze data throughout a task where a participant would look and point (the robot tip) at a randomised sequence of targets which are broadly scattered over a workspace. To generate a broad variety of target sequences, three typical workspace setups were selected: The floor, a table surface and a vertical screen.

For each target of the sequence, a participant is asked to look at a target without moving the robot and then touch the target with the robots' end effector while looking at the target. This would then be the starting posture for the next target iteration. For each iteration, we keep track of the following data:

$\boldsymbol{g}_{pri}$ / $\boldsymbol{g}_{pos}$ the true eye gaze ray to a prior/posterior target

$\Delta\phi_{gaze}$ the angular gaze shift (difference) between posterior and prior true gaze

$\boldsymbol{g}$ the measured eye gaze

ϵ the angular error of the measured eye gaze

`Tracked` True, if an eye gaze could be captured for a given target. Note that this is is different from `isGazed` and `isFixated` as those refer to a gazed object, whereas `Tracked` refers to successful gaze tracking, e.g. it would be false if the tracker got obscured or for large head angles away from the tracking device.

The relationship between those measurements can be seen in Figure 3.11. $\boldsymbol{g}_{pri}$ and $\boldsymbol{g}_{pos}$ are obtained from the eye position which is known from a motion tracked helmet and the associated target of which the position is known too. The angular gaze shift $\Delta\phi_{gaze}$ is defined as [26]:

$$cos\Delta\phi_{gaze} = \frac{\boldsymbol{g}_{pri} \cdot \boldsymbol{g}_{pos}}{|\boldsymbol{g}_{pri}| \cdot |\boldsymbol{g}_{pos}|}. \tag{3.7}$$

For each targeting iteration, these measurements are taken for the case where the robot is pointed towards the prior target as well as when the robot tip touches the posterior target.

3.3.2 Experiment Execution

We recruited 11 participants for the 1st pilot gaze estimation experiment, mainly students from different fields (4 females, $M_{age} = 25$, $SD = 4.8$). Participation was voluntary as there was no financial compensation for their time. The participants were asked to run through the aiming task for each target set-up with a random sequence of targets.

Before the experiments, the eye gaze tracker was calibrated (see Section 3.2.2) and its position on the robot adjusted to align it with the user's individual height of eyes. Each participant was asked to run through the aiming task for each target set-up, of which the order was randomised. As part of their introduction, participants were given some practice time to familiarise themselves with the robot and the trial procedure.

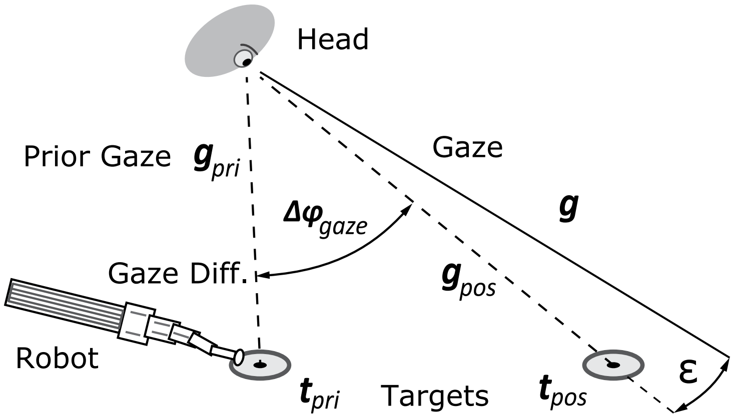


Figure 3.11: ***Illustration of a measurement iteration*** *where a participant is proceeding from looking at a prior target $\boldsymbol{t}_{pri}$ to looking at the posterior target $\boldsymbol{t}_{pos}$ before moving the robot towards it. Dashed lines represent calculated true eye gaze rays whereas the solid line is the measured eye gazes.*

For each of the three target set-ups, we took 2 data measurements for each target pair. Taking into account the measurement of the initial pose, we got 63 measurements per participant, so our final set contains 693 data points.

3.3.3 Eye Gaze Modelling Results

In order to determine the accuracy performance of the eye tracking system, we split the data into the subsets $\boldsymbol{S}_l$ (N = 330), where the participant is looking at the next target and $\boldsymbol{S}_p$ (N = 363), where the target was aimed with the eye gaze and the robot's tip at the same time. These are further split to distinguish between the cases where the eye gaze was recognised (`Tracked = true`) or not which is denoted with an additional 1/0-index (1 = `true`). That way, we get the subsets $\boldsymbol{S}_{l,1}$, $\boldsymbol{S}_{l,0}$, $\boldsymbol{S}_{p,1}$ and $\boldsymbol{S}_{p,0}$ with sizes N = 174, 156, 331 and 32, respectively.

$\boldsymbol{S}_l$ is used to investigate the effect of $\Delta\phi_{gaze}$ on ϵ, whereas $\boldsymbol{S}_p$ is used to investigate the accuracy for the case where the user's gaze is close to the tip. For the analysis, data points with a difference to the mean higher than two standard deviations are removed so that 2.55% is discarded.

For $\boldsymbol{S}_{p,1}$, we find a mean angular error of $\epsilon_0 = 1.99$ with $95\% CI[1.83, 2.16]$. The set $\boldsymbol{S}_{l,1}$ is analysed using a linear regression model. We calculate the values $c_1 = 1.243$ and $c_2 = 0.032$ for the model of the shape

$$\epsilon(\Delta\phi_{gaze}) = c_1 + c_2\Delta\phi_{gaze}, \tag{3.8}$$

where the slope c_2 is significant ($p = .012, R^2 = 0.032$). A diagram of the model can be seen in Figure 3.12.

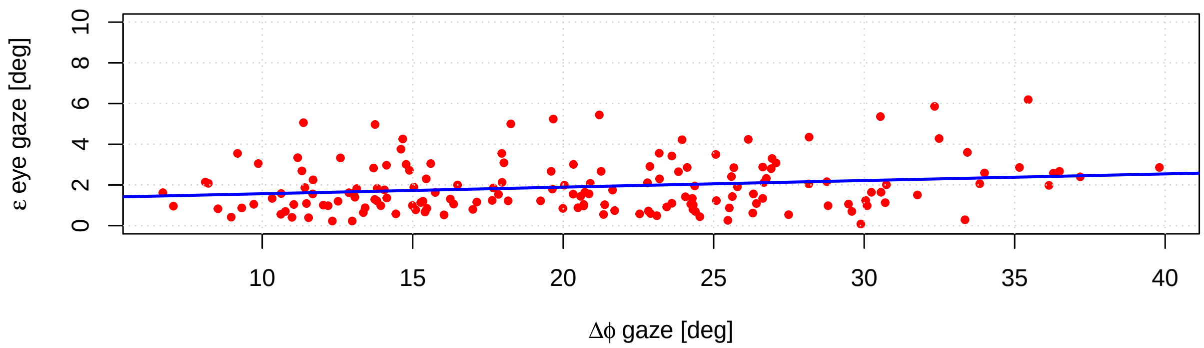


Figure 3.12: ***Eye Gaze Error over the Gaze Shift.*** *This diagram shows a linear regression model (blue) of the eye gaze error ϵ over the gaze shift $\Delta\phi_{gaze}$ (red samples).*

In order to estimate the limit of the gaze shift angle $\Delta\phi_{gaze}$ for which gaze tracking delivers reliable results, we use $\boldsymbol{S}_l$, the whole subset of samples where the participant was not looking at the tip. A logistic regression [89] is performed using the model

$$P(X) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)}, \tag{3.9}$$

where the independent variable is $\Delta\phi_{gaze}$ and $P(X)$ is the probability of the eye gaze being tracked (`Tracked = true`). To choose the optimal threshold value as a decision point for the model, we run a 5-fold cross-validation over the range $P(X) \in [.25, .75]$. As a result, we gain the decision point at $P(X) = 0.65$ for which the model fits 85.6% of the data and we get the coefficients $\beta_0 = 5.407$ and $\beta_1 = -0.177$ ($p < .001$ each) using the complete data set. By inverting the function at the decision point, we find that $P(X) > .65$ for $\Delta\phi_{gaze} \in [0, 27]$ (see Figure 3.13).

Feeding the $\Delta\phi_{gaze}$ range back into the linear model (Equation 3.8), we find $\epsilon_{27} = \epsilon(\Delta\phi_{gaze} = 27) = 2.107$ as an error prediction for the trackable range.

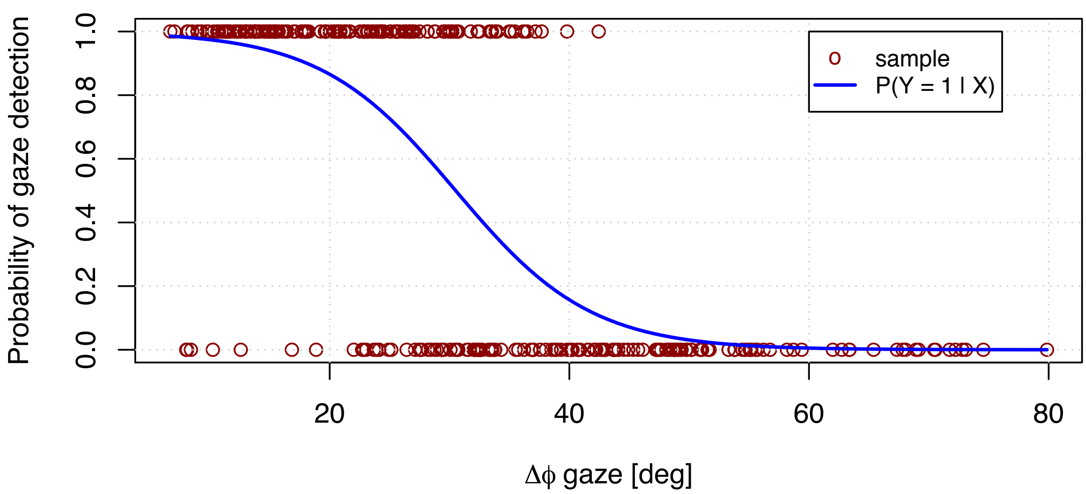


Figure 3.13: ***Angle Boundaries for Eye Tracking.*** *This shows a diagram of the logit model (blue) to estimate the probability $P(Y = 1|X)$ of successful eye tracking for a given gaze shift $\Delta\phi_{gaze}$. The (red) samples are the binary* `Tracked` *labels where* `true` *= 1 and* `false` *= 0.*

3.3.4 Gaze Tracking Discussion

In addressing the question about constraints of a workspace for eye tracking, it was found that it is limited by a maximum angle of 27 deg away from the robot's end effector. Within this constraint, we anticipate a probability of successful eye tracking above 0.65 which informs subsequent studies in terms of the limitations of experimental designs. As the angular limit goes to any direction, the workspace has the shape of a cone with a tip angle of $2 \times 27\,\text{deg} = 54\,\text{deg}$. For instance, for a user with an eye hight of 1.3 m this would result in a workspace plane of $2 \times \sin(27\,\text{deg}) \times 1.3\,\text{m} = 1.18\,\text{m}$, i.e. covering more than the robot's local workspace.

Concerning the angular accuracy of the eye tracking device in handheld robot applications, we identified a link to $\Delta\phi_{gaze}$. However, the linear regression yields a small slope coefficient indicating a small effect of the gaze direction on the error. The average error for a range below 27 deg is smaller than the maximum of the error's 95% CI of the $\boldsymbol{S}_p$ set. Therefore, an average error of up to $\epsilon = 2.16\,\text{deg}$ is anticipated.

This information is useful for estimating at what sizes and distances gazed objects can be distinguished. In order to maintain a robust response for gaze awareness of objects, i.e. detecting whether an object is being fixated by the user, an object size of at least 2 times the accuracy of the gaze model's accuracy is suggested. For example for an object at a distance of 1.3 m, the ideal size would be $2 \times \sin(2.16) \times 1.3\,\text{m} \approx 85\,\text{mm}$.

3.4 Head Gaze as a Proxy for Visual Attention?

Previous work [161] linked humans' head pose to their current target of visual focus in conversation settings. Therefore, one might assume that the head pose of handheld robot users could serve as a proxy for their current visual attention. This part of the attention study investigates whether this assumption is valid. In the first step, the head gaze is constructed and in a second step, it's relation to the eye gaze is investigated.

3.4.1 Head Gaze Tracking

Here, head gaze is defined as the direction of a person's facing. Constructing the head gaze requires the head position and the orientation of the face. As mentioned in Section 3.3.1, a tracked helmet was used to determine the head's position and its centre. As the orientation of the helmet relative to a person's head varies between users, the experiments were preceded by a calibration step to adjust the z-axis of the head gaze frame. For that, participants were asked to wear the helmet and look at a distant point at eye level. The point is known to the tracking system, which allows alignment of the z-axis of the head gaze frame with the line connecting the centre of the head and the distant point. An illustration of this head gaze construction can be seen in Figure 3.14.

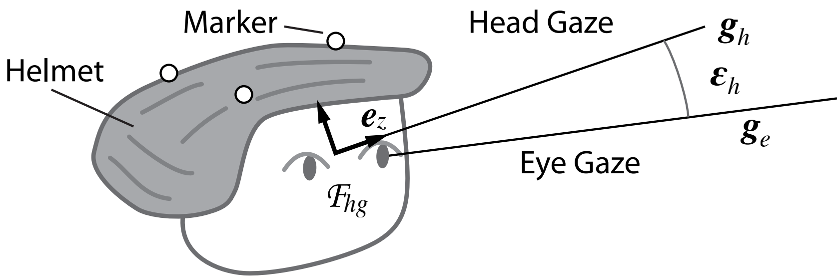


Figure 3.14: ***Head Gaze Tracking****. The illustration shows how a helmet with markers was used to track the head orientation. The z-axis of the head gaze frame $\mathcal{F}_{hg}$ was individually calibrated prior to each experiment trial. The resulting head gaze is compared with the eye gaze to derive the head gaze error ϵ_h.*

3.4.2 Collection of Head Gaze Data.

The data for the head gaze study was collected as part of the previously described eye gaze tracking accuracy study (Section 3.3.1). Recall that participants used the robot to point at predefined targets in the workspace while looking at a different target, i.e. away from the robot's tip. As previously defined, the angular difference between the robot's tip and the new target is denoted as the gaze shift $\Delta\phi_{gaze}$, as per Equation 3.7. Similar to the identification of the eye gaze error, the head gaze was recorded for each step of the target sequence. The head gaze error ϵ_h was derived as the difference between the ground truth eye gaze (see Section 3.3.1) and the measured head gaze, i.e. the difference between the head gaze and the direction of subjects' attention. As opposed to eye gaze tracking, tracking of head gaze is free from angular limits as it is derived from the motion capturing system and thus independent of the relative position to the robot. With 11 experiment participants and a sequence of 33 targets per trial, the head gaze data set contains N=363 data points.

3.4.3 Results and Discussion of Head Gaze Study

To analyse how head gaze error deviates from the current focus of attention, a linear regression was applied. For $\boldsymbol{S}_l$ we ran a linear regression with gaze angle difference $\Delta\phi_{gaze}$ as independent variable and angular head gaze error ϵ_h as observation [89]. We calculated the values $c_1 = 11.762$ and $c_2 = 0.392$ for the model Y_h of the shape

$$l(X) = c_1 + c_2 X, \tag{3.10}$$

with a significant slope ($p < 0.001, R^2 = 0.236$). A diagram of the regression model can be seen in Figure 3.15.

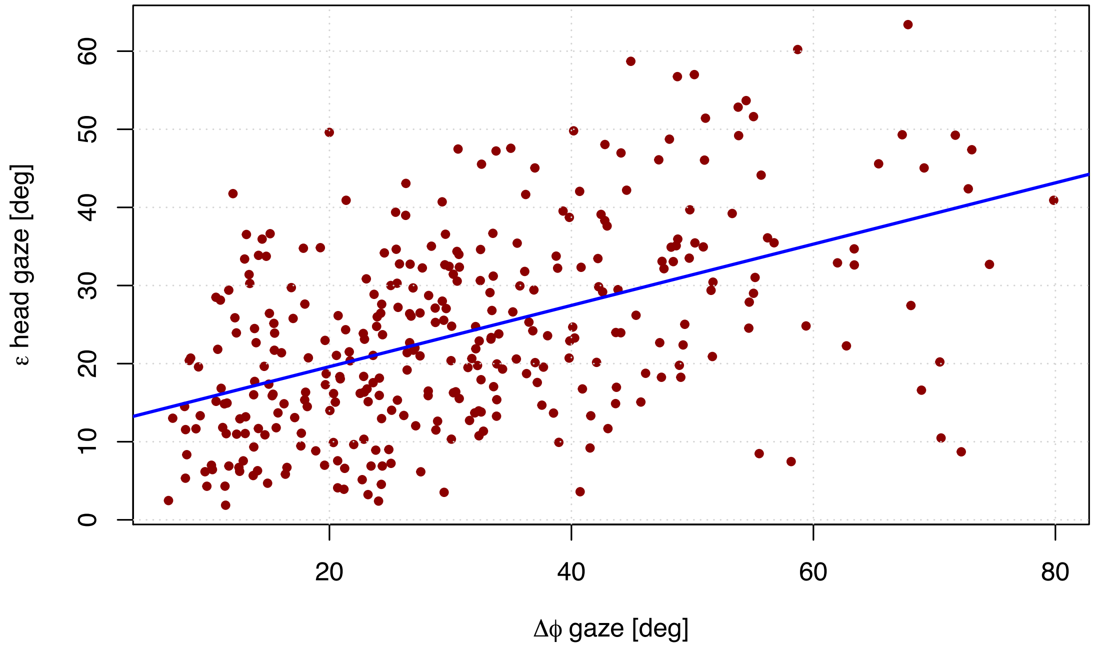


Figure 3.15: ***Head Gaze Error.*** *This shows a linear regression model (blue) for head gaze error data over the gaze shift.*

The results show that the head gaze largely deviates from the direction of a focused target. While the smallest error (i.e. $\epsilon_h > 11.72\,\text{deg}$) can be expected for targets close to the robot's tip, the error grows rapidly for targets with an increased angular distance between the target and the robot's tip. A reason for this could be that the user's body is constrained to the robot as both hands are required to hold the robot. Therefore, targets outside the field of view are aimed with the eye gaze rather than with the head gaze.

The deviation between head gaze and user's visual attention is too small to determine which object the user is focusing on for common object sizes in handheld robot applications. However, the results suggest that the head angle could be useful to estimate whether the user is currently focusing on the workspace of the handheld robot. This could be useful to guideline systems for safe interaction with the robot. As the eye gaze accuracy outperforms head gaze accuracy, this is chosen as a basis for the following attention model.

3.5 An Attention-Aware Cooperative Handheld Robot

Having determined the boundaries of the gaze tracker for its application in the handheld robot, the next step is to use the user gaze as a source of information about user attention for the robot to assist in the task. This section describes how task-relevant objects are filtered for aiming using gaze awareness and task knowledge to parametrise assistive behaviour. This is followed by user studies to investigate the attention model in action. On the remainder of this chapter, the results from the eye tracking studies are applied for incorporating the estimation of user attention.

3.5.1 The Attention Model

The attention model for the handheld robot is based on two factors: gaze awareness and task knowledge. These factors are inspired by the work of Land et al. [112] who suggest that eye gaze is closely related to the location of a person's action. Therefore, we propose the following assumptions:

A1 The part of a workspace which is watched by the user is within the user's focus of attention.

A2 A watched object is more likely to be in the user's focus of attention when it is task-relevant than when it is irrelevant to the task.

Based on these assumptions, we create a behaviour matrix with gaze awareness and task knowledge as two principal axes determining behaviour. As attention awareness is the product of both gaze awareness and task knowledge, it increases over the diagonal axis of the matrix as illustrated in Figure 3.16.

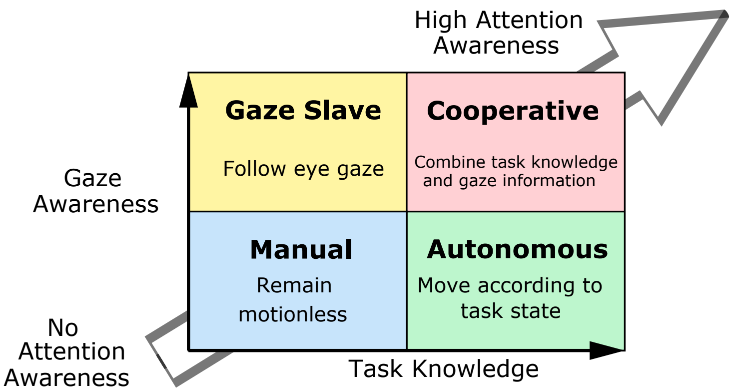


*Figure 3.16: **Parameters for the Robot's Behaviour Modes.** This behaviour diagram illustrates how the four behaviour modes of the handheld robot are linked to the attention model which is based on the level of gaze awareness and task knowledge.*

The details about each behaviour mode are described in the following:

B1 **Manual Mode**: The robot remains motionless since neither gaze nor task knowledge influences its behaviour. The user is fully in charge and can decide when and with which objects to interact with, however, the robot does not assist in the task.

B2 **Slave Mode**: In this mode, the robot ignores the status of an object or whether it is related to the task at all. Instead, the behaviour is purely determined by the estimation of the user's area of attention. This goes in line with assumption A1 so that the robot follows the user's eye gaze in the workspace.

B3 **Autonomous Mode**: The robot ignores any user actions and follows its own plan to complete the task. Choosing the sequence of task objects and finishing the job is fully automated. Furthermore, it overrides the trigger input, i.e. it decides on the time and duration of object interaction. The robot still depends on the user to carry it close enough to a target for interaction.

B4 **Cooperative Mode**: The focus of attention is modelled as the intersection between gazed-at area and location of a task-relevant object which goes in line with A2. The robot follows the eye gaze and helps to aim when a task object is fixated. While the robot finishes the job, the user can shift the visual focus to a subsequent object. The robot catches up with the eye gaze once the task with the previous object is completed.

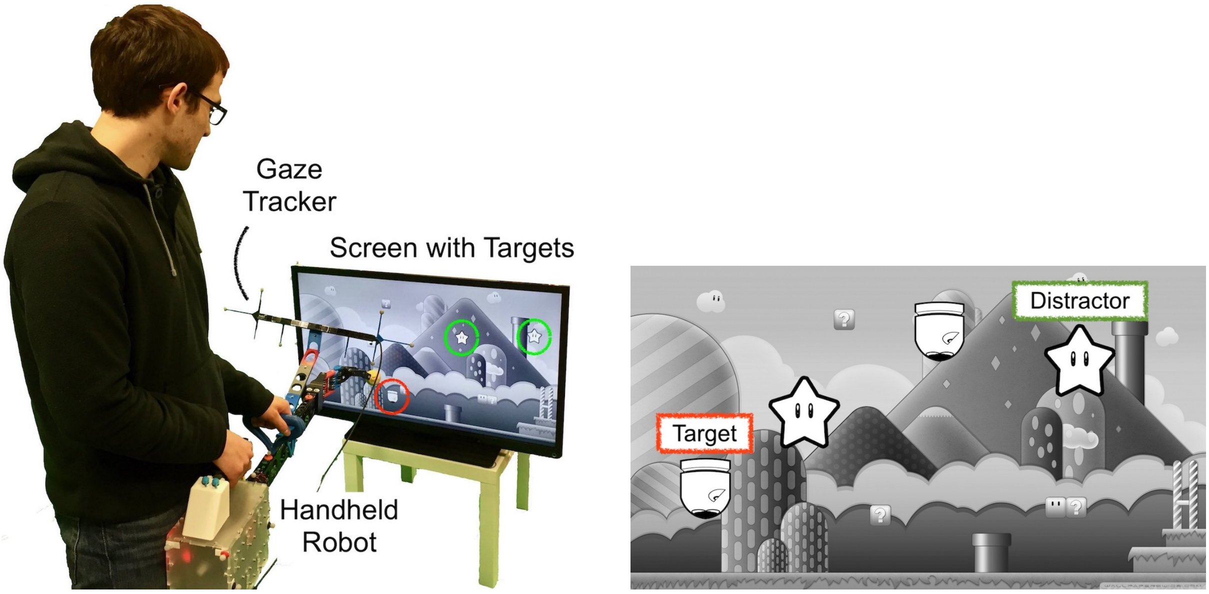
3.5.2 Validation Task

Having developed different behaviour modes based on the novel attention model, the next step is to assess the effect on cooperative task solving through user studies. This section describes the experimental task used for the study and how the robot applies the above-listed behaviour modes to assist in the task. The design of the experimental task was guided by the following list of requirements:

- **Cooperative.** The robot should be able to assist in the task using the attention model and its task knowledge. The task should be easy enough to be carried out by novice users. At the same time, the robot would depend on being moved around by a user, i.e. it could not solve the task by itself.
- **Within the Gaze Tracking Constraints.** The Workspace should fit within the constraints of the field of view in which eye tracking works reliably and task objects dimensioned with regards to gaze tracking accuracy.
- **Cognitive Load**. A design is required where the robot's behaviour modes can be assessed in the context of different cognitive loads.

With these in mind, a semi-simulated validation task was developed where participants use the robot to catch targets on a screen. The game principle is inspired by *Space Invaders*[7] and displayed on an LCD screen with 105 cm diagonal. An overview of the task setup can be seen in Figure 3.17.

[7] Available for example at *http://www.pacxon4u.com/space-invaders/*



(a) (b)

*Figure 3.17: **Attention Experiment Task and Setup**. Figure (a) shows the testing setup with the interactive task. The user has to stop the dropping targets (red) using the robot to increase the game score while distractors (green) can pass. The robot assists in the task based on the user's eye gaze. Figure (b) shows an example of the screen content. The targets randomly appear with varying speeds and frequencies.*

Markers were attached to the screen to determine its position relative to the robot and the eye tracker. That way, the intersection points between rays (e.g. gaze or robot tip direction) can be calculated in real-time and be fed to the control unit of the robot. Another advantage is that the positions of every pixel of the screen are known in 3D space. This framework offers a lot of flexibility for task design because the positions of targets are known, even when moving, without the necessity of individual marker setups.

In the game setup, targets travel with constant speed from the upper edge of the screen to the bottom line. The aim of the game is to stop as many targets as possible before they reach the bottom.

The robot's tip emits a virtual laser that can be used to stop a target. The laser needs to be activated via a trigger on the robot's handle and the tip needs to be close to the target (< 100 mm). It takes some lasering time to stop a target but less time for quicker targets since otherwise, it would be impossible to complete it before it reaches the bottom line. The outcomes of the accuracy experiments are used to dimension both screen size and target diameter.

To stimulate best possible performance, a scoring system was introduced where the score increases for caught targets and decreases for missed ones, i.e. $+100$ and -100 respectively. Over the series of trials, participants were motivated to beat their current high score.

The design of the game is greyscale coloured to avoid a disadvantage for colour blind people. The background of the game is a cartoon landscape and alongside the targets, similar objects are dropping which cannot be stopped but are used as distractor stimuli (see Figure 3.17). While 50% of the targets are spawned randomly, the rest are part of a challenging scenario. For example, the targets would drop in line or triangle formations or in arrangements, where some slower targets are being taken over by quicker ones. This exposes users to situations where there are targets with equal priority. Equally, they would face situations where a target's priority suddenly becomes exceeded by a new faster and thus more time-critical target. The purpose of those arrangements is to assess the robot's behaviour modes in scenarios with non-trivial solutions, i.e. there will likely be times where the robot's plan does not initially meet users' intuition.

Concerning the modes of the robot (B1-B4) with regard to the concrete experimental task, the robot behaves as follows: The user is fully in charge in the manual mode. This mode is used as a reference to compare against for the other three modes. As the robot's tip remains motionless (see B1), users have to perform both tactical and aiming motion to catch targets, which also includes pulling the trigger for laser activation at the right time. In slave mode (B2), the robot keeps moving towards the point of intersection between user's eye gaze and screen. That way, users can use their eye gaze to steer the robot towards a target of their choice. In cooperative mode (B3), the robot uses its task knowledge to assist in the aiming task. Fixated objects are aimed at by the robot and it overrides the laser trigger to assist with timing. The robot finishes the local job once it started even if the user looks away, e.g. because they look out for the next target. In autonomous mode (B4), the robot automatically decides which targets to aim for. The user has to follow and help the robot reach the chosen target to complete the task. The robot uses its task knowledge to optimise the target sequence to complete as many as possible in the given time. Its decisions are based on an underlying greedy approach that prioritises time-critical targets (i.e. that are close to the bottom line) and optimises for short ways. The robot also knows whether a target is not worth aiming for, i.e. when the completion time would exceed the target's time to travel to the bottom line. Unlike cooperative mode, the autonomous mode does not take into account the user's eye gaze, i.e. its strategy is independent of user attention.

3.5.3 Attention Experiment

For this new experiment of the attention study, we recruited 15 participants (6 females, $M_{age} = 25.5$, $SD = 5.6$). Many were students from technical courses, however, there was no expertise required to solve the task. 3 participants, i.e. 15% have participated in eye tracking accuracy study in Section 3.3. However, every participant was trained to use the handheld robot extensively, hence we argue that knowledge advantages from previous experiments are negligible. There was no financial compensation for their time, however, many volunteers were thankful for trying the robot and the game and they were offered some refreshments. Each participant ran 3 game trials in each behaviour mode for a duration of 80 s each. They were exposed to targets with varying speeds and stopping times and the order of the behaviour modes was randomised to cancel out training effects. Before starting the experiment session, the participants were given an explanation and demonstration for each mode plus some time for practising to get familiar with them.

For an objective measure of cooperative performance, the share of completed targets was recorded, i.e. how efficient the task was solved. During trials, we kept track of the count of targets that were completed as well as the total number of targets the subject was exposed to. For the event of a target being stopped by the user or reaching the bottom line (i.e., the user failed to stop it), a data point was created, which contains the target's speed and the current behaviour mode of the robot. Over the 180 trials (i.e. 15 participants $\times$ 4 robot modes $\times$ 3 speed ranges), we collected 17,000 target samples in total.

The game runs with an update rate of 70 Hz and when one of the gaze-based modes is used (B2 - B4), we registered whether the tracker recognised the eye gaze. This was later used to analyse the reliability of the gaze model during task execution.

After each trial, the participants were asked to complete a questionnaire to assess the current trial and mode. The main parts of the questionnaire are the *NASA* TLX criteria [63], which were used to measure the subject's task load (see questionnaire in Appendix D). This standardised questionnaire was used in previous studies to assess the cooperation quality of the handheld robot [58, 60]. The test consists of a series of task load criteria, measuring mental/physical/temporal demands, effort, subjective performance and frustration on a continuous scale. Furthermore, we asked participants to what extent they agree with the statements: *The robot helped me with the task* and *The robot obstructed me during the task* on a 5-point Likert scale (*Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly disagree*). Also, they were given the chance to provide feedback comments for the current trial.

3.6 Results of Attention Study

The target data set is used to assess subject performance for each behaviour mode. Here, performance is defined as the proportion of completed targets over the total targets presented. Furthermore, the set is split into three speed ranges $R_{1,2,3}$: $[70, 200)$, $[200, 330)$ and $[330, 490]$ (in mm/s) for a separate analysis. The lower bound of 70 was chosen because below that, most people could stop every target in every mode, hence this would not serve us with any valuable information. Initially, the ranges were then equally divided between 70 and 460, i.e. with an equal bin size of 130. The fastest bin was extended to 490 because a few participants managed to catch targets with such a high speed value, i.e. those edge cases would have been excluded otherwise. 2.1% of the data points are outside of these ranges and are thus discarded. This is due to some malfunctioning of the randomisation function of the C# standard library that was used to vary the target speeds. A few percentage of the output is much higher or lower than the predefined boundaries.

3.6.1 Mode Performance

The effect of the speed range and the behaviour mode on the performance is determined using a two-way factorial repeated measures Analysis of Variance (ANOVA). As the results yield a significant effect for each factor ($p < 0.001$), they are further explored using post-hoc pairwise t-tests with applied Bonferroni correction. The mode-dependent differences in performance for each speed range and associated t-test results can be seen in Figure 3.18.

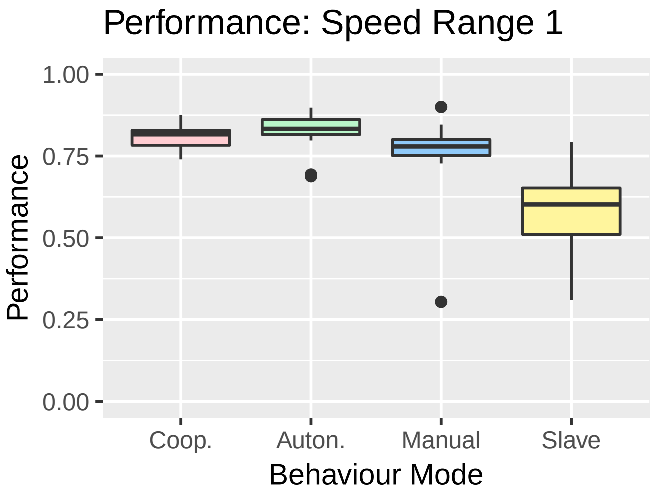
For every speed range, slave mode is outperformed by the other modes and the performance yields a significant difference to each. Cooperative mode and autonomous mode outperform manual mode for each speed range, however, significance can only be determined for the two higher speed ranges R_2 and R_3 and not for R_1. In no case could a significant difference between the performance of the cooperative mode and the autonomous mode be found (see Figure 3.18)

We do not find any correlation between performance and age, gender, hours per week that video games are played or whether vision aids such as glasses or contact lenses were used.

Considering the set of game update frames where the eye tracker did not recognise the eye gaze, we note that those add up to a share of 49.9%. However, they are evenly distributed over the trial time so that in only 5.1% of the time, these frames locally add up to over 150 ms.

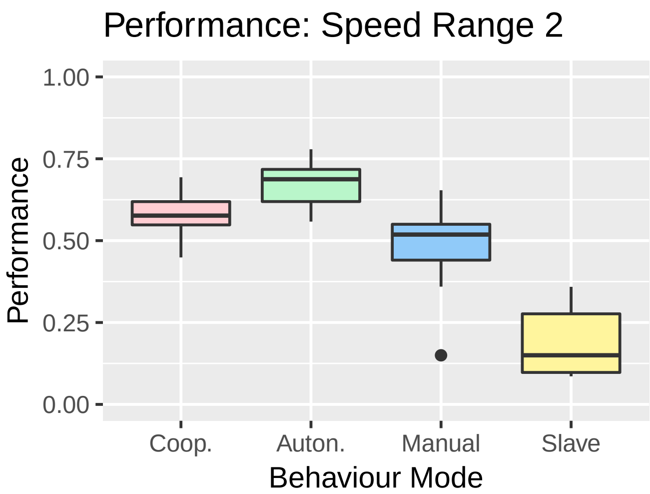
3.6.2 Task Load Index

For the analysis of the modes' effect on the perceived task load, i.e. the combined NASA TLX results, we applied an ANOVA for the combined dataset and determine significance ($p < 0.001$). We proceed with a post-hoc pairwise t-test between the modes where the p-values (displayed in Table 3.2) are Bonferroni corrected. The mode-dependent differences of TLX results can be seen in the diagram in Figure 3.19. The results for slave mode yield



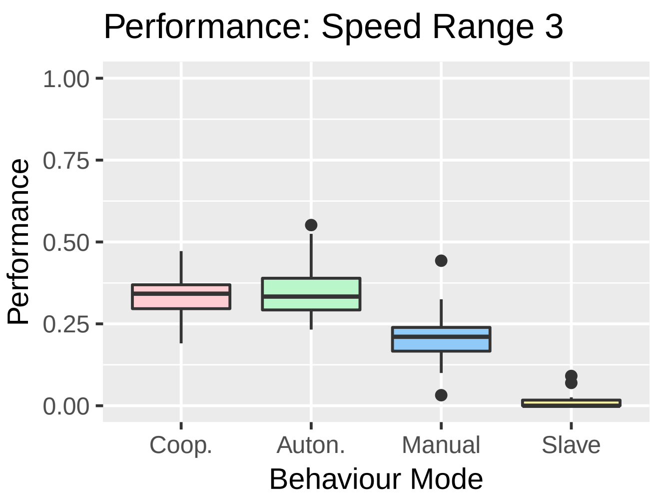
	Coop.	Auton.	Manual
Auton.	1	-	-
Manual	0.82	0.33	-
Slave	**1.40E-07**	**2.70E-08**	**3.90E-05**

(a) Speed Range 1: Slow



	Coop.	Auton.	Manual
Auton.	0.058	-	-
Manual	**0.038**	**5.60E-06**	-
Slave	**2.40E-15**	**< 2E-16**	**6.80E-11**

(b) Speed Range 2: Medium



	Coop.	Auton.	Manual
Auton.	1	-	-
Manual	**0.00043**	**2.40E-05**	-
Slave	**3.10E-15**	**<2E-16**	**2.30E-08**

(c) Speed Range 3: Fast

Figure 3.18: ***Comparison of Performance for the Robot's Modes****. The performance (higher is better) is measured in completed targets over total targets for each mode and speeds. The speeds range from slow (1) to fast (3). The tables display the respective Bonferroni corrected p-values of pairwise t-test results. Significant ($p < 0.05$) values are displayed in bold face. Note that the cooperative and autonomous mode outperform the reference mode (manual) for higher speeds. The manual job outperforms the slave mode consistently while no significant difference was observed between the autonomous and cooperative mode.*

a significant difference to the cooperative mode and the autonomous mode. Furthermore, the cooperative mode significantly outperforms the manual mode .

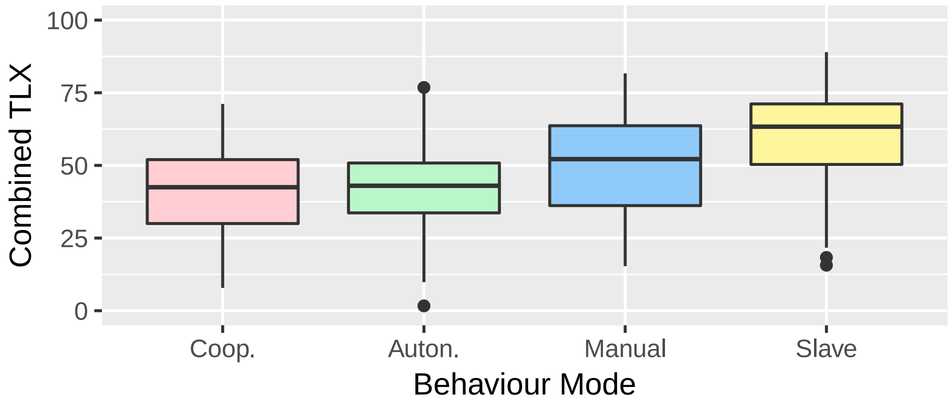


Figure 3.19: ***Results of the Task Load Index Questionnaire.*** *Perceived task load for each behaviour mode measured by the combined NASA TLX (lower is better). Associated t-test results can be seen in Table 3.2*

	Coop.	Auton.	Manual
Auton.	1	-	-
Manual	**0.038**	0.102	-
Slave	**1.30E-05**	**6.00E-05**	0.202

Table 3.2: ***Pairwise t-test results.*** *This table shows the Bonferroni corrected p-values of pairwise t-test results for the mode-depended mean differences of TLX outcomes. Significant ($p < 0.05$) values are displayed in bold.*

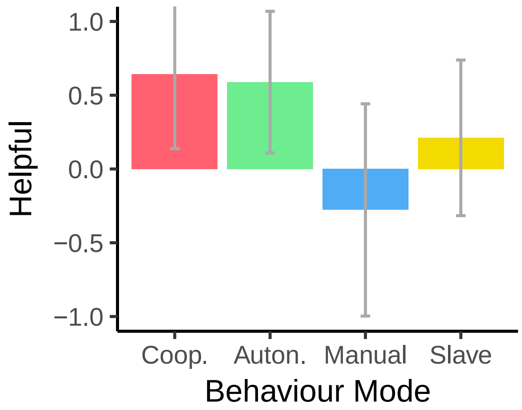
3.6.3 Helpfulness and Obstruction

The 5-point Likert scales (from *strongly disagree* to *strongly agree*) for the statements about the robot's helpfulness and obstruction are scaled to numeric values on the interval $[-1, 1]$. As can be seen in Figure 3.20a, the robot is rated most helpful in the cooperative (0.64) and autonomous (0,59) mode followed by the slave mode (0.21) while the manual mode tends towards unhelpful (-0.28).

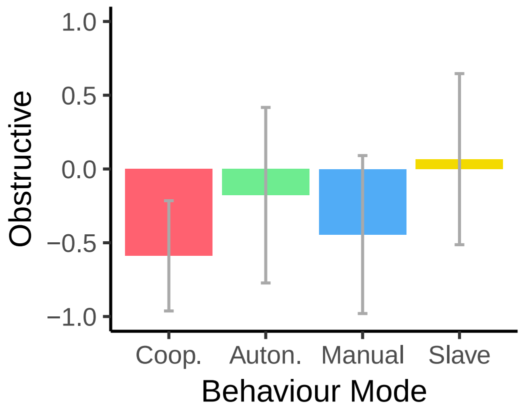
3.6.4 Qualitative Feedback

As commenting on the trial behaviour was optional, the number varies for the different modes. We received many comments on the cooperative and autonomous mode and a few for the slave mode, whereas the manual mode remained mostly uncommented.

Comments on the cooperative mode are mostly positive and often refer to collaboration experience, e.g. *I feel comfortable with the robot's assistance* or *It feels like a team* and



(a) Higher is better



(b) Lower is better

Figure 3.20: ***Helpfulness and Obstruction for the Behaviour Modes****. Participant's rating of the robot being helpful (left) or obstructive (right) where 1 is strongly agree and -1 is strongly disagree. Whiskers indicate standard errors.*

The robot helped me with accuracy once I chose a target. Also, participants often pointed out that they felt in control, e.g. *I like the shared control* or *I feel more in control with it*.

For the autonomous mode, comments are positive when referring to accuracy, e.g. *The robot is better than me* and negative (mostly for fast targets) in terms of the robot's predictability, e.g. *I was irritated when the robot changed plans* and *Why are you going there?*

Within the slave mode, participants were complaining when eye tracking was faulty, e.g. *There was some offset, the robot did not follow accurately* and *I saw a target but [the robot] did not follow.*

The few comments on the manual mode were addressing physical workload, e.g. *It was exhausting* and *The robot feels heavy.*

3.6.5 Qualitative Observations

Over the set of trials, there were a few issues that occurred repeatedly in the respective modes. For example, in slave mode it was a common observation that the robot's tip motion was irritating for users when they were exploring the scene, i.e. before aiming for targets. Furthermore, a feedback effect could be observed in some instances when the user focuses on the robot's tip. In that case, the robot's tip started to drift due to the inaccuracy of the gaze tracker.

In autonomous mode, users sometimes struggled with mismatches between their plans and the robot's autonomous actions. This occurred more frequently in settings with higher

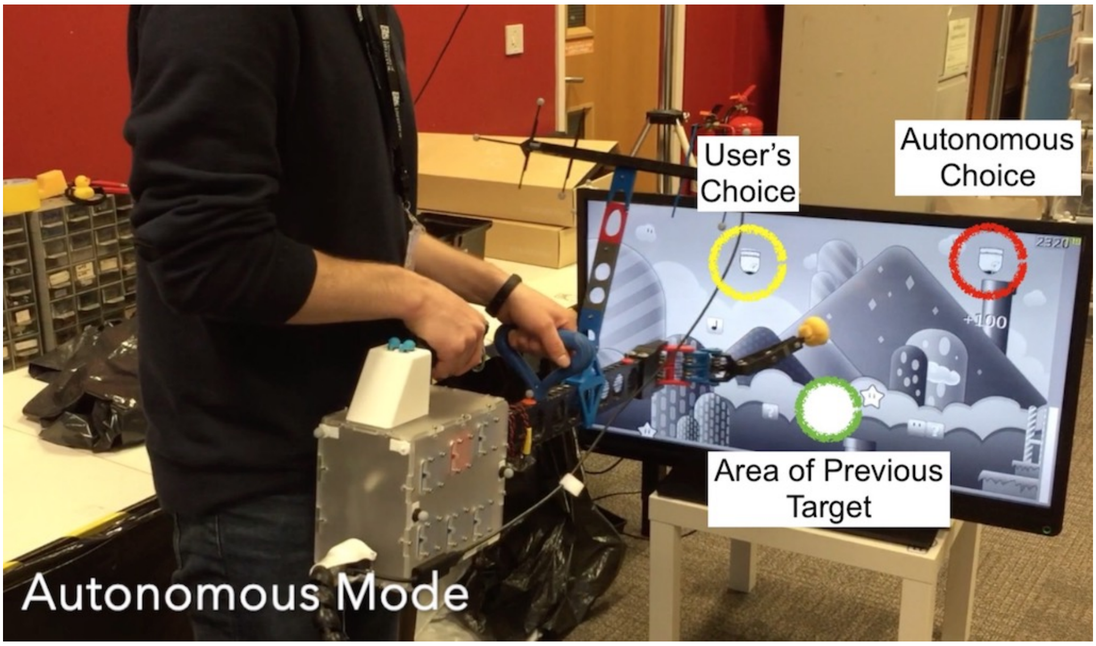


Figure 3.21: ***Conflicting Plans in the Autonomous Mode.*** *This figure shows a typical situation for a conflict between the user's and the robot's plan. After completing a previous target (green), there are two targets available with similar priorities. The user points the robot towards their chosen target (yellow) but the robot points towards its own choice (red) as it does not take into account the user's focus in the autonomous mode.*

speed. In those cases, the user rushed to the next target and moved the robot towards it, whilst the robot's tip moved to a different target (see Figure 3.21). This often resulted in missing both targets because the robot could not reach either of them in time.

Compared to slave mode, the gaze-dependent robot motion appeared to be less of a problem but rather beneficial. Recall that the robot aims for a target once it got fixated through the user's gaze. That way, the robot only moved when the focus shifted to another target definitively, rather than moving and jittering during visual exploration. This also cancelled the aforementioned feedback issue that occurred in slave mode. Another benefit of the cooperative mode is that it is more robust concerning occasionally occurring small time gaps of a few hundred milliseconds where the eye gaze could not be detected by the eye tracker. In slave mode, these instances would lead to a delayed or jittery following of the robot's tip along the gaze trajectory. However, since the cooperative mode is fixation-based and takes into account task knowledge, it remains locked to a chosen object even when there is no gaze data available for short periods.

Compared to the autonomous mode, the robot seemed to react more predictable in the cooperative mode. Often, the user's gaze was one step ahead in the task, i.e. while the robot finished one target the subject already fixated the subsequent selection (see Figure 3.22). This made task solving more fluent even though their choices were not always as efficient as the choices made by the robot during autonomous mode.

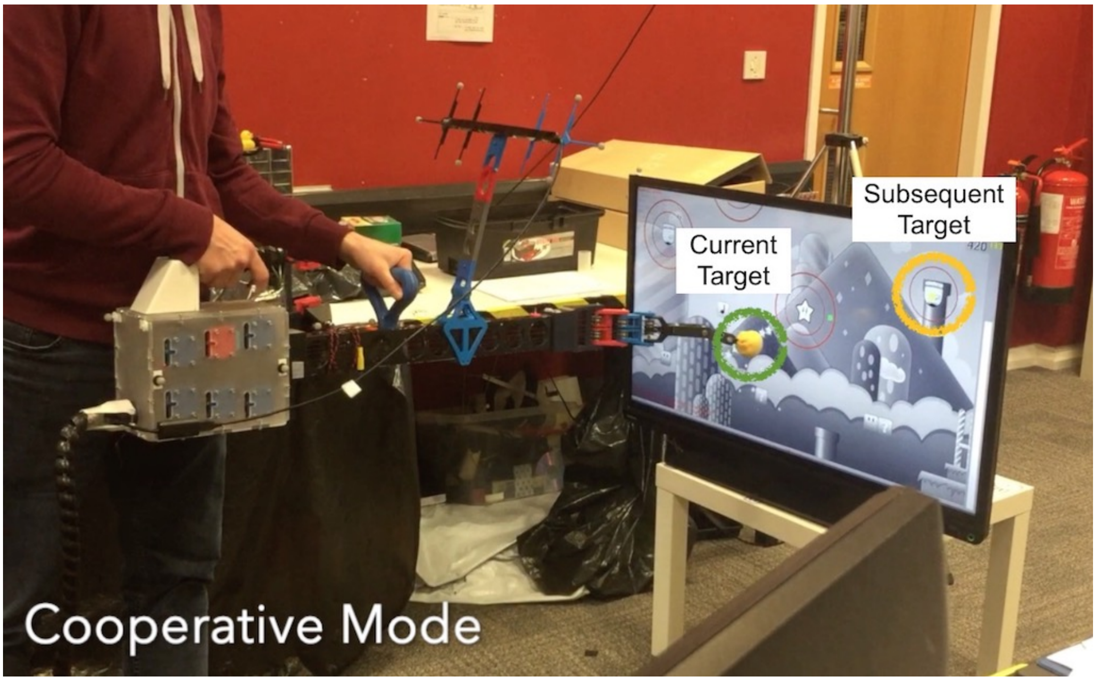


Figure 3.22: ***Plan Adaptation in the Cooperative Mode.*** *This figure shows a typical task solving strategy with the robot in the cooperative mode. While the robot is finishing the current target (in greed circle), the user already fixates a subsequent target (circled yellow). The robot follows once the previous target is completed.*

3.7 Discussion of Attention Study

This study aimed to explore how estimating user attention can benefit collaboration of a handheld robot with a human and in particular, how the incorporation of attention affects teamwork performance and users' perceived task load.

In addressing research question **Q1**, two gaze-based attention models were introduced where one additionally takes into account the robot's task knowledge. These modes were tested against the robot in a fully autonomous mode and for the scenario where the same job was done manually.

Regarding **Q2**, the modes were analysed for performance and task load. We found that using the primitive approach, where the robot is following eye gaze, performance decreases while the workload is increased in comparison to the manual case. One explanation might be the lack of accuracy for eye tracking and that peripheral view could not be taken into account. Moreover, it was observed that the robot's motion towards the focus of gaze influenced the gaze behaviour which in turn caused more tip motion. This sometimes led to off-set errors and jittering during the use of the slave mode which was also reported in the feedback.

In contrast, we found that the attention-based cooperative behaviour of the robot exceeds the manual analogue in terms of performance and decreases workload, which makes it appear similar to the autonomous mode. This statement, however, is constrained to the requirement of a certain level of temporal demand for the effect to become apparent. The specific speed constraints for each mode are subject to further investigation.

When the cooperative mode is compared to fully autonomous, the statistics do not yield a difference in terms of task performance. However, we note that the autonomous mode is modelled with full omniscience which might not be possible outside the lab environment. For example, the robot might know which objects are task-relevant, while only the user knows the right sequence of task steps. In that case, the attention model can inform the robot's control unit. Furthermore, qualitative feedback indicates that the robot is more predictable in the cooperative mode, making it more preferable.

We note that the terms *autonomous* and *cooperative* were used interchangeably in previous work [60] as the robot's cooperation was solely based on the novel autonomous task decisions, which were derived from task states but did not take into account user states. In this chapter, we distinguish between autonomous and cooperative behaviour as we start to see the benefits of the robot's adaptation to a user's plans. While the difference in performances is small, we suggest that the feeling of working together is important for humans in collaborative work with machines, hence why this will be further addressed in the subsequent Chapter 4

3.8 Chapter Conclusion

This chapter presents a system for estimating user attention for handheld collaboration. Gaze information and task knowledge are used as two factors of the attention model. First, we developed a gaze model that estimates a 3D gaze ray from 2D gaze information and motion capturing. The gaze model was used to inform a subsequent attention study in which attention incorporated gaze-based behaviours with varying levels of robot's task knowledge. Performance and task load in these modes were compared against a fully autonomous mode and a manual mode.

Results indicate that cooperative behaviour is more effective than completing the task manually for cases where there is a high demand for speed. We also found that task load is reduced when cooperative behaviour is based on both task knowledge and eye gaze, i.e. with incorporated attention.

These findings can be seen as a big step in human-robot collaboration. As modern tools become more complex and intelligent, their adaptation to users becomes more relevant and capturing user attention is a first step towards a tool that understands the users' plans.

The information of user attention could be used as a starting point for a more sophisticated intention prediction system. The attention model is useful to determine low-level decisions, e.g. in pick and place applications or other instances where users make quick decisions, which the robot could in turn adapt to. More complex tasks that involve related parts would require an intention model that can link task objects to a high-level solution strategy. This new research question will be the subject of the subsequent chapter.

Chapter 4

Rebellion and Obedience: The Effects of Intention Prediction in Cooperative Handheld Robots

This chapter presents a model for user intention prediction and its effect on collaboration with handheld robots. It employs the attention model that was introduced in Chapter 3 as main predictor. While both Chapter 3 and this chapter focus on single-user applications, in Chapter 5 we will look at multi-user applications. The outcomes of this chapter are summarised in the supplementary video[1] (scan QR code).

The main results of this chapter were presented at the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) [210].

4.1 Introduction

A handheld robot shares properties of powered hand tools while being enhanced with autonomous motion as well as the ability to process task-relevant information and user signals (see also Section 2.2 and 2.3). Since the robot holds task knowledge, such a system could help cutting workers' training times, as less user expertise is required for task solving. At the same time, the robot benefits from humans' natural navigation and obstacle avoidance capabilities. While this can arguably be beneficial for task performance, the high proximity between user and robot also leads to codependencies that create the need of communication methods between user and robot for efficient collaboration.

[1] Chapter 4 Summary Video: https://youtu.be/H245WdJpNpE

Earlier work in this field explored robot-human communication for improved cooperation [58, 60]. Such one-way communication of task planning, however, is limited in that the robot has to lead the user and as users exert their will and decisions, task conflicts emerge. This, in turn, inflicts user frustration and decreases cooperative task performance (see Section 2.2).

Efforts towards involving user perception in robot's task planning were made in our recent work on estimating user attention described in the previous chapter (see Chapter 3). This allows the robot to estimate the user's point of attention via eye gaze in 3D space during task execution. While the estimation of users' visual attention helps just-in-time planning, we lack an intention model which would allow the robot to infer the user's goal in the proximate future and go beyond reacting to immediate decisions only.

The results from Chapter 3 show that an estimate of users' visual attention informs the robot about areas of users' interest. This is particularly helpful for the robot's control during tasks with high temporal demand. As opposed to an intention model, the attention model would react to the current state of eye gaze information only, rather than using its history to make predictions about the user's future goals. What is necessary for cooperative solving of more complex tasks like assembly where there is an increased depth of subtasks is a system that goes beyond reacting to immediate decisions, i.e. a system that predicts further than one step ahead.

In recent years, promising solutions for intention inference have been achieved through observing user's eye gaze [74], body motion [179] or task objects [123] and affordance [103]. An extensive review of methods for intention prediction is presented in the background Chapter 2 (Section 2.6)

Using eye gaze as a predictor for actions is based on findings from studies that link gazing behaviour to future actions. Land et al. [112, 113] found that fixations towards an object often precede a subsequent manual interaction by a few hundred milliseconds, depending on the type of interaction. Experiments with virtual [12] and physical [169] tasks show that humans gather information through vision just in time rather than planning and memorising a task solution strategy upfront. Notably, Huang et al. [75] used gaze information from a head-mounted eye tracker to predict customers' choices of ingredients for sandwich-making using an SVM as a prediction model with 76% accuracy. It was later used as a basis for a robot's anticipatory behaviour, which led to more efficient collaboration compared to following verbal commands only [74]. This supports our motivation of replacing explicit commands with control through intention anticipation in the context of handheld robot collaboration.

Another aspect of action prediction is the current task state. For example, Liu et al. [122, 123] use information about object locations and about the sequence of previously

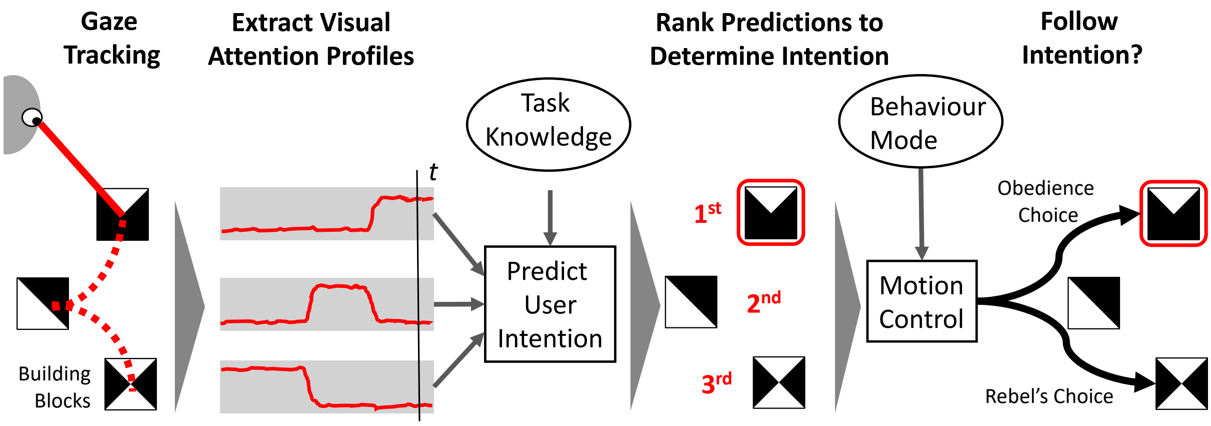


Figure 4.1: ***The Intention Prediction Model****. This shows an overview of the modules of the intention prediction model and its application in the validation task. The system captures the users' eye gaze during their decision process and derives a visual attention profile for each object in the scene, e.g. in this case, the building blocks. The Prediction model uses gaze data and task knowledge to calculate the intention probabilities for each object. Finally, the handheld robot uses the predictions to bias its behaviour during task solving. Depending on the behaviour mode, it follows the intention (obedience choice) or chooses a different target (rebel's choice) in the validation experiment.*

used objects to derive an estimation for a subject's subsequent step in block assembly tasks. To varying extents, this task knowledge is already available to the majority of intelligent tools (Section 2.3) because it is required for their autonomous control. Using this data to further support a prediction system could improve accuracies in comparison to mono-modal models.

The above methods improve cooperation in a turn-taking human-robot collaboration setup. However, we lack knowledge about their effect on cooperation performance within a shared control setup such as we face with handheld robots. To the best of our knowledge, such a scenario was never tested with a system for intention prediction that involves both human behaviour cues and task knowledge. Thus, the question remains open whether there is a model which suits the setup of a handheld robot, which is characterised by close shared physical dependency and a *working together* rather than a *turn taking* cooperative strategy. Hence, this is explored in this chapter, which is guided by the following research questions:

Q3 How can user intention be modelled in the context of a handheld robot task?

Q4 To what extent does intention prediction of users affect the cooperation with a handheld robot?

For our study, we use the open robotic platform, introduced in [59], combined with the eye tracking system introduced in Section 3.2.2. The 3D CAD models of the robot design are available from [1]. Within a simulated assembly task, which was inspired by [12], eye gaze information is used to predict subsequent user actions. Figure 4.1 shows an overview of the proposed system.

At the core of this chapter is the validation of the intention model. This is a particularly challenging problem, because of the physical codependencies that emerge from the shared-control characteristics of the handheld robot. Namely, the robot cannot reach a target without the user's help and vice versa. Now, when the robot uses intention anticipation to bias its decisions towards an aim for interaction, it is hard to judge, which party actually took the lead. The fact that the actions of the two parties converge, i.e. the large scale tactical motion of the user and the small-scale pointing movements of the robot, does not immediately imply that this is due to a correct prediction of the user's goal. This is due to possible feedback loop because one could ask:

- Did the robot indeed follow the user's will?
- Or did the user adapt their plans towards the robot's (perhaps faulty) best guess?

In an attempt to overcome this obstacle, this chapter introduces the comparison of obedience and rebellion as a paradigm to validate the prediction of intention. To clarify the reasoning behind this concept, the reader shall be exposed to the following thought experiment. Suppose the robot's majority of predictions were correct. Such a robot would have the power to choose to obey or to rebel against the user's plans. As suggested in [58], disobedience is a cause of users' frustration. Comparing frustration levels within these two conditions serves as implicit evidence for successful predictions, i.e. the accuracy of the intention model. For this reason, rebellion and obedience are used as parameters to test the handheld robot's anticipation capabilities.

The two principal parts of this chapter consist of modelling user intention, followed by the aforementioned validation procedure. The remainder of this chapter is organised as follows: The development of the intention prediction model and the acquisition of training data through experiments are described in Section 4.2. Section 4.3 and 4.4 outline strengths and weaknesses of the model and assess its accuracy, which is further discussed in Section 4.5. The model is then applied and tested in practice through an assistive pick and place task in Section 4.6 to 4.8, where the model is used for the robot's anticipatory behaviour. The chapter closes with conclusions in Section 4.9.

The contribution of this chapter is an intention prediction model with real-time capabilities that allows for human-robot collaboration through online plan adaptation in assistive tasks. Contribution details are summarised in the following list:

- We propose an online intention model, which predicts users' interaction location targets based on eye gaze and task states.
- For data collection and model validation, we propose an experimental setup of a block copying task to emulate an example of assembly.

- In the absence of universally accepted psychophysical metrics, we propose to measure the frustration induced through the robot's rebellion to validate intention predictions.

4.2 Prediction of User Intention

In this section, we describe how intention prediction is modelled for the context of a handheld robot based on an assembly task. The first part is about how users' gaze behaviour is captured and quantified within an experimental study. In the second part, we describe how this data is converted into features and how these were further filtered through task knowledge and used to predict user intent.

The first challenge in this work is to design an example task for experimental purposes that covers the requirements needed for an investigation of the research questions **Q3**/**Q4**. The core of this research is to understand users' gazing behaviour during complex task execution with one of the main assumptions being that task-relevant objects receive users' attention prior to interactions. For example, suppose the robot is used in an assembly task for manufacturing. Then the user might decide to pick a nut, required for a threaded rod. According to Land et al. [112] and Huang et al. [75], it can be expected that this interaction is presided by gaze fixations at the nut, which in turn allows for a prediction of interaction or at least its selection, given a set of options.

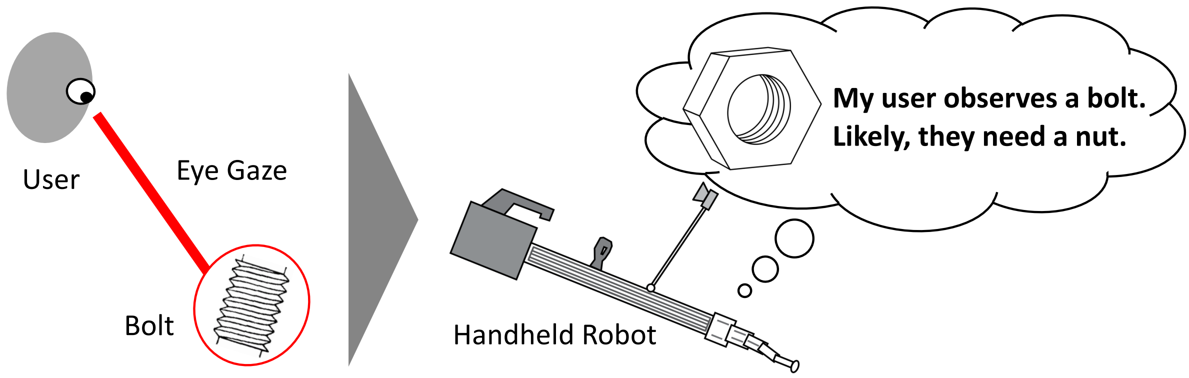


*Figure 4.2: **Linking Task Objects for Predictions.** This demonstrates the new prediction paradigm introduced in this study. The user gazes at a a task object (e.g. a bolt) and the robot predicts the need of a complementary object (e.g. a nut).*

The concept of this work goes one step further and tests whether this assumption extends to objects linked to another subtask. Specifically, for the aforementioned assembly example, one could assume that users decide on picking a nut when they observe a counterpart at the assembly object, where one is required, say a threaded rod (Figure 4.2). Arguably, a cooperative robot should be able to use this task knowledge to link fixations of related objects, e.g. the robot should prepare for nut picking when the user observes a threaded rod. This could result in more accurate predictions at an earlier point in time.

For this reason, a testing setup is required that fulfils the following criteria:

- It needs to be diverse enough to allow for a broad range of task solution strategies.
- It needs task-related dependencies between subsets of task objects.

Furthermore, the intention model requires data about object fixations and how the objects are related to each other. The following describes the implementation of these criteria concerning both the experimental setup and the intention model.

4.2.1 Data Collection

We chose a simulated version of a block copying task, which has been used in the context of work in hand-eye coordination [12, 169], but was redesigned to fit the purposes of a handheld robot setup. As this study focuses on the user's interaction with the robot rather than the robot's interaction with objects the proposed setup refrains from physical grasping of objects. Instead, participants use the robot to move blocks that are simulated and displayed on an LCD TV display. The display is integrated with the motion capturing system and so the robot's interactions with the screen surface can be detected without a need for additional sensors.

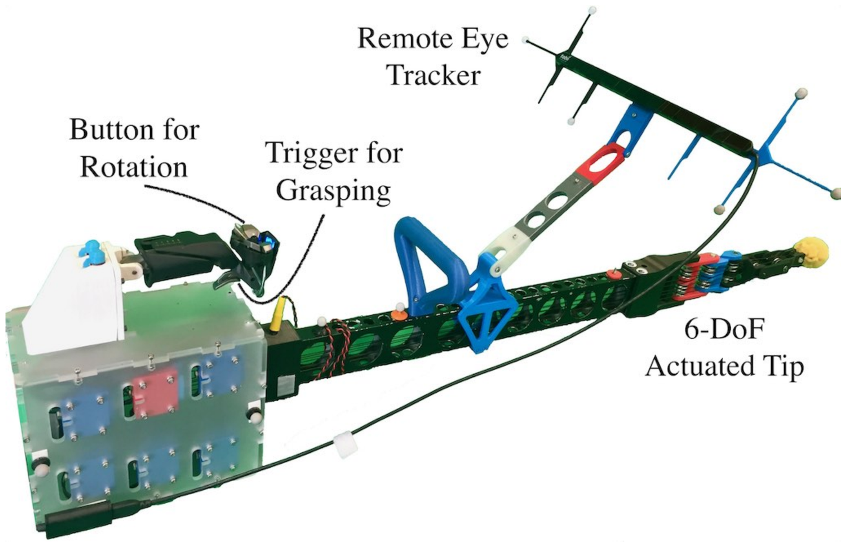


*Figure 4.3: **Handheld Robot used in this Study.** It features a set of input buttons and a trigger at the handle and a 6-DoF tip [59]. The intention prediction model is based on user perception through gaze tracking, which was introduced in Chapter 3.*

Participants of the data collection trials were asked to use the handheld robot (Figure 4.3) to pick blocks from a stock area and place them in the workspace area at one of the associated spaces indicated by a shaded model pattern. Hence, for a given block type, the relationship between the piece in the stock area and the pattern spot where it needs to be placed is equal to matching parts in an assembly process (e.g. as demonstrated in Figure 4.2). While the user decides on which part of the pattern to complete, their gaze information could already inform the intention model concerning a subsequent selection of stock part.

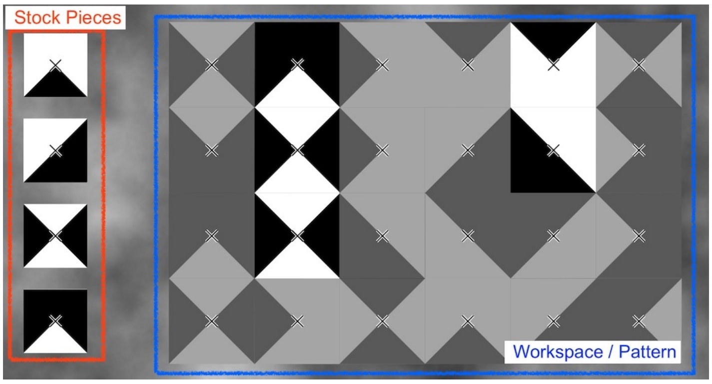
Inspiration for the task was drawn from the block-copy task presented in [12], where coloured blocks are used which subjects use to copy a pattern. However, for this study, the block design was changed from colours to grey-scale patterns. This design adds complexity due to the demand for matching orientation and eliminates any problems related to colour blindness in participants. Similar designs can be found in tasks for IQ tests [147]. Tasks in this field are characterised by high complexity, while the rules can be learnt swiftly. These properties are beneficial for the presented study as well, since subjects face high cognitive demands at the level of decision making, while training times remain short.

For the initial data collection experiment, the robot remained motionless to avoid distraction and the risk of user decisions being influenced by the robot's motion. Using the intention model for robotic motion will be subject to later validation (see Section 4.6). An overview of the task can be seen in Figure 4.4a and example moves in Figure 4.4b.

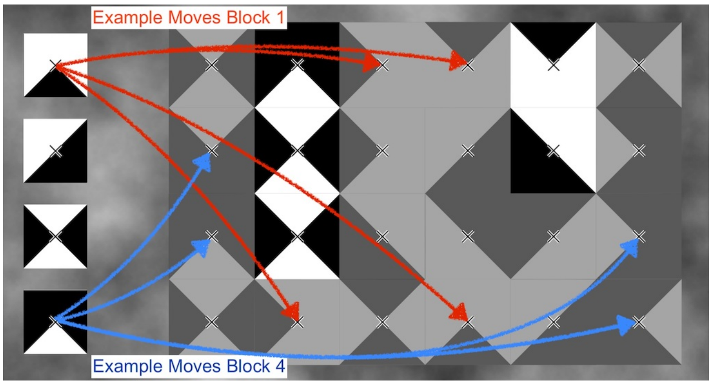
To pick or place pieces, users have to point the robot's tip towards and close to the desired location and pull/release a trigger in the handle. The position of the robot and its tip is measured via a motion capturing system[2]. A detailed description of the motion capturing setup is presented in the previous chapter (Section 3.2.2 and Figure 3.7). Another button in the handle allows the user to rotate a grabbed piece. The handle of the robot houses another button which can be used to rotate the grabbed piece. The opening or closing process of the virtual gripper is animated on the screen. If the participant tries to place a mismatch, the piece goes back to the stock and has to be picked up again. Participants are asked to solve the task swiftly and it is completed when all model pieces are copied. Throughout the task execution, we kept track of the user's eye gaze using a robot-mounted remote eye tracker in combination with a 3D gaze model from Chapter 3. Figure 4.4c shows an example of a participant solving the puzzle.

For the data collection, 16 participants (7 females, $m_{age} = 25$, $SD = 4$) were recruited to complete the block copy task, mostly students from different fields. Participation was voluntarily and there was no financial compensation for their time. Each completed one practice trial to get familiar with the procedure, followed by another three trials for data collection, where stock pieces and model pieces were randomised before execution. The pattern consists of 24 parts with an even count of the 4 types. The task starts with 5

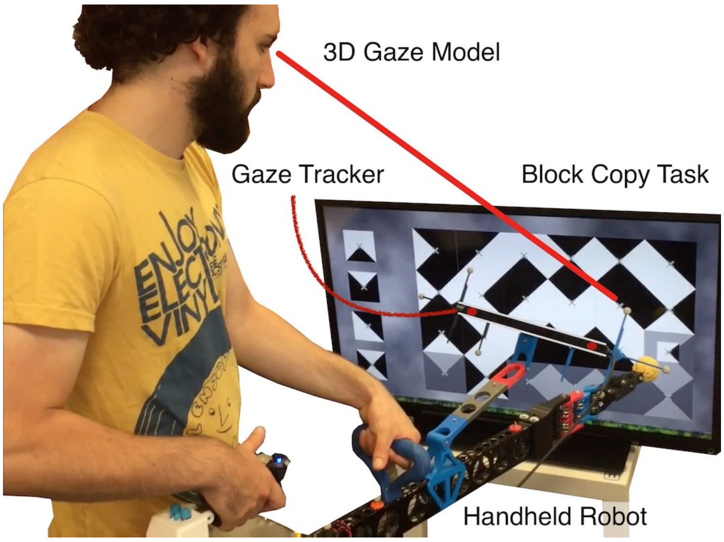
[2] OptiTrack: optitrack.com, 7 camera setup (model Flex3)



(a) Layout of the block copy task on a TV display. The area is divided into stock (red) and workspace (blue). The shaded pattern pieces in the workspace area have to be completed by placing the associated pieces from the stock using the real robot. Some blocks are pre-completed at the beginning of the task to break trivial completion strategies (e.g. line-by-line filling).



(b) This shows a couple of possible example moves for block 1 and 4. Using the robot, a piece from the stock (left column) has to be moved to an associated piece in the pattern (shaded blocks) and match the model's orientation to complete it.



(c) This picture shows a participant within our user intention prediction study who solves the assembly task and is about to decide where to place the currently held block. Using the eye tracker, the prediction system extracts the user's gaze pattern, which is used for action prediction.

Figure 4.4: ***Overview of the Block Copy Task****. Layout and example moves are shown in (a) and (b). (c) Shows an experiment participant during task execution.*

pre-completed pieces to increase the diversity of solving sequences leaving 19 pieces to be completed by the participant. That way, a total amount of 912 episodes of picking and dropping were recorded.

With 912 recorded episodes, 4 available stock pieces and 24 pattern parts, we collected 3,648 gaze history data points for stock parts prior to picking actions and 21,888 for pattern pieces.

4.2.2 User Intention Model

In the context of our handheld robot task, we define intention as the user's choice of which object to interact with next, i.e. which stock piece to pick and on which pattern field to place it. The studies on human gaze behaviour (introduced in Section 2.6.3) inspired the use of gaze data for action prediction and form the basis of our assumptions for the intention model listed in the following:

A1 An intended object attracts the users' visual attention prior to interaction.

A2 During task planning, the users' visual attention is shared between the intended object and other (e.g. subsequent) task-relevant objects.

Moreover, as noted in [58], a mismatch between the robot's plans and the user's intention inflicts user frustration. Hence, with regards to the model's experimental validation (Section 4.6), we also assume that

A3 If the predicted intention is the true intention, a robot that rebels against following the predicted goals induces user frustration.

Our method is constrained by the assumption that full task knowledge is available to the system. This includes information about task objects' positions and their relationships such as task-specific matching.

As a first step towards feature construction, the gaze information for an individual object was used to extract a Visual Attention Profile (VAP), which we define as the continuous probability of an object being gazed. Let $\boldsymbol{x_{gaze}}$ be the 2D point of intersection between the gaze ray and the TV screen surface and $\boldsymbol{x_i}$ the 2D position of the i-th object in the screen. Then the gaze position can be compared to each object using the Euclidean distance:

$$d_i(t) = ||\boldsymbol{x_{gaze}} - \boldsymbol{x_i}||. \tag{4.1}$$

As a decrease of d implies an increased visual attention, the distance profile can be converted to a VAP using the following equation:

$$P_{gazed,i}(t) = \exp(\frac{-d_i(t)^2}{2\sigma^2}). \tag{4.2}$$

Here, σ defines the gaze distance resulting in a significant drop of P_{gazed}, which is set to 60 mm based on the pieces' size and tracking tolerance, which was identified in the eye tracker accuracy study of the preceding chapter (Section 3.3). The intention model uses the VAP of the interval $T_{anticipate} = 4$ s before the point in time of the prediction. This is to account for the average duration of subtasks (see Section 4.3). When tweaking the window size, for the intention model, turned out that data points beyond 3.5 s would not add much to the accuracy of the models, whereas real-time prediction becomes slower due to increased input dimension. For this reason, 4 s were chosen as a compromise between the two. Due to the data update frequency of 75 Hz, the profile is discretised into a vector of 300 entries (see example in Figure 4.5).

The prediction for picking and placing actions was modelled separately as they require different feature sets. As mentioned above, earlier studies about gaze behaviour during block copying [12] and assembly [145] suggest that the eye gathers information about both what to pick and where to place it prior to initialising manual actions. For this reason, we combined pattern and stock information for picking predictions for each available candidate, resulting in the features selection:

- F_1 The VAP of the object itself.
- F_2 The VAP of the matching piece in the pattern. If there are several, their VAPs are combined using the element-wise maximum function.

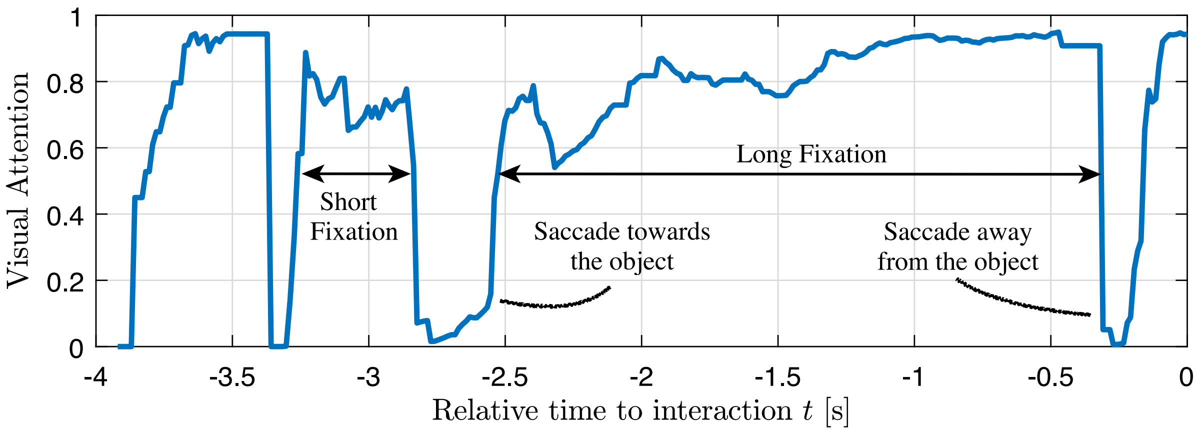


Figure 4.5: ***Visual Attention Profile (VAP).*** *This is an example for a VAP of a single object in the scene. It demonstrates the characteristic changing visual attention over time within the anticipation window of the prediction model for an individual object.*

This goes in line with our assumptions **A1**, **A2**. Both features are vectors of real numbers between 0 and 1 with a length of $n = 300$. For the prediction of the dropping location,

A2 is not applicable as the episode finishes with the placing of the part, hence why only F_1 (a vector with length $n = 300$) is used for prediction. Note that this feature contains implicit information about fixation durations as well as saccade counts.

An SVM [68] was chosen as a prediction model as this type of supervised machine learning model was used for similar classification problems in the past, e.g. [75]. A strength of SVM models is that they generalise well for complex problems, even for small training sets [68]. Furthermore, the calculation for predictions is fast due to its linear mathematical characteristics. The model analysis further down will also cover how the SVM compares to other prediction methods, such as ANN and Logit model. Model design and training were done using the MATLAB *Deep Learning* and *Statistics and Machine Learning* toolboxes.

We divided the sets of VAPs into two categories, one where the associated object was the intended object (labelled as `chosen = 1`) and another one for the objects that were not chosen for interaction (labelled as `chosen = 0`). Training and validation of the models were done through 5-fold cross-validation [99].

The accuracy of predicting the `chosen` label for individual objects is 89.6% for picking actions and 98.3% for placing. However, sometimes the combined decision is conflicting e.g when several stock pieces are predicted to be the intended ones. This is resolved by selecting the one with the highest probability $P(\texttt{chosen} = 1)$ in a one-vs-all setup [182]. This configuration was tested for scenarios with the biggest choice, e.g. when all 4 stock parts (random chance = 25%) would be a reasonable choice to pick or when the piece to be placed matches 4 to 6 different pattern pieces (random chance = 17-25%). This validation set X_{valid} includes 540 picking samples and 294 placing samples. The one-vs-all validation of the SVM results in a correct prediction rate of 87.9% for picking and 93.25% for placing actions.

	Accuracy [%]
SVM	87.9
Attention Only	77.1
ANN	75.9
Logit	70.1

*Table 4.1: **Intention Prediction Performance.** This shows the accuracy of the tested prediction models, namely SVM, ANN and a Logit model. The SVM yields the highest accuracy of the prediction models and outperforms the straight forward "choose last attended object" approach.*

For a comparison to alternative prediction models, the same procedure was repeated using a Logit model and an ANN. The ANN has 2 interconnected hidden layers with the size of the input data (i.e. 300 nodes for each feature F_1, F_2). Table 4.1 shows that the SVM outperforms the model alternative. Furthermore, it outperforms the straight forward use of the attention model, i.e. using the latest fixated object as a prediction result. Further adjustment of the ANN's hyper parameters, e.g. adding nodes and layers, might improve its performance. However, this also bears the risk of overfitting. Therefore, we decided to proceed with the SVM as a basis for the intention modelling analysis.

4.3 Results of Intention Modelling: Quantitative Analysis

The analysis of the intention model's performance is divided into two parts, a quantitative analysis and a qualitative assessment. The quantitative analysis focuses on the accuracy of the intention model in comparison to established methods while the qualitative part outlines common reasons for correct performance or failures.

Having trained and validated the intention prediction model for the case where VAPs range over $T_{anticipate}$ prior to t_0, the time of interaction with the associated object, we are now interested in knowing to what extent the intention model predicts accurately at some prior time $t_{prior} < t_0$. To answer this question, we extend our model analysis by calculating a t_{prior}-dependent prediction accuracy where respective predictions are based on data from the time interval $[t_{prior} - T_{anticipate}, t_{prior}]$. Within a 5-fold cross-validation setup, t_{prior} is gradually decreased while predictions are calculated using the trained SVM model and compared against the ground truth at t_0 to determine accuracy. The validation is based on the aforementioned set X_{vaild} so that the random chance of correct prediction would be $\leq 25\%$. The shift of the anticipation window over the data set is done with a step width of 1 frame (13 ms). This is done for both the case of predicting which piece is picked up next as well as inferring intention concerning where it is going to be placed. For the time offsets $t_{prior} = 0$, 0.5 and 1 seconds, the prediction of picking actions yields an accuracy a_{pick} of 87.94%, 72.36% and 58.07%, respectively. The performance of the placing intention model maintains a high accuracy over a time span of 3 s with an accuracy a_{place} of 93.25%, 80.06% and 63.99% for times $t_{prior} = 0$, 1.5 and 3 seconds, respectively. In order to interpret these differences in performance, we investigated whether there is a difference between the mean duration of picking and placing actions. We applied a two-sample t-test and found that the picking time (mean = 3.61 s, $SD = 1.36$ s) is significantly smaller than the placing time (mean = 4.65 s, $SD = 1.34$ s), with $p < .001, t = -16.12$.

As the prediction model of the picking actions implements the novel aspect of adding the VAPs of related objects, its comparison to existing methods is of particular interest. Figure 4.6 shows a comparison of the proposed model (where both features F_1 and F_2 are used) against the case where F_1 is the single basis for a prediction, such as the model recently explored in [75]. It can be seen that both models well exceed the chance of picking randomly. Notably, the proposed model outperforms the existing one shortly after the subject ends the preceding move and presumably starts planning the next one. To further

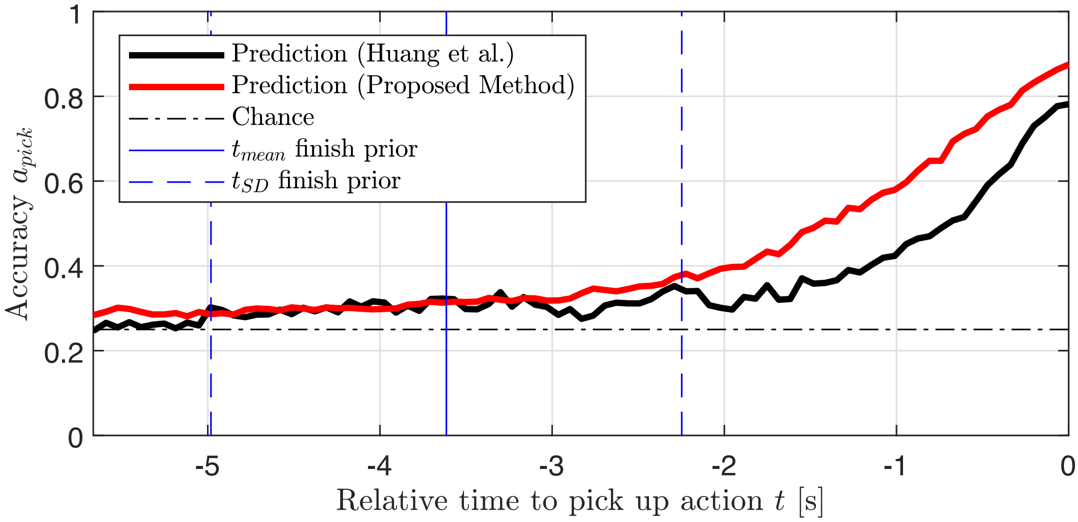


Figure 4.6: ***Prediction Performance of the Proposed Model and How it Compares to Existing Models.*** *This diagram shows the performance of predicting pick up actions averaged over 912 samples for two models: our proposed model (red) and an SVM (black), which is based on the feature F_1 only, such as proposed by Huang et al. [75]. It can be seen how both models perform better than chance (dashed black) and predict the actions with increasing accuracy as the prediction time t approaches the time of the action's execution $t = 0$. t_{mean} (with temporal SD t_{SD}) is the mean time of completing the last block and hence the earliest meaningful time of predicting picking as a subsequent action.*

investigate the effect of the chosen model on the prediction performance, a two-factorial ANOVA was applied where the prediction time t relative to the action and the model were set as the independent factors and the performance as the dependent variable, which reveals that the correct prediction rate of the proposed model is significantly higher ($p < .001$) than the one of the existing model. Moreover, the proposed model is able to predict intention more reliable (improvement from 68.34% to 87.94%) and at an earlier point in time than the existing method. The proposed model reaches the 60% accuracy mark at $t = -0.92$s, i.e. 470 ms earlier in time than the existing model with $t = -0.45$s (see Figure Figure 4.6). The prediction model performs fast enough to fit the time constraints set by the eye tracker (30 Hz), which allows for real-time predictions during task execution, as demonstrated in Figure 4.7.

4.4 Results of Intention Modelling: Qualitative Analysis

For an in-depth understanding of how the intention models respond to different gaze patterns, we investigate the prediction profile, i.e. the change of prediction over time, for a set of typical scenarios.

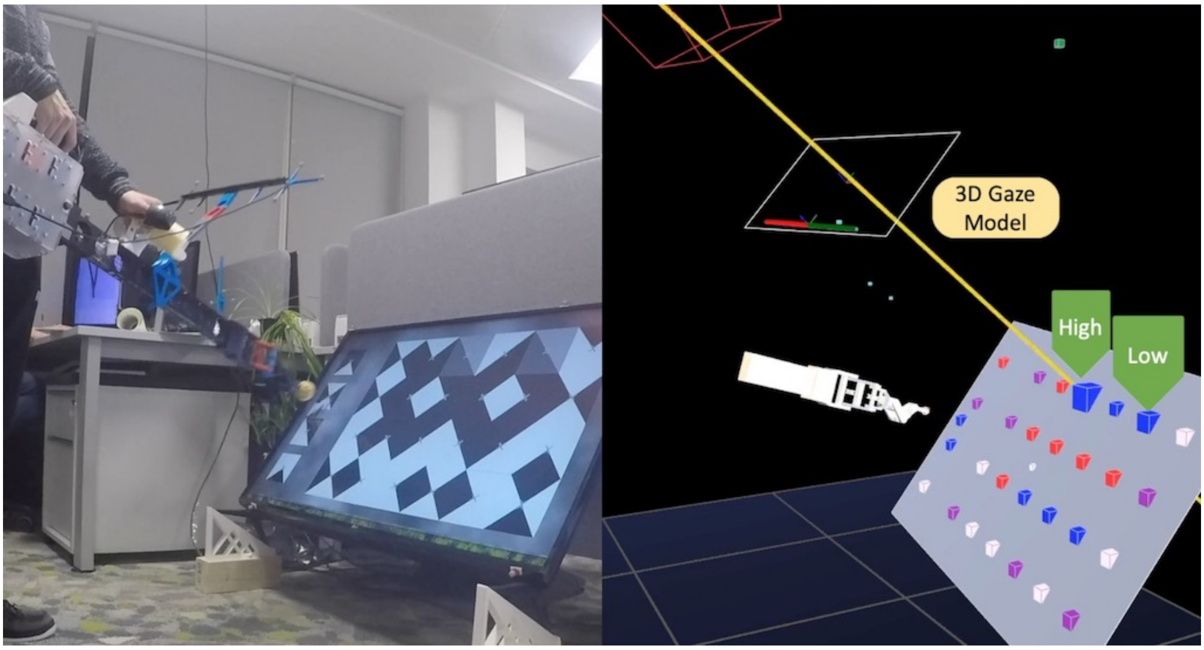


Figure 4.7: ***Real-Time Intention Prediction.*** *This shows an example for a prediction during the pick-and-place task. The picture on the left shows a snapshot of the experiment setup. To the right, is the visualised prediction for the different task objects (cubes). The probabilities are indicated by the objects' sizes. In this case, there are two matching alternatives and the model predicts a higher probability for the object that the user is currently gazing at. Notably, the user's eye gaze is one step ahead of the task, i.e. the user observes a spot to place the block in the future rather than focusing on the picking process. This enables the intention model to make predictions for the proximate future.*

4.4.1 One Dominant Type

A common observation was that the target object perceived most of the user's visual attention before interactions, which goes in line with our assumption **A1** and findings from [75]. An example of these *one-type-dominant* samples can be seen in figure 4.8a. Furthermore, note that the prediction remains stable for the event of a short break of visual attention, i.e. the user glances away and back to the same object (Figure 4.9b). This is a contrast to an intention inference based on the last state of visual attention only, which would result in an immediate change of the prediction. For the majority of these one type dominant samples both the picking and placing prediction models predict correctly.

Of particular interest are examples where the model made connections between objects of the same type, i.e. when the model combines gaze data from pattern space with the data from the matching stock part. An example for such a case can be seen in Figure 4.8c, which shows the VAPs of a stock piece and its counterpart in the pattern right before the stock piece get selected by the user. Notably, the VAP of the stock piece alone is characterised by large time gaps that yield no attention. However, because the pattern counterpart is gazed at during those intervals, the model (correctly) calculates a high prediction value. This supports assumption **A2** and thus the proposed new paradigm of linking task-related objects for more accurate predictions.

4.4.2 Trending Choice

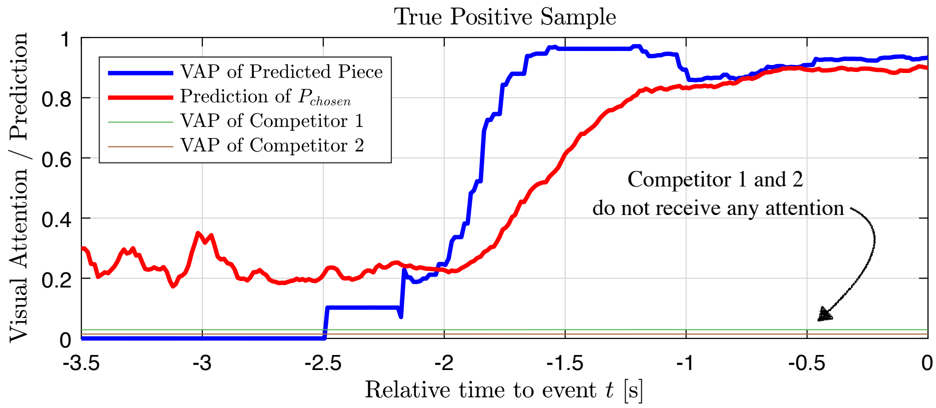
While the anticipation time of the pick-up prediction model lies within a second and is thus rather reactive, the placing intention model is characterised by a slow increase of likelihood during the task, i.e. it shows a low-pass characteristic. Figure 4.9 demonstrates that the model is robust against small attention gaps and intermediate glances at competitors, however, the model requires an increased time window to build up confidence.

4.4.3 Incorrect Predictions

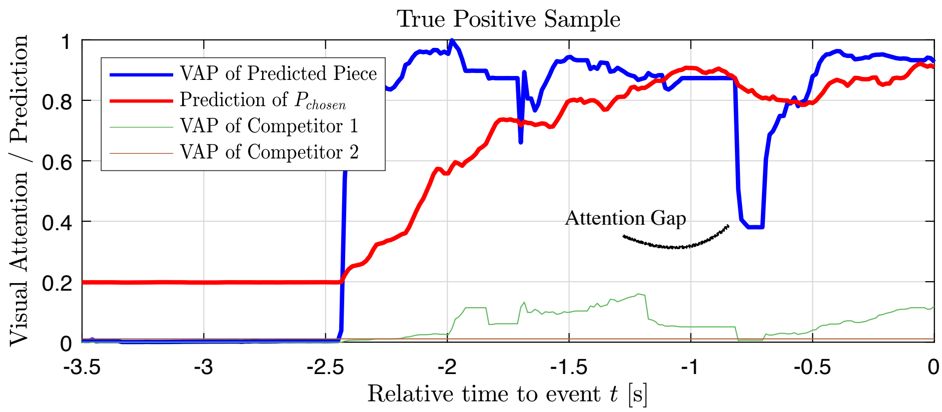
There is a number of reasons for occasional incorrect predictions with some examples displayed in Figure 4.10. Most commonly, a close-by neighbour received more visual attention and was falsely classified as the intended object. For example, Figure 4.10a shows how the user's gaze switches between a set of candidates without attending the final selection long enough for the model to make a correct prediction. Hence, one of the competitors is favoured by the model leading to an incorrect prediction.

In some rare cases, there were no intended fixations recorded for the candidate prior to the interaction, which can occur due to offset errors in the eye tracking system or due to the user making decisions based on peripheral view information (see Figure 4.10b).

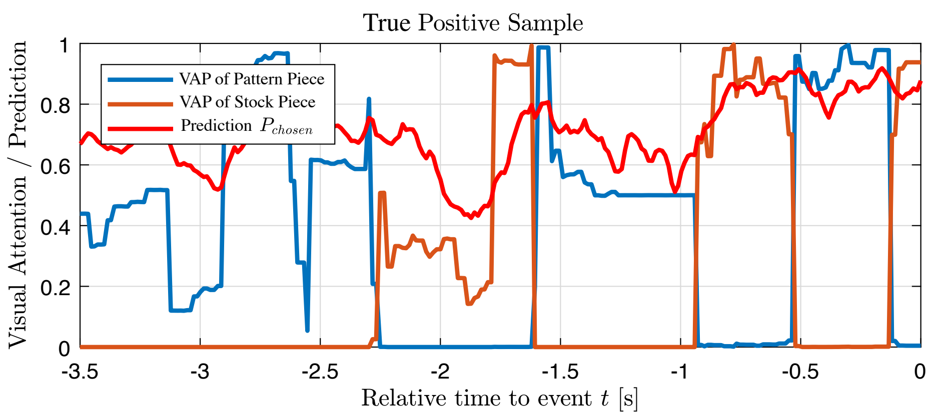
In other few samples that led to faulty predictions, the eye tracker could not recognise the eyes, e.g. when the robot is held so that the head was outside the trackable volume or the head angle range. In that case, the tracking system is unable to update the gaze model, which would lead to over or underestimation of perceived visual attention as can be seen in Figure 4.10c.



(a) One piece receives most of the user's visual attention prior to placing. The two highest competitors do not receive any attention, which leads to an accurate prediction.

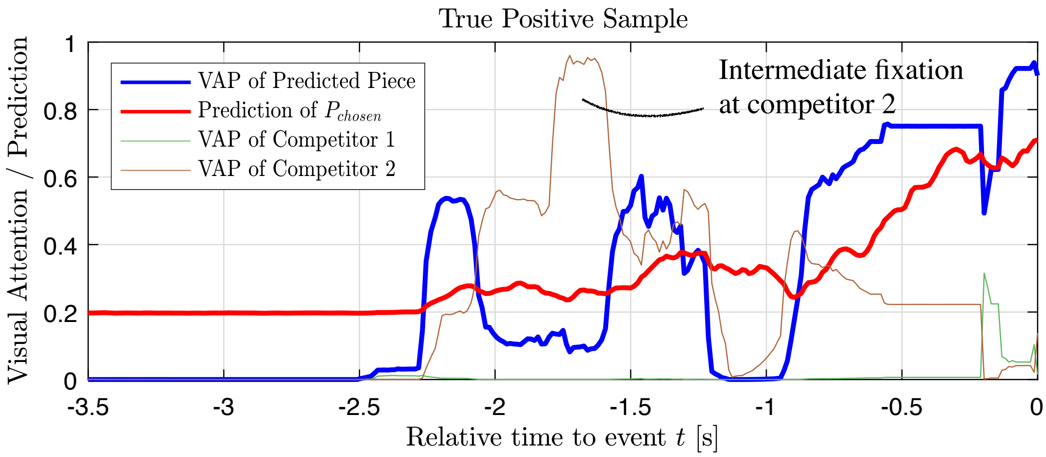


(b) This demonstrates the model's robustness for cases of small attention gaps. While the attention drops for a few hundred ms, the prediction remains stable. Dominance of visual attention with break gap: the prediction model maintains the prediction.

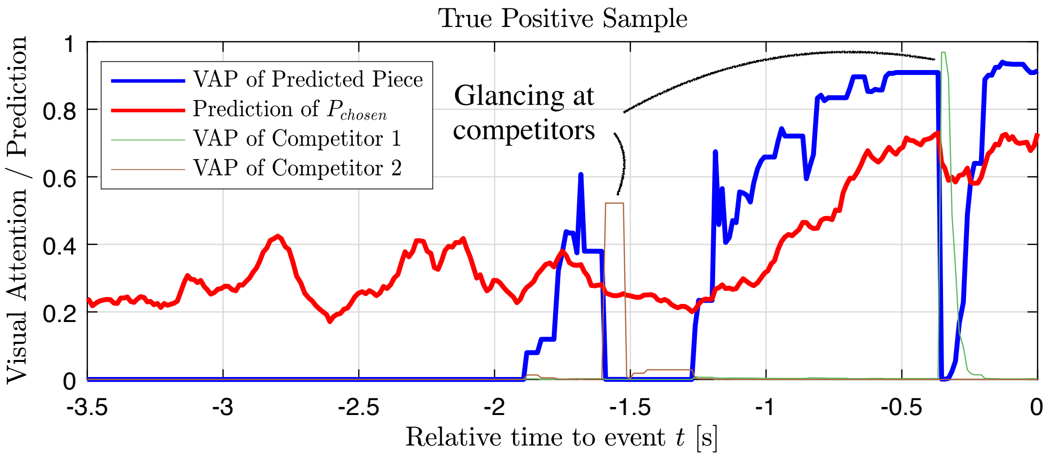


(c) User gaze alters between stock piece and matching workspace location. This was a common observation prior to piece selection. The fact that P_{chosen} is high towards the end, shows that the model is able to make a connection between VAPs of task-related objects.

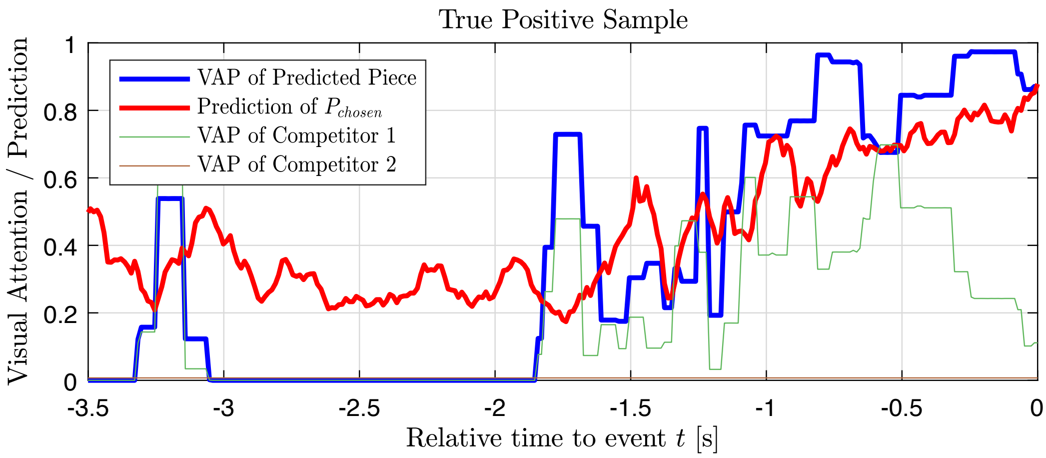
Figure 4.8: ***Correct Predictions with One-Type-Dominant Characteristic.*** *These diagrams show examples for correct predictions of one type dominant samples. They demonstrate how a long fixation time (blue) result into a high (red) probability (a), that the model is robust against a short duration of an absence of user attention (b) and how pattern glances and stock fixations are combined for predicting future interaction (c).*



(a) Increasing visual attention over time prior to object selection.

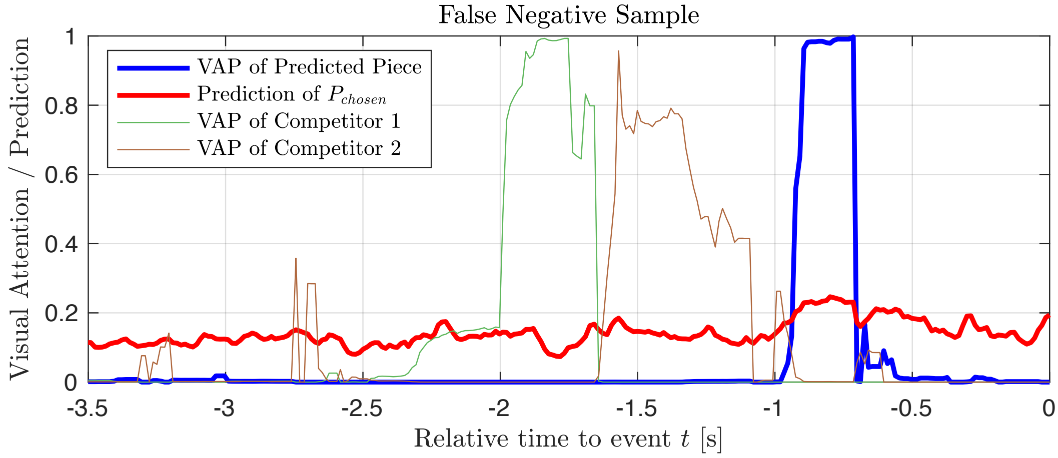


(b) This shows a typical characteristic of increasing user attention of the selected object over the time prior to the their decision. The model is robust against cases where competitors receive short intervals of attention (P_{chosen} remains stable).

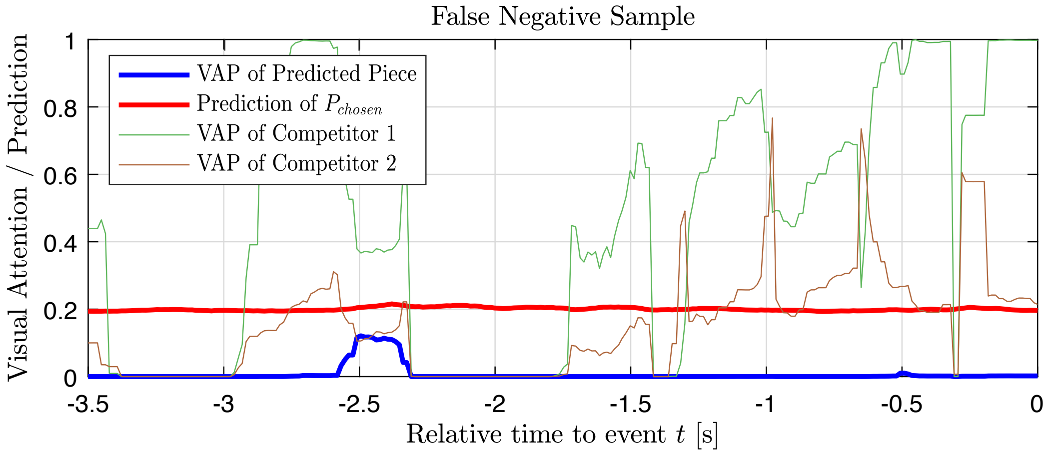


(c) The user's attention alters between a competitor and the selected piece. However, the attention shifts towards the selected object as the user approaches their decision.

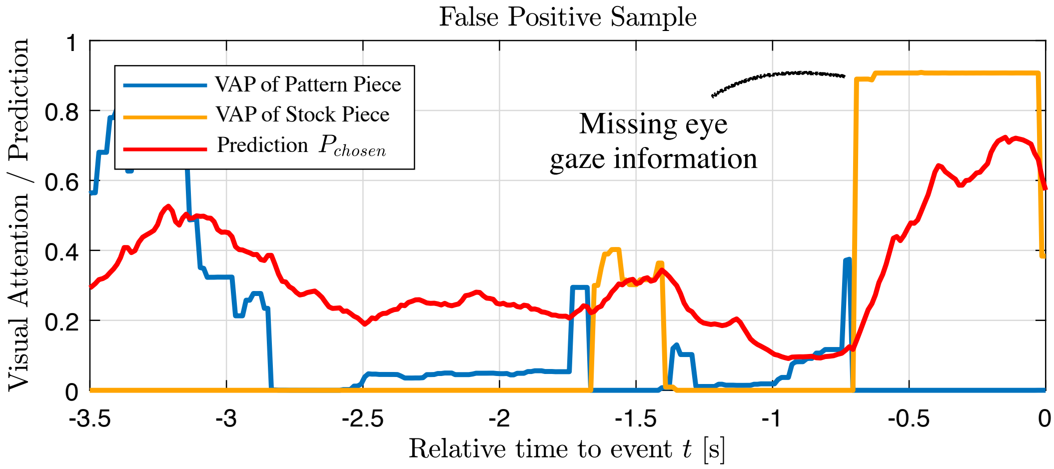
Figure 4.9: ***Correct Prediction with Attention Build-Up.*** *These examples illustrate how the visual attention (blue) of an object builds up during the user's decision process in which case the intention prediction (red) remains undecided (i.e. $P_{chosen} < 0.5$) for a longer time compared to the case where no competition receives fixations (see Figure 4.8).*



(a) Favoured competing choice: underestimation of the piece selected by the user, due to competitors receiving more attention in the immediate past.



(b) Underestimation due to missing visual attention for the intended object. This can occur when users use their peripheral view or due to an offset error of the gaze model.



(c) Overestimation due to missing gaze information. This can occur, e.g. for extreme head angles away from the gaze tracker.

Figure 4.10: ***Reasons for Incorrect Predictions.*** *This shows examples for incorrect predictions due to the lack of fixations (a). The predicted probability (red) for this object remains low despite being chosen by the user. (b) and (c) show situation that lead to the model's underestimation or overestimation, respectively.*

4.5 Discussion of Intention Modelling

Recall that the intention prediction modelling process aimed to derive an estimate of which object the user wants to interact with next so it could later be used to bias the robot's actions.

In addressing research question **Q3**, we proposed a user intention model based on gaze cues for the prediction of actions, which was assessed in a pick and place task. As a novel aspect introduced through this study, the predictions are not only based on saccades and fixation durations of an individual object but also those of related objects. In other words, assessing the attention on objects in the workspace helps to predict which piece outside the current workspace is needed next. When the subject turns their attention towards the piece, the model interprets this as a confirmation rather than the start of a selection process. This shows that the proposed model can create a link between attention spans towards related objects. This helps to cut the time required for the model to gather relevant gaze information and makes predictions more reliable than traditional models.

We showed that, within this task, the prediction of different actions has different anticipation times, i.e. the model allows predictions 500 ms before picking actions (71.6% accuracy) and 1500 ms prior to dropping actions (80.06% accuracy). This can partially be explained by the fact that picking episodes are shorter than placing episodes. More importantly, we observed that users planned the entire pick-place cycle rather than planning picking and placing actions separately. This becomes evident through the qualitative analysis, which shows altering fixations between the piece to pick and where to place it. That way, the placing prediction model can already gather information at the time of picking.

The fact that there is a difference between anticipation times for specific actions makes it hard to estimate to what extent the proposed approach generalises to new tasks. This is less surprising given that previous research showed that human gazing behaviour differs between tasks, concerning how far the eyes go ahead of the task [113]. However, the presented results indicate that actions can be predicted significantly earlier when task knowledge is used to link related objects and their gaze fixations. Arguably, this might generalise to other tasks where objects are linked through subtasks, e.g. pick and place, assembly or when a specific tool is required to interact with a part that has attracted the user's attention. This could be useful for assistive robotic systems beyond handheld robots as well, particularly when close collaboration is required.

The prediction model depends on a continuous gaze stream as an input. While the results show that the model is robust against small gaps in the data stream, predictions will fail for extreme cases, e.g. when users would turn their head far away from the gaze tracker. This rarely happens because most of the time, the user focuses on objects close to the tip and the sensor is aligned accordingly. Therefore, large head angles away from the sensor usually occur when the user aims for an object that is farther away. In most cases, the robot is able to capture the eye gaze and make a prediction as the user approaches the target and the gaze tracker gets realigned. This gives the robot enough time to react. Extending the system with additional sensors could improve the trackable range of the eye tracking system, which would make it more robust for these rare cases.

In terms of the system's limitations, we point out that it is unclear how well the model generalises and performs for new tasks as the differences in performances for the two example actions (picking and dropping) indicate that the model is task-dependent. Furthermore, the links between task objects were pre-defined, i.e. matching objects and their locations were labelled. The classification and localisation of task-relevant objects is a complex problem and thus subject to future work.

The results are encouraging for testing the prediction model in a real-time application. Therefore, we proceed with an experimental study where the intention model is used to bias the robot's action for cooperative behaviour. The robot's knowledge about the user's intention allows for a rebellion against user plans. The following section shows how this counter-intuitive twist can be used to validate the intention model in action.

4.6 Intention Prediction Model Validation

In the second part of this study, we validate the proposed intention model for the case where it is used to control the robot's behaviour and motion. While the aforementioned experiments and analysis demonstrate that the intention model is capable of predicting users' short term goals while having full control over the robot's tip, it is unclear whether this is true for the case where the robot reacts to these predictions. For example, users might adapt their intention to the robot's plans just by seeing it moving towards a target that might differ from their initially intended move. That way, labelling the robot's predictions as being correct or incorrect in the same way as we did in the previous study (Section 4.2) becomes invalid due to the lack of ground truth.

For this reason, we propose to assess the intention model indirectly instead by observing users' reactions to the predictions with a focus on frustration. We base our experimental validation on assumption **A3** and use frustration as a measure of correct predictions. For the case of the validation task, this means that the robot predicts where the user wants to reach next. The robot can then choose to avoid the predicted goal with the idea being that if this frustrates users, this could serve as evidence of the prediction being correct. Arguably, frustration can only be induced through disobedience if the robot rebels against the true intention. Equally, low user frustration is expected when the robot aims for predicted objects, given the prediction was correct. Therefore, a correlation between frustration and the robot's prediction-based obedience would implicitly validate the proposed intention model in action.

Rebellion in Artificial Intelligence (AI) is a subject that gained more importance over the recent years as robots have become more complex and have reached a point where social behaviour can be part of their interaction with humans [34]. For this reason, researchers have been working on a framework with various types of rebellion for AI agents. For example, Aha and Coman [6] introduce the concept of *outward-oriented* and *inward-oriented* rebellion. In the first case, the agent rebels against a person's (e.g. a user) behaviour and objects to it. The inward-oriented rebellion is characterised by the agent's actions conflicting with expectations towards it. A rebel type suitable for the presented study is rebellion through *non-compliance* [6]. This is a special case of the inward-oriented rebellion, i.e. where the robot reacts through behaviour: after a request, the robot refuses to respond with the requested behaviour. In the context of the presented experiment task, the request is substituted with the prediction of user's intention, i.e. the request is derived implicitly.

The following describes how the level of rebellion can be quantified and implemented for user studies.

4.6.1 Intention Affected Robot Behaviour

For the experimental validation of the intention model, we used the aforementioned (Section 4.2.1) block copy task and introduced an assistive behaviour to the robot, which is controlled based on the predictions of a user's intended subsequent move, i.e. which piece the user wants to pick up next or at which location they want to drop it. We created 3 different behaviour modes: *Follow Intention*, *Rebel* and *Random*. For each mode, the robot retreats to a crouched position while there is a low probability for each available target. When the probability of the target with the highest probability reaches a threshold, the robot reacts as follows:

- **Follow Intention:** The robot moves towards the target with the highest predicted intention. It estimates, which block the user wants to pick up next and assists with reaching it. After the block is picked, the intention model is used to estimate where the user wants to place it in the pattern and moves towards it (i.e. *Obedience Choice*, see Figure 4.1).
- **Rebel:** After a prediction, the robot moves towards the target with the lowest

intention prediction score (i.e. towards the *Rebel's Choice*) and thus avoids the one with the highest prediction value. This goes for the selection of stock pieces as well as for predicted target locations to place the block in the pattern (see Figure 4.1).

- **Random:** This is the reference mode where the robot randomly chooses which block to pick and where to place it. This is independent of predictions made, i.e. in some cases, the choice randomly goes in line with the user's intention.

For each mode, the target chosen by the robot will always be in the feasible set defined by the task. For example, the robot would never move towards a stock piece that is already completed throughout the model pattern.

As per assumption **A3**, we argue, that an observed reduction of user frustration in the *Follow Intention* mode compared to the *Rebel* mode would validate that the predicted user intention went in line with the true intention. A demo of the behaviour modes can be seen in the supplementary video[3] of this paper and on our webpage [1].

In rare occasions, none of the targets exceeds the probability threshold that invokes the robot's reaching actions. In that case, the robot chooses a random valid target after a small period of anticipation time to prevent a deadlock.

4.6.2 Experiment Execution

We recruited 20 new participants (6 females, $m_{age} = 26$, $SD = 4$) for the validation study of which 2 were later removed from the set for data analysis due to malfunctioning gaze tracking. The participants were mostly students from various fields, though an academic background was not required. 2 participants, i.e. 10%, have participated in one of the experiments in Chapter 3. However, due to extensive training with the robot prior to experiments, we do not expect any impact of this circumstance on the outcomes. Participation was on a voluntary basis as there was no financial compensation. Each was asked to first complete the task without the robot moving for familiarisation with the rules and the robot handling. This practice session was followed by 3 trials where, for each, the robot's behaviour was set to a different behaviour mode. The block pattern to complete as well as the order of the behaviour modes were randomised. Furthermore, 5 (out of 24) randomly chosen blocks were pre-completed to stimulate some diversity in solving strategies, e.g. to prevent repeated line-by-line completion.

The participants were told to solve the trial tasks swiftly and that their performance was recorded, i.e. they were encouraged to outperform their own preceding trials concerning completion time. They did not receive any information about the behaviour modes but

[3] youtu.be/H245WdJpNpE

were told that the robot will move and try to help them with the task. Each trial was followed by the completion of a *NASA* TLX form [63] and 3 min resting time (see questionnaire in Appendix D).

4.7 Results of Model Validation Study

The analysis of the experiment results is divided into two parts. The first part covers the quantitative analysis, which focuses on the statistical validation of the intention model, while the qualitative review summarises common reactions to the respective behaviour modes.

4.7.1 Quantitative Analysis

To determine the effect of the robot's behaviour mode on the subjects' frustration level, we performed an ANOVA with the mode as the independent variable and the frustration component of the TLX as a dependent variable. As the analysis yielded a significant effect ($p = .023$), it was further explored using a post-hoc pairwise t-test with applied Bonferroni correction. The frustration mean for the *Rebel* group was identified as being significantly higher than in the *Follow Intention* group ($p = .019$). No significant mean differences were found when comparing the *Random* group to the others. The results can be seen in Table 4.3 and Figure 4.11.

We extended our analysis to both, the combined TLX results, which serve as an indicator for perceived task load, and the measured performance, which is defined as the number of completed blocks per minute. However, an applied ANOVA did not yield an effect of the robot's behaviour model, neither on the combined TLX nor on the performance.

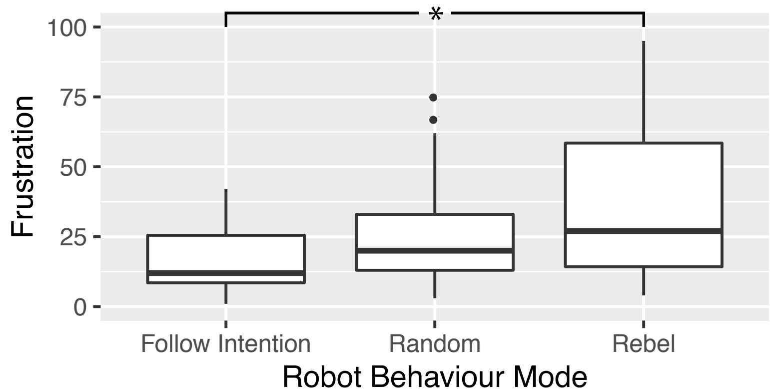


Figure 4.11: ***Frustration Results****. Perceived frustration from the TLX results for each of the tested behaviour modes. The mean values (see Table 4.2) of starred groups yield a significant difference (see Table 4.3).*

	Frustration Means and SDs
Follow Intention	16.6 (13.2)
Random	27.0 (21.7)
Rebel	37.2 (27.6)

Table 4.2: ***Mode-Dependent Frustration Levels.*** *Means and standard deviations of the frustration component of the TLX questionnaires*

	Follow Intention	**Random**
Rebel	$p = .019$ *	$p = .495$
Random	$p = .469$	-

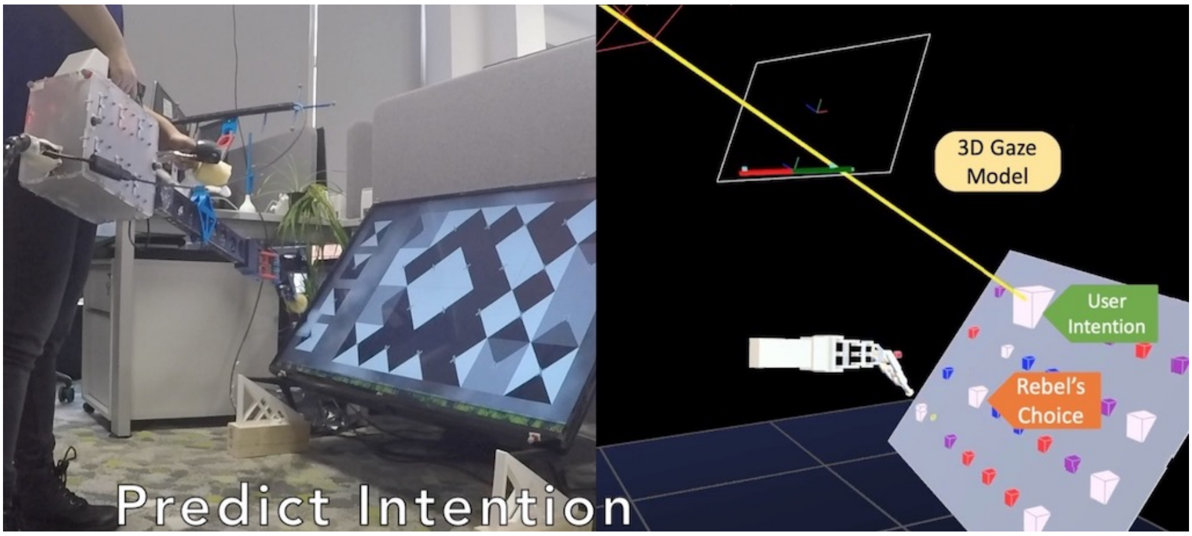
Table 4.3: ***The t-test Results.*** *Bonferroni corrected p-values of pairwise t-test results for the differences in mode depended frustration means. The starred value is significant ($p < .05$).*

4.7.2 Qualitative Analysis

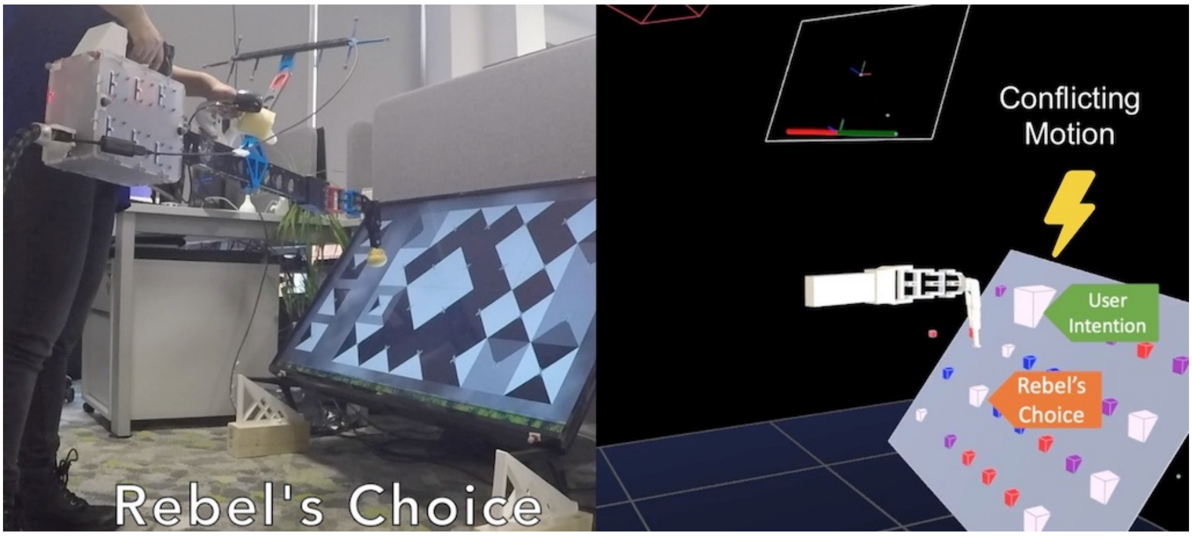
As part of a qualitative review of the robot's behaviour, there were some common reactions of subjects for the respective modes. Notably, an increased number of corrective motion was observed with the robot in the *Rebel* mode (see Figure 4.12). Often, the user rushed towards the intended target but then they had to change their decision as the robot tried to reach a different target. Some participants started to give up on planning entirely and started following the robots' rebellion choices.

In contrast, the collaboration between users and the robot was characterised by a sense of *working together* when the robot was in the *Follow Intention* mode. Because the robot moved towards intended targets, its movement complemented the tactical motion of the user in a timely manner, which lead to shortened travel paths (see Figure 4.13). Some participants expressed their surprise during task execution and wondered *how does it know what I want to do?*

Some participants commented on the behaviour modes after completing the task. The *Follow Intention* mode was often preferred (e.g. "I liked being in charge and the robot was helpful" and "The robot followed my decisions"), whereas the *Random* mode lead to irritation in some users (e.g. "First I thought it would go where I wanted but then it started moving unpredictably"). For the *Rebel* mode, we observed divergent reactions. While some subjects struggled because of the mismatch between the robot's motion and their plans, others started following the robot's lead. This was also reflected in the comments, e.g. "Now the robot does its own thing, I don't like it" versus "It was easier because I did not have to think much". Notably, some participants questioned their abilities (e.g. *The robot is better than me [at coming up with task solutions]*) even though they performed better in the *Follow Intention* mode. This might be due to the underlying assumption that robots are better in term of efficient task solving.

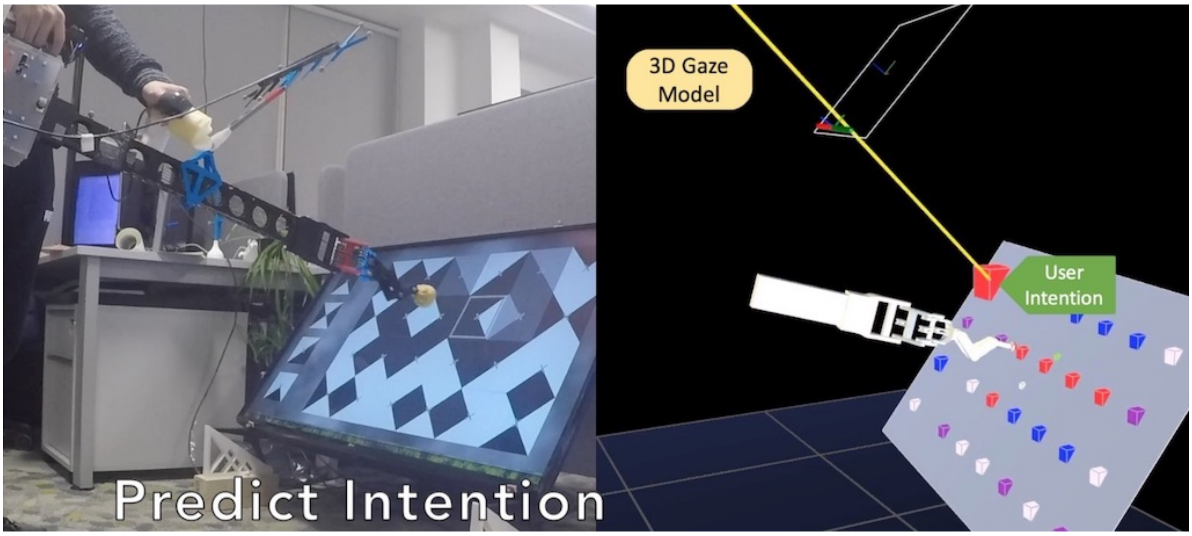


(a) The robot predicts where the user wants to place the piece that they are currently picking. Due to the gazing history, the model predicts the highest probability value for the upper option (green arrow) and the lowest value for the bottom option (orange arrow).

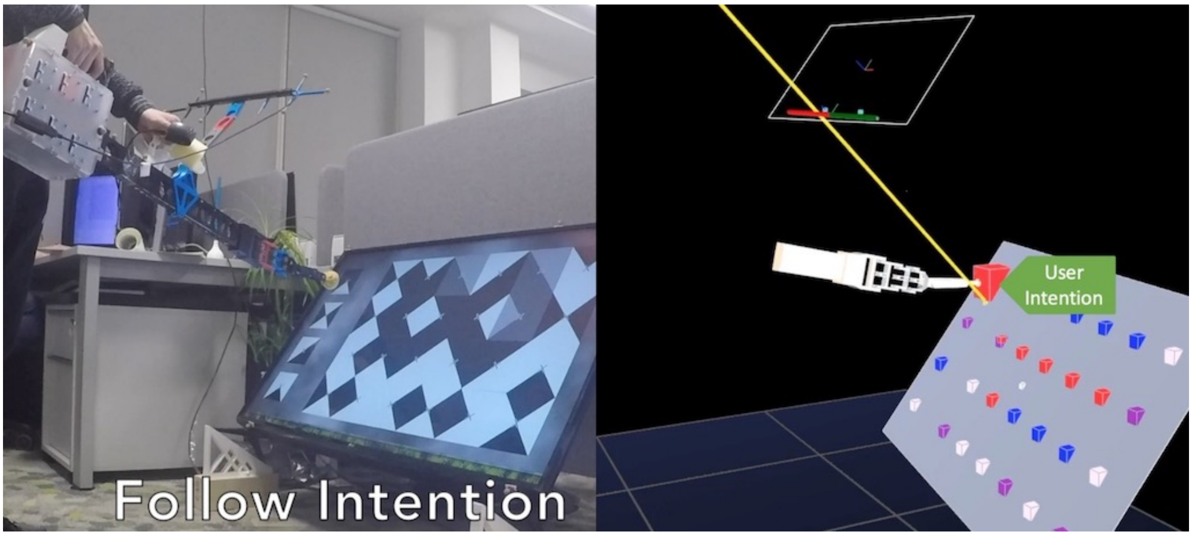


(b) Avoiding user intent leads to a mismatch with the user's tactical motion. The robot avoids the intended piece, which leads to conflicting motion as the robot aims for the lower option (Rebel's Choice) while the user rushes towards the intended option (User Intention). The user is forced to follow the robot's plans to continue.

Figure 4.12: ***The Rebel Mode.*** *These figures demonstrate a common mismatch scenario that was observed when the robot was in the rebel mode. While the user picks up a block from the stock area, the robot uses the eye gaze data (yellow line) to derive a prediction about where they want to place it (a). In the rebel mode, the robot moves towards the object with the lowest prediction value. Because the prediction was correct, this leads to a mismatch between the user's subsequent motion (toward the upper option) and the robot's aiming motion (towards the lower option). The occurrence of these frustrating situations serves as indirect evidence for the robot's prediction being correct. (See more details in the supplementary demo video: youtu.be/H245WdJpNpE)*



(a) Prediction of the red piece during placing of the previous piece as a result of its preceding attention history. Note that that the user is already looking at a subsequent piece while completing the current block.



(b) The robot's motion goes in line with the user's intention as it adapts its plans. Both the user and the robot move towards a common target.

Figure 4.13: ***The Follow Intention Mode.*** *These figures illustrate the systems' underlying intention estimation (right) during task execution (left) and how the different modes affect cooperation. The users' eye gaze model is represented as a yellow line while the estimated probability for a piece to be chosen by the user is indicated by its size. (larger objects have higher probability values). In these figures, the robot is in the cooperative obedience mode and thus follows the highest prediction. (a) shows that the robot predicts which block the user is about to pick up while they are finishing placing the previous one. (b) shows the robot after the user's tactical reaching movement. As the robot points towards the (correctly) predicted block, the robot's motion matches user movements leading to a smooth task flow. After picking the block, the prediction episodes starts over again with the robot predicting where the user intends to place it.*

For a qualitative impression of participants' behaviour for each mode, the reader is referred to the supplementary video[4] of this chapter.

4.8 Discussion of the Intention Validation

This part of the study aimed to validate the intention model implicitly, using levels of the robot's rebellion and users' associated frustration responses as a basis of evidence in the validation process.

The observed difference in frustration ratings between the mode where the robot supports the user's predicted intention versus avoiding it is evidence for most of the intention predictions matching the true intention. This validates the proposed intention model and its application in assisted reaching. With regards to the question, to what extent the prediction of user intention affects the robot's collaboration (research question **Q4**), our interpretation of the results is that during the *Follow Intention* trials, the robot did follow the users' preferred sequence rather than the users adapting it to the robotic motion. That way, the intention model enhances cooperation concerning action anticipation between collaborators. This is an important element of a cooperative robot because it is required for its response on time [29]. There are more layers to it such as intention communication and the adaptation to other user preferences, which leaves space for future exploration.

The difference in frustration levels is smaller than expected. Notably, the Standard Deviation (SD) of frustration levels in the *Rebel* treatment is large compared to the other two modes. Presumably, this is because some individuals did not blame the robot for mismatching motion but themselves, hence why they would not feel frustrated about the robot's choice but believe that it knows better and so they were willing to follow its implicit suggestions.

Mean frustration for the *Random* mode being between the other two modes is expected, given that the robot's choices contain both predicted and non-predicted targets. The effect size is too small for a reliable distinction within this group size. Our analysis furthermore shows that user frustration is more sensitive to the robot's intention prediction than perceived task load and performance. Therefore, we suggest that collaborative agents should follow user intention when there are subtasks with similar priorities for enhanced cooperation.

[4] https://youtu.be/H245WdJpNpE

4.9 Chapter Summary and Conclusion

This chapter aimed to investigate the prediction of user intention as an approach to overcome obstacles in handheld robot collaboration that result from the characteristics of high physical proximity and dependency between the robot and its user. Namely, conflicting movements that arise from mismatches between user's intention and the robot's plan previously lead to frustration in users and was in the way of smooth cooperation. What is necessary is a model that makes swift predictions to allow the robot to react to the user's intention and adapt its plans where possible in a timely manner.

In search of a method that suits the handheld robot setup, the proposed prediction model builds on the previously introduced visual attention system that allows gathering user eye gaze information with respect to task-relevant objects in the work scene. The proposed method goes beyond already established gaze-based prediction models. Here, eye gaze data is combined with context information that links scene objects that are related through subtasks following the key-and-lock principle.

To gather training data for the model, a block copy task was introduced where the robot was used to pick and place blocks to reconstruct a pre-defined pattern. Experimental data was then used to train an SVM-based model for the anticipation of user actions. The results show that picking actions can be predicted with up to 87.94% accuracy at 500 ms ahead and dropping actions with 93.25% accuracy at 1500 ms ahead in real-time. We show that the prediction accuracy benefits from combining gaze data with context information, leading to a significant improvement.

Using the block copy task setup, the model's performance was validated through user studies where the robot reacts to anticipated user intentions. The knowledge about the user's plans enabled it to choose to comply with it or to work against it. This allowed for an indirect validation of the model's accuracy as a significant difference in frustration levels was observed in these two conditions. The ability of the robot to induce frustration through rebellion is evidence that the predicted user plans the robot was objecting to were the true intentions in users. The introduction of frustration via rebellion was used as a new way to investigate the usually complex aspect of effectiveness of intention prediction in human-robot interaction. This approach, together with the model proposed, could be useful in other cooperative robot studies.

The proposed intention prediction model has real-time capabilities that allow the robot to adapt its plans and motion planning to the user's intention fast enough to assist in reaching towards objects the user intends to interact with. This outcome demonstrates that the robot's cooperation could be enhanced through the intention model given the availability of gaze data and context information, e.g. in assembly tasks with complementary parts or when a specific tool is required and needs to be fetched to complete work on specific objects in the work environment. As such, the principle of using multiple sources of information to make predictions is applicable to other domains that are characterised by close collaboration and interaction with humans for example in shared workspace assembly and assisted living.

The quality of a robot's cooperation in collaborative task setups depends on many aspects that are required for an efficient exchange of information and interaction that is safe, efficient and reliable. The proposed method of intention prediction is a contribution to the robotics research community that brings it one step further on its way towards a future of assisting robots that humans can use effortlessly thanks to the evolution of the machine's counterpart to human intuition.

This chapter focused on the robot's interaction with a user who essentially knows how to solve the task hence why the prediction of intention became relevant. Another case to consider is the scenario of a user with limited task knowledge. This raises the research question, how task knowledge could be obtained and communicated to the user. Often, another person (e.g. an expert) holds task knowledge that would enable the carrying user to complete the task. For this reason, the subsequent chapter will bring the interaction with the handheld robot to a new level that involves two people controlling the robot simultaneously.

Chapter 5

Reach Out and Help: Assisted Remote Collaboration through a Handheld Robot

In the Chapter 3 and 4, we investigated interaction concepts for single-user handheld robot setups. In this chapter, we extend our exploration to a setup with two users that collaborate via the handheld robot in a helper-worker setup. This addresses current problems in remote assistance as the aspect of physical remote interaction facilitates efficient collaboration, which opens up new handheld robot applications. The outcomes of this chapter are summarised in the supplementary video[1] (scan QR code).

The main results of this chapter are currently in submission for publication. The preliminary article is available as preprint version [211].

5.1 Introduction

This work introduces an assistive collaboration system that allows a remote expert to guide a local user through a task using a handheld robot which is carried by the local user. The aim of this research is to explore how the robot could work as a mediator between the two human parties based on the proposed collaboration concept. That is, the expert can perceive and explore the work environment through the robot and then delegate task operations to the robot that in turn completes them in part autonomously. The proposed system is tested using the maintenance of a plumbing system as an example task, which covers a broad variety of common human-robot-human interaction problems.

[1] Chapter 5 Summary Video: https://youtu.be/cTJ8tNJJXV0

We show that the robot's autonomous contributions to the task leverage a workflow that is characterised by efficient interaction.

5.1.1 Challenges in Remote Collaboration

Collaborative remote assistance tasks usually involve a less experienced person (local worker) who has to manipulate a set of physical objects with the help of a remotely situated expert. In the context of a complex problem, the local worker has limited knowledge about required operations and relies on instructions of an expert. Examples for such tasks are maintenance [53] and inspection [121] of remotely located systems and expert-guided surgery to train remote novice surgeons [10, 162]. Current systems solve collaboration challenges such as spacial referencing and communication, however, they limit the expert's view to a camera that is either stationary or worn by the local user [54]. What is missing is a system that allows the remote user to explore the environment for inspection and grants physical access to decouple the solution from the local worker's competences.

A new approach to remote collaboration is to use a robot to enable a remotely located user to help with a task in the local user's workspace. For example Veronneau et al. [226] present a remotely actuated waist-mounted robot that can assist in various tasks through the help of a remote assistant. With these recent developments, the question is raised to what extent a robot's partial autonomy and task knowledge could be used to leverage remote interaction. Answering this question is particularly relevant for more complex tasks, i.e. in tasks where the solution is not immediately obvious to the local user, e.g. due to a lack of task expertise. With the robot as a partially autonomous assistant, such a setup would go beyond a master-slave relationship and more towards a *working together* paradigm with the users and the robot each contributing to the task.

Remote maintenance is of particular interest for industrial applications [21]. Modern products and plants are characterised by increasing complexity which requires high expertise to diagnose and solve problems. However, it might be expensive to get an expert on site and relying on the consultation of a manual is inefficient [231]. Therefore, the development of remote assistance systems has caught researchers' attention in recent years (Section 2.7 and 2.8). Some solutions have considered remote guidance through AR [21, 69] and semi-autonomous telemanipulation systems [139], i.e. systems that solve some subtasks such as grasping autonomously. However, what is missing is a remote assistance setup that combines the advantages of physical access through telemanipulation with the ones of cooperative guidance and task solving.

With humans and robots working together in complex tasks, defining the collaboration style in a scenario that is characterised by mixed-initiative interaction is of particular importance to enable each party to contribute to the task with their respective strengths at suitable times [67]. As outlined in Section 2.5, traded control is a subdomain of mixed-initiative interaction, where a human is in charge of controlling the robot for some parts of the tasks, while the robot is supervised but autonomous during other parts of the task [105]. This concept could benefit interaction with remote maintenance robots: The remote expert could control the robot during the inspection part, in which they explore the scene, while the robot could take over control once the expert makes a diagnosis and decides on what needs to be done. Regarding the proposed pipe maintenance test case, examples for such semi-autonomous operations could be to adjust valves in the system or to inspect gauges and pipe elements. What is necessary for the autonomous execution is the robot's knowledge about the task, i.e. the location and identity of task objects and how to interact with them. For example, adjusting a valve is a different form of interaction than checking a pipe element for cracks.

Previous work on handheld robots [58–60] demonstrates that task knowledge and environment information can be used to assist in collaborative task solving, which brings together the robot's precision and the natural navigation competences of human users (see also Section 2.2). Arguably, a device with these characteristics could bridge the aforementioned gap between remote guidance and telemanipulation, with the handheld robot helping both effective communication between the workers and task outcomes.

While the introduction of the robot's task knowledge brings advantages, a crucial aspect remains unexplored, namely where this knowledge might come from, e.g. whether it could be learned [3] or derived from a remote expert and mediated through the robot [106]. In this work, we start with exploring to what extent a handheld robot is suitable for a remote assistance human-robot-human setup and whether the benefits of the robot's partial autonomy can be observed analogous to aforementioned work.

Essentially, remote assistance systems are characterised by two key aspects: remote guidance (introduced in Section 2.7) and telemanipulation (see Section 2.8). Remote guidance systems allow a remotely located helper to assist another person through directions and instructions. Usually, the remote helper can perceive the worksite through a camera system, e.g. as part of portable systems [69, 121, 162, 177, 231] or in some instances controlled by the helper [61, 108]. While these solutions enable communication and instruction, they do not allow for direct physical access, i.e. the local instructed users have to carry out the tasks themselves by following directions. Physical manipulation of task-objects through the instructor is not part of the aforementioned designs. The counterpart to remote guidance is the field of telemanipulation. Here, remote physical manipulation is the central aim. Instead of instructing another person, operators solve the task remotely through a robotic system that can be controlled in a precise manner. Examples can be found in applications for surgery [62] and for tasks in maintenance and inspections in hazardous environments such as fusion power plants [204], the deep sea [232] and outer space [195]. These works are milestones for telemanipulation of highly specialised tasks and for environments that are inaccessible for humans. However, the research question remains open, how concepts of remote guidance and telemanipulation could be combined in a single system for collaborative task solving.

5.1.2 Chapter Overview

This chapter explores a remote collaboration system that enables a remote expert to assist a local worker in a task using the handheld robot as a basis. Its tooltip can be used for spacial referencing, i.e. navigation through pointing gestures and the work environment can be explored remotely using a steerable camera. Moreover, the system allows for the remote expert's physical access to the work through the tooltip interface and intuitive control.

The spacial coordination between the remote user, the robot and the local user is a challenging problem. Here, we explore multimodal communication and the robot's ability to assist through world stabilisation as a starting point to address this problem. We see potential in the robot's accuracy and motion capabilities to delegate small-scale tasks to the robot. That way, remote users can focus on strategies and their communication to complete the task rather than wasting their capacity on fine-tuned tool tip motion required for the task execution.

The development of a remote assistance system requires the achievement of the following objectives:

- A sensory extension of the existing handheld robot to allow for the remote perception of the environment, e.g. through cameras.
- A workstation that allows a user to remotely interface the handheld robot including visual feedback, tooltip control with the option of task delegation to the robot.
- To leverage task execution, we propose semi-autonomous assistive features for the handheld robot, which are based on position correction and task knowledge. These allow remote users to chose an object for interaction, while the local task execution is taken over by the robot.
- For validation of the system, we propose a collaborative maintenance task with a partially simulated pipe system. It covers common teleassistance problems such as inspection, manipulation and coordinated navigation.

We explore this setup guided by the following research question:

Q5 How does the handheld robot's autonomy and task knowledge affect performance and communication in a remote assistance setup?

This work involves an extended version of the open robot platform, introduced in [59], which is coupled to the proposed workstation (Section 5.2.5) for telepresence access. As the interaction between the handheld robot and its carrier has been investigated extensively in previous studies [58, 60, 209, 210], the focus of this work lies on the experience of the remote access, i.e. the role of the remote expert. We investigate the subjects' interaction in this human-robot-human setup and assess proposed semi-autonomous assistive features of the robot. An experiment overview can be seen in Figure 5.1. The outcome of the studies are summarised in the following list of contributions:

- As the main contribution, a new paradigm for remote assistance through handheld robot collaboration is introduced. It allows for remote-teamwork with the handheld robot as a mediator of instructions and physical interaction.
- Another key contribution is the proposed concept of local task delegation through the robot's assistive features. We show that their use improves team productivity and usability, hence why we suggest that they could benefit other telepresence systems too.
- The qualitative analysis reveals a repeated sequence of exploration, guidance, local task solving and retraction as a common emerging strategy in this novel collaboration setup. This knowledge is useful to guide future developments of similar human-robot-human setups.

A demonstration of the remote assistance system and examples for the robot's semi-autonomous assistance features can be seen in the supplementary video material[2].

5.2 Remote Assistance Study

In this study, we propose and test a remote assistance system which consists of two main parts. On the local workspace site, a camera-equipped handheld robot with 5-DoF motion capabilities (displayed in Figure 5.2) is carried by a local user. A remote user accesses the robot through a remote interface, which allows them to control the robot for inspection, manipulation and gesturing.

[2] https://youtu.be/cTJ8tNJJXV0

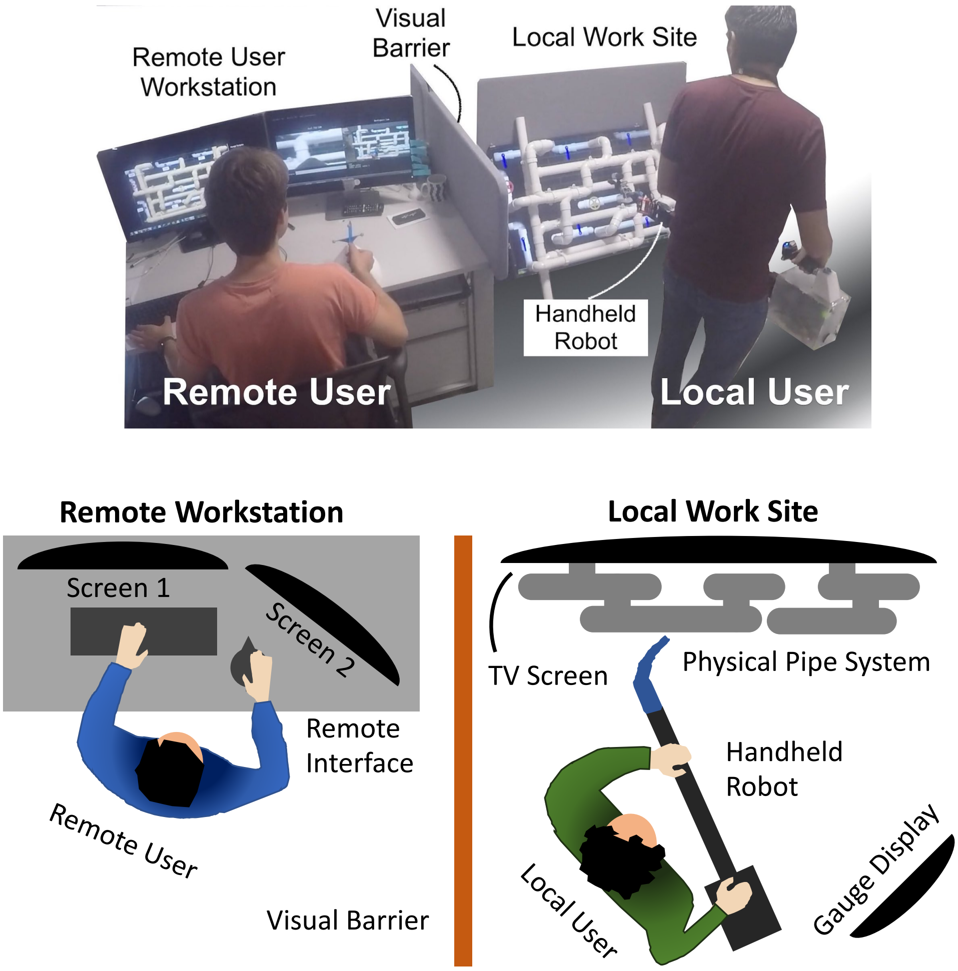


Figure 5.1: ***Overview of the Remote Assistance Experiment Setup.*** *The remote assistance system allows a remotely located user to assist a user at the local work site through a remote-controlled semi-autonomous robot, which allows for inspection and physical interaction. Using the remote workstation, the local workspace can be perceived through robot-equipped cameras. The local user helps with reaching through moving the robot to specific locations in the scene as they follow the remote user's instructions. Detailed views of the respective setup elements can be seen in Figures 5.3 to 5.7.*

5.2.1 Study Design

We investigate the collaborative interaction between the three agents involved in the task, i.e. the remote user, the local user and the handheld robot. Our main focus lies on the effect of the robot's semi-autonomous assistance features on the collaborative task performance and communication strategies. Figure 5.1 shows an overview of the experiment setup. The experiment trials were divided into an initial phase to assess the training procedure and a subsequent study to investigate the effect of the robot's condition on performance and user perception using a within-subject design (Section 5.3). In each, the remote user and local user pairs use our proposed remote assistance system in two different conditions:

Non-Assisted

The robot is steered manually by the remote user to guide the local user to task objects. After reaching an object, the remote user completes the task manually. Some tasks require verbal requests for information about the system's current state from the local user.

Assisted

Initially, the robot is steered manually for guidance. When approaching a task object, the remote user can select it to delegate the interaction to the robot. The robot assists through locally fulfilling the task within its workspace and takes into account the system's current state (detailed description in Section 5.2.6).

In each condition, the local user's responsibility is to follow the lead of the instructions of the remote user and to assist in reaching through moving the handheld robot towards task objects.

The setup is a semi-simulated pipe system, which we use as an example for a collaborative maintenance task. Solutions for solving this task require elements of common real-world assistance problems, such as inspection, diagnosing, instructing and manipulation.

5.2.2 Hypotheses

Concerning the effect of the robot's assistance features on performance and collaboration, we hypothesise that with those features enabled:

H1 The time to complete the task would be reduced as the robot's assistance saves time.

H2 The robot's task knowledge and autonomy reduce the required total amount of communication and reduce the share of the local user.

H3 The perceived workload of both users would decrease as it gets transferred to the robot.

H4 The users' rating of the system's usability would be increased.

H1 goes under the assumption that the robot performs local tasks faster when solving them autonomously, due to its accuracy and precise timing, i.e. the time for aiming and fine-tuning of motion is shorter and the robot can react faster. Moreover, the operator can already start to plan a subsequent step, while the robot completes a previous step. With regards to **H2**, it can be expected, that communication concerning coordination between both parties gets reduced as the robot's autonomy helps to regulate spacial inaccuracies. Additionally, in autonomous mode, the robot accesses information about the workspace site that is otherwise exclusive to the local user. Hence, a decrease in the local user's share of verbal communication is expected. As a result, executing the task becomes easier (**H3**) and the robot more user friendly (**H4**).

5.2.3 Collaborative Setup

The task setup consists of two main areas, the local workspace site and the remote workstation (see Figure 5.1). For the experiment, remote and local users were located in the same room, however, a visual barrier prevented them from direct interaction. They were allowed to speak to each other as if they were on a phone call.

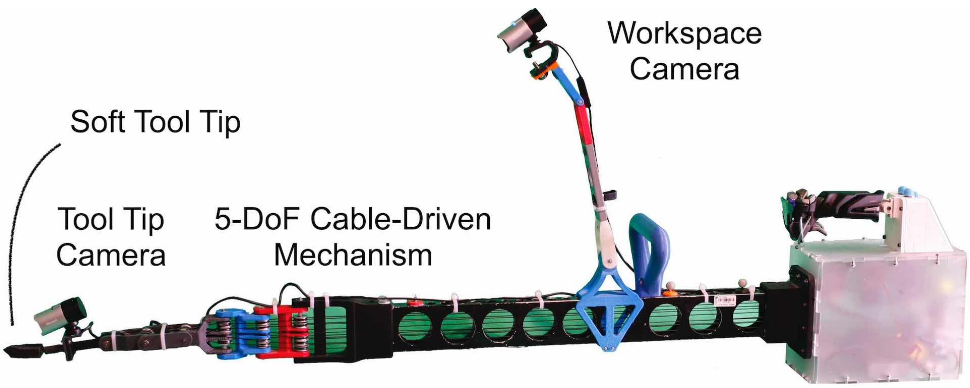


*Figure 5.2: **Handheld Robot in the Remote Setup**. It has a remote-controlled tooltip with 5 DoF, equipped cameras and a tooltip for sensor simulation. The workspace camera grants the remote user a broad overview of the scene, while the camera at the tooltip allows for close observation of tooltip operations.*

For the experiment, we used the handheld robot reported in [59], of which the mechanical design is publicly available on our research website [1]. The robot was adapted to suit the requirements for a remote control setup. Inspired by examples from the field of remote assistance [233], we followed a two-camera design to combine a detailed view with a more distant scene view. As can be seen in Figure 5.2, the robot features a 5-DoF actuated tip and two cameras. The first camera is fixed to the robot's frame delivers the overview of the current workspace. The second camera is positioned close to the tooltip so that it can be directed for exploration, whilst allowing a detailed view on tooltip operations.

The remote user sat at a desk equipped with the remote interface, which allows the perception of the robots' workspace and features a 5-DoF input for remote control as well as information about the system and its required goal states. The local user is located in the workspace where the physical task has to be completed. The user holds the robot in place for inspection and helps the remote user to reach objects for manipulation and diagnosis. The collaborative setup is shown in Figure 5.1 and 5.3, additionally, a task demonstration is included in the supplementary video[3].

5.2.4 Experiment Design

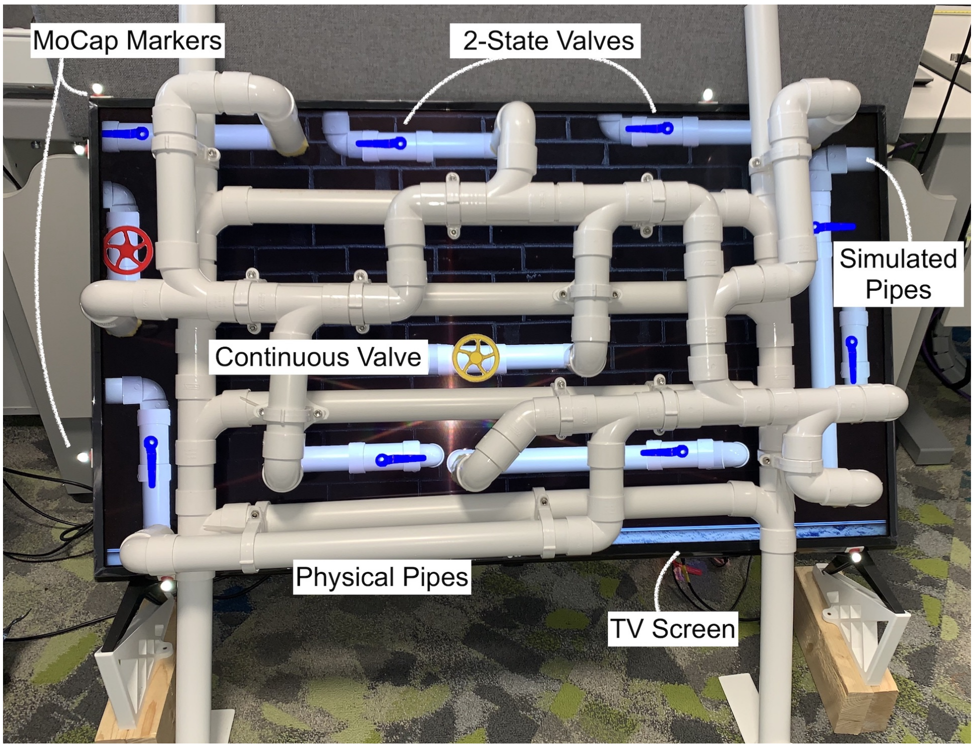
To test the proposed hypotheses **H1** to **H4** (introduced in Section 5.2.2), the experiment setup requires the following elements:

- Common remote collaboration tasks such as inspection, diagnosis, remote guidance and manipulation.
- Some information about the task goal should be exclusive to the remote user. This is their expertise in the system and part of the preceding training.
- Some information about the current state of the local work site should be exclusive to the local user. This is the worksite specific knowledge.

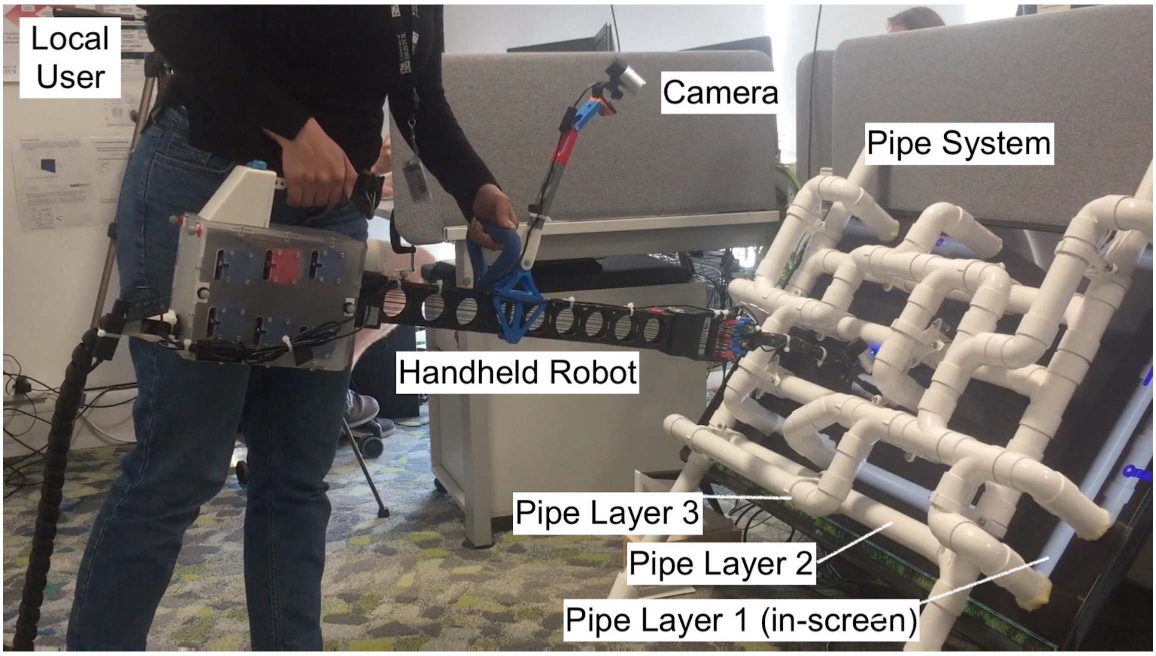
Following the above criteria, a semi-simulated maintenance task of a pipe system was chosen. The task is simple enough to fit the scope of a user experiment and allows adding cognitive load to test more complex scenarios. The setup consists of a network of pipes, valves and gauges. While the majority of the pipes are a physical system, the valves are simulated through a display in the background of the pipe system (Figure 5.3). The gauges are also simulated, but on a separate screen (Figure 5.4) diagonally behind the local user. That way, the remote user cannot look at the gauges while carrying out work on the pipes, i.e. information about their current state is exclusive to the local user. The values of the gauges depend on the state of the valves.

The system contains 8 discrete valves, 2 continuous valves and 3 gauges (see Figure 5.3 and 5.4). There are two different kinds of valves, the first kind has two discrete states i.e. open and close while the wheel-shaped valves are continuous.

[3] https://youtu.be/cTJ8tNJJXV0



(a) Front View



(b) Side View

Figure 5.3: ***Overview of the Mockup Pipe System****. The pipe system was used for the maintenance experiment task. Its main elements are the (blue) two-state valves, the (yellow and red) continuous wheel valves and the pipes. The piping is organised in three layers, one is simulated in the TV screen (see front view) and the other two layers made of physical pipes in front of it (see side view). A top view of the experimental layout is presented in Figure 5.1.*

Depending on their predefined contribution value, the two-state valves add 10 or 100 units to the associated gauge when they are turned open and 0 when they are closed, i.e. the blue gauge displays the sum of contributions for each opened blue gauge. This models measuring combined currents with the valves controlling the sources with varying current contributions [238]. The continuous valves contribute between 0 and 10 to the respective gauges.

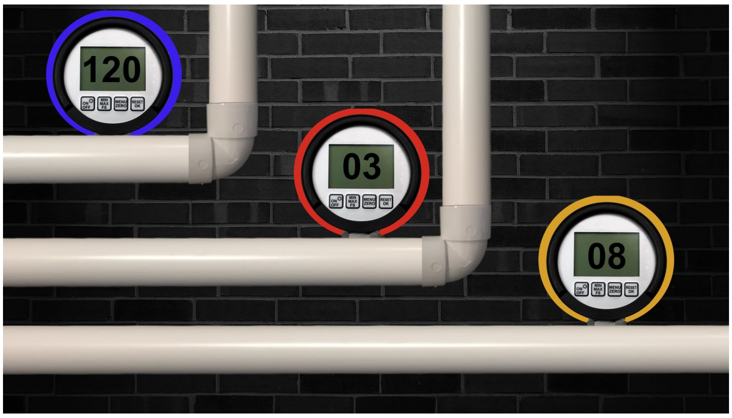


Figure 5.4: ***Gauge Display with Pipe System Response****. Three gauges indicating the current state of the pipe system. The readings result from the valve configurations, i.e. which of the (blue) two-state valves are open or close and to what degree the (red and yellow) continuous valves are turned open. Adjusting the valves to get the gauge values to a specified target value is one of the main steps within the maintenance task. The position of the gauge display with respect to the overall setup is presented in Figure 5.1.*

The experiment consists of two main tasks, adjusting the valves and checking the pipes for cracks. For the first task, the valves need to be changed so that the gauges get to a predefined target value. As an example, suppose the target value was 230, 7 and 5 for the blue, red and yellow gauges, respectively and all valves were closed. The operator would then have to open 5 of the two-way valves, 2 with a contribution of 100 and 3 with the contribution of 10. The continuous valves would have to be turned open until they match the target values 7 and 5.

The remote user holds knowledge about the target values of the gauges and the contributions of the respective valves and consequently knows what changes need to be done. However, initially, the remote user does not know the pipe system's current state, i.e. valve states and readings of the gauges. Retrieving this information requires either a visual exploration of the work scene or verbal requests.

When the soft tooltip of the robot touches the valve, the remote user can press an activation key to turn it open or close. This manipulation is simulated through a 2D animation of the valve handle/wheel in the screen and the associated gauge value changes accordingly. For simulation purposes, the touch of the robot's tip is registered using motion capturing[4], which enables 3D localisation of the handheld robot and the screen surface.

[4]OptiTrack: optitrack.com

For more details of the motion capturing setup see Section 3.2.2 and Figure 3.7.

For the pipe checking task, a sonar sensor is simulated. The procedure of taking a measurement is inspired by [16] where a sensor tip is placed on a machine part for a short duration to check the condition of the material, e.g. for crack detection. Similarly, here, the robot's tip needs to be in contact with a pipe to be checked for a few seconds while the remote user activates the sensor reading.

There is no predefined order in which valves have to be opened or closed and pipes to be checked. In that way, the remote user has to come up with an individual strategy for a solution, which brings the task closer to real-world problems. The maintenance task is completed when all gauges display the desired target values and a predefined set of pipes is successfully checked with the sensor.

The task setup stimulates cooperation between the users since neither party could solve the task on their own. The local user lacks knowledge about the specific task goals and relies on the remote expert's guidance, who would not be able to access the workspace without the help of the local user.

5.2.5 Remote Interface

The workstation of the remote user consists of three main units: the robot control system, a display with the robot states and another one containing task system information. This design of the visual interface is in essence derived from solutions reported in several remote assistance studies, e.g. [21, 54, 140], while the spacial input is inspired by work on remote manipulation [204]. An overview can be seen in Figure 5.5.

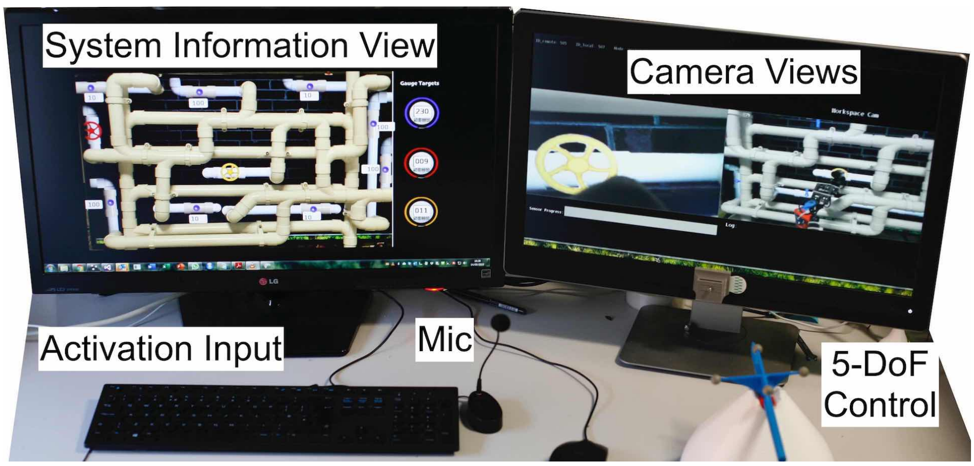


Figure 5.5: ***Remote Workstation.*** *The remote user has access to specific system information (see system information view) and can see the local work site through the camera views. The robot is remote controlled using the activation input and the 5-DoF control. The position of the remote work station with respect to the overall experiment setup is presented in Figure 5.1.*

Input Control Unit

The main part of the control unit is the 5-DoF input. It is realised through a wand, which is tracked through motion capturing. Its relative position and pose to the input base socket is replicated by the robot arm with respect to its home position as a local reference frame. To account for the limits of the robot's workspace, the wand is wired to the input base. When the wand rests on the input base, i.e. in its initial position, the robot is half crouched, i.e. at the centre of the task space.

The initial position allows the remote user to either reach out or retreat as demonstrated in Figure 5.6. Tooltip operations such as manipulation and sensor activation can be triggered by pressing the space bar of a keyboard.

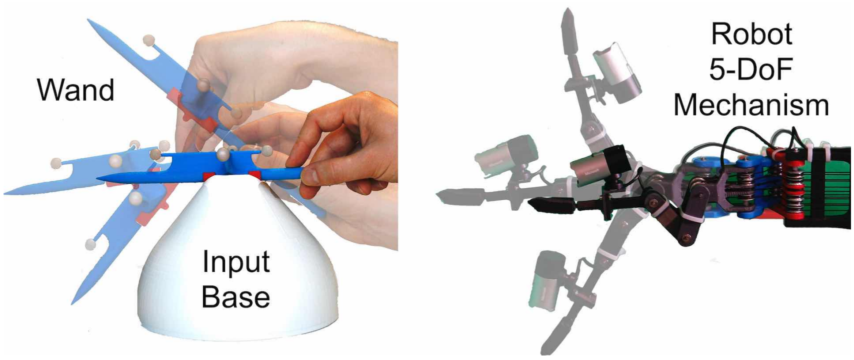
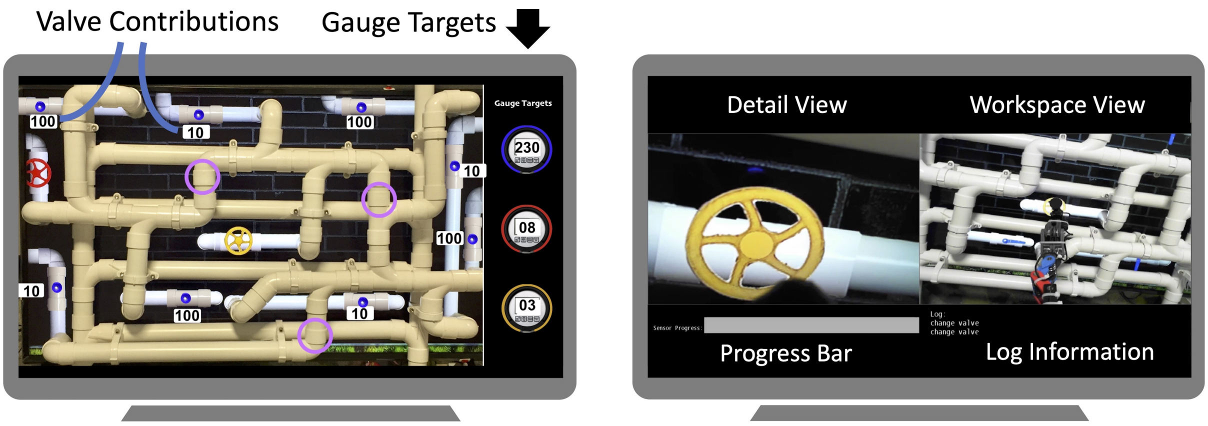


Figure 5.6: ***Remote Control of the Handheld Robot****. Example of a remote user's 5-DoF spacial input and the robot response (transparent images). The non-transparent position is the home position, which is centred within the robot's work space.*

Control Views

Two displays of the workstation provide the remote user with system information (Figure 5.7). The view for the robot states contains a split view for the two cameras, a progress bar for the simulated sensor state and a log protocol of robot actions. The screen with system information contains essential details that are required for the remote user to develop a strategy. That way, the display serves as a manual for the remote user, where they can pick up detailed system information e.g. how the components are connected through pipe routing. Additionally, the display shows valve contributions, respective gauge target values and indicates which pipes require checking. For diversity, these change randomly between trials. Quickly looking up the details for the individual tasks is part of the remote user's training prior to the experiments. does not show the current state of the system on the working site, e.g. valve configurations and current gauge readings. This information has to be gathered as part of the diagnosis process.



(a) Screen 1 *(b) Screen 2*

*Figure 5.7: **Illustration of Workstation Views:** The system information view (a) provides the remote user with information that is required to derive a solution. This includes valve contributions, the target values for the gauges and which of the pipes need to be checked (purple circles). The screen for robot states, contains a split view that shows both camera perspectives and at the bottom section features progress information. The screen setup is part of the remote workstation (see Figure 5.5).*

Experimental Limitations

The setup described here is a first attempt of collaborative telepresence with two humans in the loop with a focus on the subjects' interaction. As the robot is wired to the remote interface, the proposed testing setup does not take into account possible lags of the camera or time-delays in tip actuation, which might affect collaboration in a long-distance setup, e.g. over the internet. Dealing with time-delays is a complex problem in telemanipulation [95], which exceeds the scope of this work. Furthermore, some existing solutions for spacial input that are specialised for telemanipulation provide haptic feedback in the form of force reflection [153] in addition to direct control. However, since this work focuses on the investigation of collaboration, object manipulation was simulated and so haptic feedback becomes non-critical. Therefore, this feature was not implemented in the presented solution.

5.2.6 Robot Assistance Condition

When the robot is in the assisted condition, a set of features are enabled which incorporate task knowledge and navigation capabilities. In this state, the remote user is no longer required to complete detail motion. Instead, they select an object to interact with and the robot aims for it when the activation key is pressed. For example, the remote user could roughly direct the local user to a valve, activate the assistant and the robot completes the manipulation. The robot has knowledge about the gauges target values and can, for example, turn open the continuous valves until the associated gauge matches the required value. Similarly, the robot helps with a world-stabilised positioning of the sensor on the pipes' surface during crack detection and retreats when the measurement is complete.

5.2.7 Trial Procedure and Data Collection

Both, the initial training study (Section 5.3.1) and the remote collaboration study (Section 5.3.3) are based on the same trial procedure. However, they differ in the distribution and matching of participants to satisfy the respective study purposes (see Figure 5.8). Therefore, this section describes the general trial procedure, while specifications about participant counts and distributions are reported separately.

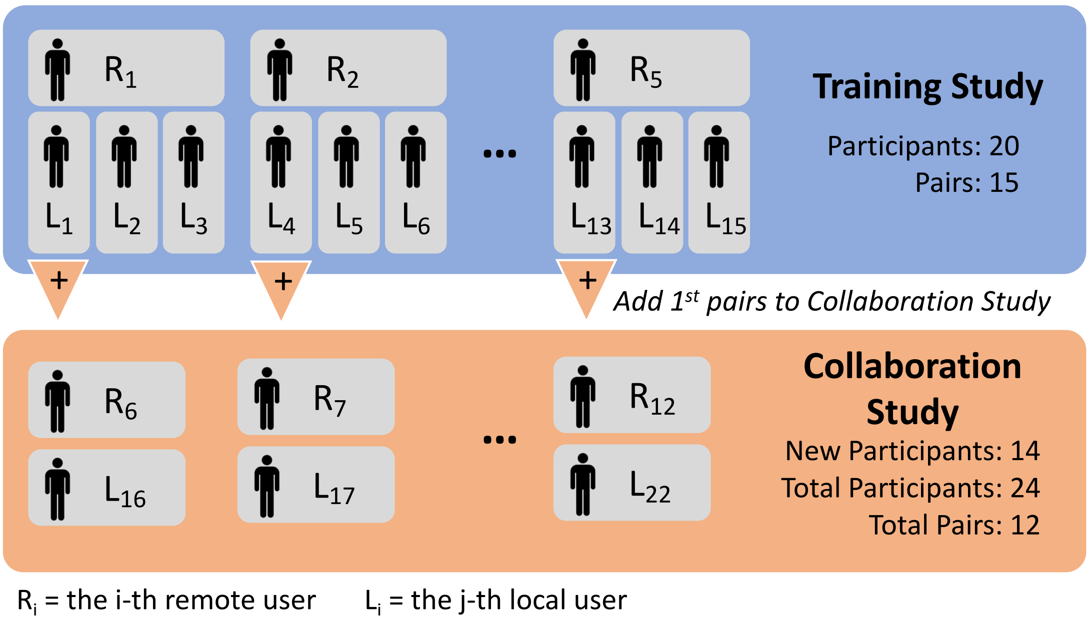


*Figure 5.8: **Illustration of Participant Matching** to create subject pairs for the training study and the collaboration study, e.g. R_1 was paired with L_1, L_2 and L_3, respectively. Each pair completed the task once for both conditions: assisted and non-assisted (randomised). In the training study, the experiment was repeated twice for each remote user to investigate possible improvement over time. The collaboration study contains unique pairs only and focuses on user perception. Note that every first pair of the training study was added to the collaboration study to increase statistic power.*

For a given remote-local pair of subjects, the experiment task was executed one time for each of the conditions *non-assisted* and *assisted*, which were counterbalanced across the trials. This was done to cancel out within-pair training effects and preferences. The initial state of the valves, the gauges' target values and which pipes would require checking were randomised in such a way that the number of actions required to solve the task remained constant for each trial. Participants were asked to complete the task swiftly and were informed that their completion time was recorded to measure their performance.

Investigating the proposed research question **Q5** requires metrics for the effectiveness and quality of human-robot interaction. As the focus of this work is on collaboration rather than solving the specific task, using the result of a task as a measure of performance (e.g., traditionally in surgical robotics [219]) is unsuitable here. In accordance with the evaluation criteria established by Nielsen et al. [156], we use task completion time as a measure of performance and perceived task load and usability as measures of users' involvement.

With this in mind, the completion of the experiment task was followed by recording the *NASA* TLX [63] and a SUS [27] for both subjects and for each condition (see questionnaires in Appendix D).

As mentioned in Chapter 3 (Section 3.5.3), the *NASA Task Load Index* (TLX) [63] is a standardised test that measures an individual's perceived task load based on an average rating of the six aspects: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort and Frustration. The test quantifies perceived task load with a score from 0 to 100, i.e. low to high (lower is better), respectively. The NASA-TLX has been in use for over 30 years now and it has already proven useful in previous handheld robot studies, e.g. in [58, 60] and in our work presented in Chapter 3 and 4.

The System Usability Scale (SUS) [27] is a technology-independent rating based on 10 Likert scale questions focusing on an individual's experience during usage. The test quantifies usability on a scale from 0 (low) to 100 (high), where a higher value is better than a lower one. The test has been in place since the late 90s and has been used to assess telerobotic systems before [5].

Furthermore, the required time to complete the task was recorded as well as voice recordings from microphones, which were placed close to the respective users. The audio material was later transcribed to derive word counts for the analysis. As an estimate of prior experience with video games, all subjects were asked for the average weekly amount of hours they usually spend on gaming. Furthermore, video recordings of the trials were taken for a qualitative assessment.

5.3 Experiment Execution

The experiments were executed in two parts an initial phase to assess the preparative training of participants and a subsequent collaboration study. The results of the initial phase were used to inform and continue with the second study to assess our hypotheses (Section 5.2.2) based on the research question **Q5**.

5.3.1 Training Effect on Performance

We aim to simulate a scenario where a skilled remote user works in collaboration with a local user. This requires suitable training that allows local workers to familiarise with the handling of the handheld robot. Furthermore, the training aims to qualify remote user subjects as experts in the given task. Therefore, we repeat trials in a pilot study for a small group of remote users to assess whether their improvement went into saturation after the training procedure. The following lists the steps taken during the training:

S1 General introduction and system demonstration for both subjects ($\sim$ 10 min).

S2 System demonstration with the remote user subject in the position of the local user and the experimenter being the remote operator ($\sim$ 10 min).

S3 Training of the remote user with the experimenter being the local user for both robot conditions as per Section 5.2.1 ($\sim$ 20-30 min).

S4 Training of the local user with the experimenter being the remote user for both robot conditions ($\sim$ 10 min).

S5 Training of both subjects in their respective roles ($\sim$ 10 min).

Step **S3** to **S5**, respectively, were repeated until both subjects yielded confidence.

20 participants (4 females, $m_{age} = 30.3$, $SD = 4.9$) were recruited and split into two groups for the role of the remote user ($n = 5$) and the local user ($n = 15$). The volunteers were staff and students from our department, however, no technical knowledge was required for either of the roles. There was no benefit or financial compensation in exchange for participation.

Each of the 5 participants of the remote user group was matched with 3 unique participants of the local user group. The roles were never swapped and local users were never matched with another remote user to avoid additional hierarchical dependencies in the experimental data. That way, 15 pairs were created and each received the training as per Section 5.3.1. Subsequently, each pair completed two trials (see Section 5.2.7), where both robot conditions (*non-assisted* and *assisted*) were tested in randomised order to cancel within-pair training effects. If it was a remote user's second or third trial, the pre-experimental training solely consisted of step **S5**, i.e. without the steps of the individual training. The experiment series was completed with:
15 pairs $\times$ 2 conditions = 30 data points (see Figure 5.8).

5.3.2 Training Effect Results

This analysis aims to assess the effectiveness of the proposed training procedure, i.e. whether it takes the remote users to a point of expertise in the task where they no longer improve. Therefore, experiment data were analysed concerning completion time as a measure of performance with regards to the number of trials completed by the remote user.

We performed a two-way ANOVA with the robot's mode and the trial number as independent factors, respectively, and the time to complete as a dependent variable. One data point ($\sim 3\%$ of the total set) was considered as an extreme outlier as its difference from mean exceeded three standard deviations. The results yield that the completion time significantly increases when the robot is in the non-assisted mode compared to the assisted mode ($p = .016$). Whereas no evidence ($p = .681$) was found for further improvement over the trials after the initial training and there is no significant interaction ($p = .331$) between mode and trial count. A diagram of the data can be seen in Figure 5.9.

We conclude that the training procedure is effective enough to take the remote user subjects to a level of performance where they no longer improve over time, which motivates further trials. Furthermore, the results imply that the robot's autonomous features improve the teams' task performance.

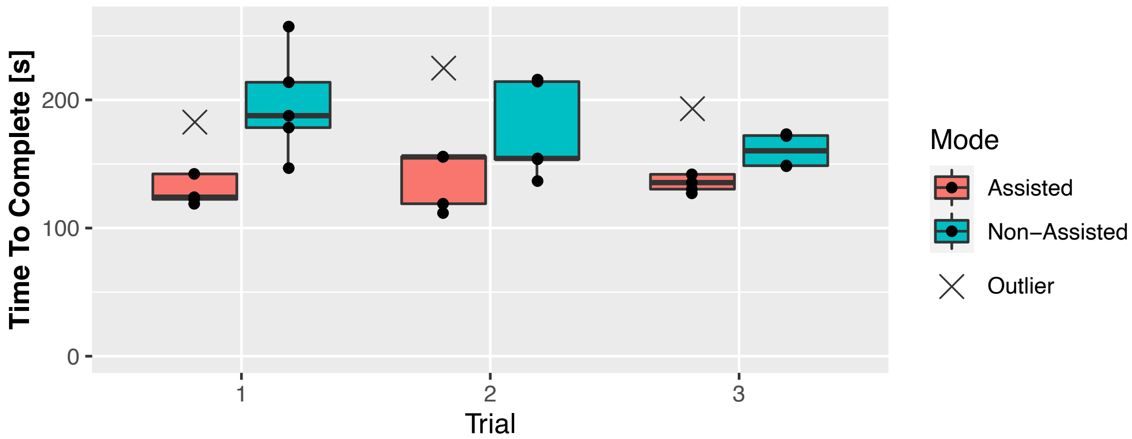


Figure 5.9: ***Results of the Training Study****. Task completion time (lower is better) over subsequent trials for two different robot conditions. A set of 3 trials represents the progression of 1 remote user matched with 3 different local users. The even performance for a given mode indicates that the remote user's performance has levelled out during the training while it is greatly impacted by the robot's condition. The outliers are outside 1.5 times the interquartile range of the respective distribution.*

The assessment of the effect of robot condition on the two user groups (remote and local), requires a paired within-subject study design [190]. Therefore, we take the subset of every first trial of this study, i.e. 5 pairs (see Figure 5.8) and continue testing in unique pairs in the subsequent collaboration study.

5.3.3 Remote Collaboration Study

We recruited 14 new participants (all male, $m_{age} = 26.1$, $SD = 4.1$), who were teamed up to form 7 new unique pairs. Most were students from our Computer Science Department, however, technical knowledge was not required. Again, there was no financial benefit for their voluntary participation. They completed the same training and experiment procedure as the 5 pairs from the previous study, i.e. one trial for each mode in randomised order. This allows for merging the two data sets so that the final set contains:
12 pairs $\times$ 2 conditions = 24 data points (see Figure 5.8).

3 out of the 24 participants, i.e. 12.5%, participated in experiments of previous chapters (Chapter 3, 4). However, the experiment in this chapter is based on a completely different setup to which previous experience does not apply to an extent where one would expect an impact on the results.

5.4 Results

The analysis of the experiment data is divided into two parts, a quantitative and a qualitative assessment.

5.4.1 Quantitative Analysis

To assess the effect of the robot's condition on task performance and collaboration, we compare completion time and dialogues' word counts as well as TLX and SUS results between the two condition groups. Concerning these metrics, a series of paired t-tests was performed with the robot's condition as an independent variable. The results are summarised in Table 5.1 and illustrated in the diagrams of Figure 5.10.

Concerning task performance, a significant ($p < .001$) decrease of mean completion time, from 189.3 s to 138.2 s, is observed when the robot is in assisted mode compared to the non-assisted mode (see Figure 5.10a). Hence, the teams performed the task 37% faster in assisted mode. Furthermore, no significant correlation (as per Pearson [18]) between completion time and demographics was found for either of the participant groups, which is also true for the prior gaming experience.

Regarding the dialogue required for task coordination, from non-assisted to assisted mode there is a significant drop of word count for the remote users ($p = .005$) as well as for the local users ($p = .008$) leading to a total word count reduction of 38% (Figure 5.10b,

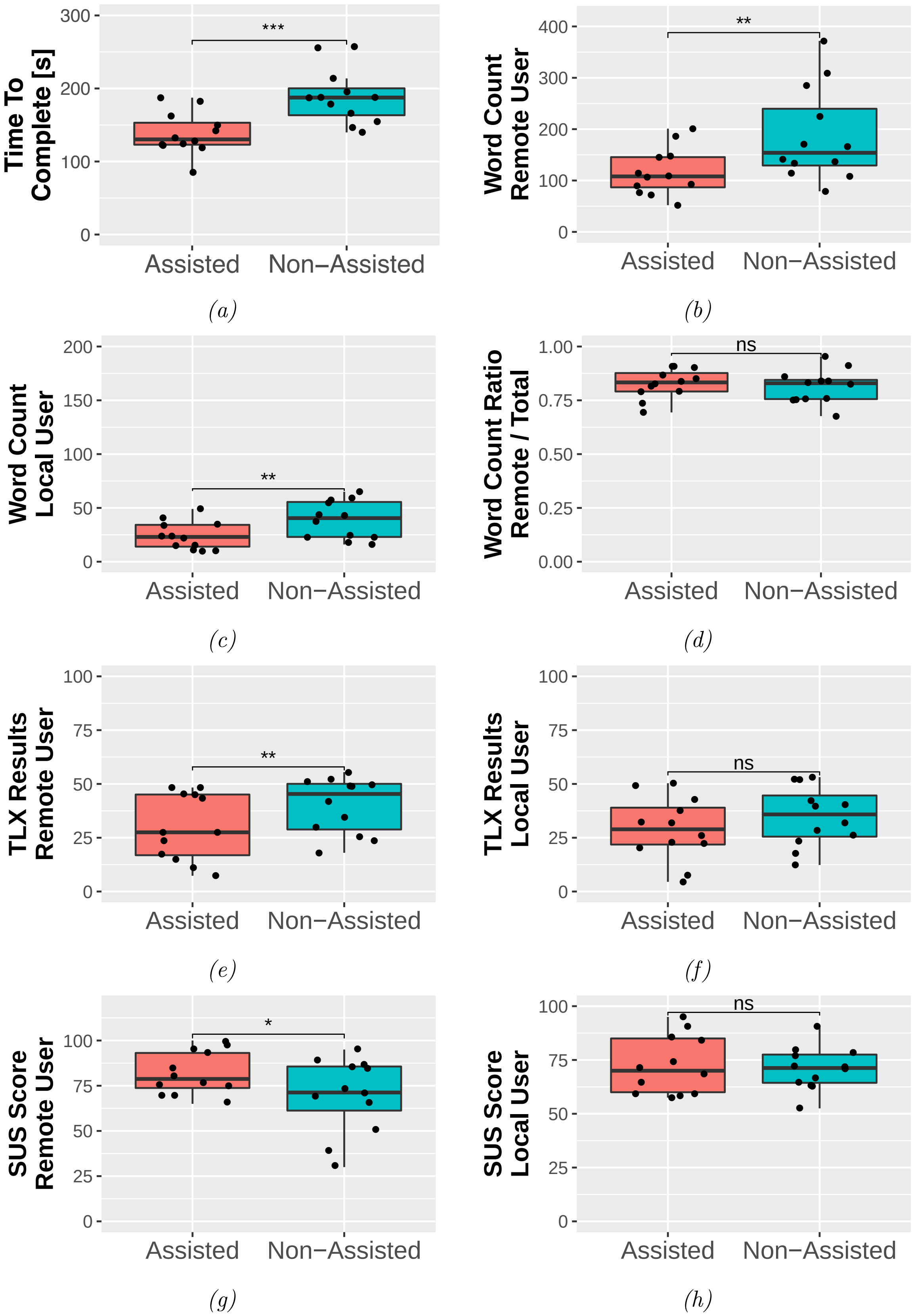


Figure 5.10: ***Comparative Analysis of Robot Conditions.*** *The diagram shows completion times, word counts, TLX and SUS scores for the assisted and non-assisted condition of the robot. Starred samples yield a significant within-subject difference depending on the robot's condition (see Table 5.1). Levels of Significance: $p \geq .050$; * $p < .050$; ** $p < .010$; *** $p < .001$.*

		Assisted		Non-Assisted		t-test Results		
		Mean	SD	Mean	SD	t	p	
Time To Complete [s]		138.2	28.8	189.3	37.9	t(11) = -6.010	p < .001	***
Word Count	Remote User	116.1	45.6	186.8	90.7	t(11) = -3.476	p = .005	**
	Local User	24.2	13.0	38.8	17.5	t(11) = -3.186	p = .008	**
	Total	140.3	54.0	225.7	95.2	t(11) = -3.857	p = .002	**
	Ratio [%]	82.8	6.7	81.3	7.7	t(11) = 0.723	p = .484	
TLX	Remote User	30.0	15.4	40.0	13.0	t(11) = -4.281	p = .001	**
	Local User	29.0	14.7	35.0	14.7	t(11) = -2.113	p = .058	
SUS	Remote User	81.9	11.7	70.0	20.7	t(11) = 2.828	p = .016	*
	Local User	72.5	13.3	70.8	9.9	t(11) = 0.795	p = .443	

Table 5.1: ***Summary of Quantitative Analysis.*** *t-test results for the analysis of differences in average completion time, word count, TLX and SUS scale depending on whether the robot's assistive features were enabled. Starred values indicate a significant difference. The distribution of the data can be seen in the associated diagrams in Figure 5.10. Levels of Significance:* ** $p < .050$; ** $p < .010$; *** $p < .001$.*

5.10c). The effect of the robot's condition on the word count ratio between remote and local user is non-significant.

In terms of participants' perceived workload, the robot's assistance significantly ($p = .001$) decreases from an average TLX score of 40 to 30 (lower is better), i.e. by 25%. However, no significant effect ($p = .058$) is identified for the TLX score of the local user (see Figure 5.10e and 5.10f). In both conditions, the TLX ratings by the local users are close to the lower boundary of the scale.

The overall system usability, as per the SUS questionnaire, is rated significantly higher ($p = .016$) by the remote users when the robot is in Assisted Mode. Per the TLX ratings, the robot's condition (i.e. Assisted/Non-Assisted Mode) has no significant ($p = .443$) effect on the SUS ratings for the local user group and they approach the upper boundary of the scale in both cases. The SUS results are summarised in Figure 5.10g and 5.10h.

We note that for the Assisted Mode, the perceived task load, as per TLX, was rated low, i.e. ≤ 30.0 out of 100 (lower is better) for both user groups. Furthermore, the SUS ratings were high for both groups, i.e. with usability scores ≥ 72.5 out of 100 (higher is better).

5.4.2 Qualitative Analysis

The major novelty of the proposed remote assistance scenario is the remote user's ability to physically interact with the workspace environment. This introduces new solution strategies and behaviours which are reflected in the collected video material.

Across the different participant pairs, a general problem-solving strategy could be observed, which consists of four phases: scene exploration, spacial guidance, local task solving and retraction. During scene exploration, the remote user uses both camera views to diagnose a problem. Once the problem is identified, the local user is guided to a scene object through tooltip gestures and verbal instructions. The ratios between the usage of those two means of communication vary strongly between remote users. While some participants gave many verbal instructions, others preferred the use of the tooltip for directions. Notably, in one instance, the remote user was able to guide the local user through the pipe checking task without using any verbal instructions. Instead, they used pointing gestures with the manipulator for navigation to a pipe that required checking and made the robot retreat to a crouch position to signal to the local user that the checking process was completed.

Another observation was that in this phase the remote user can in some instances control the motion speed of the local user by the amount of tip deflection. The bigger the angle with which the remote user steers the tip to one side, the higher the speed of the local user following it. This multimodal guiding strategy led to a more smooth interaction compared to ones where the remote user would solely use verbal instructions for navigation.

After the object of interest is reached, there is a transition from shared control over the robot to remote user control as the local user holds the robot in place during manipulation. After a subtask is solved, the local user retreats away from the workspace to allow the remote user further exploration of the scene and the working cycle starts over again.

5.5 Discussion of Remote Assistance Study

In terms of task performance, we found that the handheld robot's autonomy and task knowledge contributes to a more efficient collaboration as it reduces the time to complete the task, which supports **H1**. Regarding communication between the remote user and the local user, the qualitative results show that the remote control of the robot's tip extends the means of communication as it can be used for deictic gestures, such as pointing for navigation, which is true for both modes of the robot.

H2 is partially supported as the robot's assistance features reduce the amount of verbal communication required to solve the task. We suggest that the reason for this is that the robot's aiming feature replaces the remote user's instructions for low scale motion. For example, in the non-assisted condition, remote users often relied on verbal instructions such as *Go a bit more to the left, please* or *Up, up, up* during navigation to a task object. Whereas, when aiming was assisted, it was more clear to the local user where to go since they could just follow the tip's direction until reaching their aim. Furthermore, the robot's task knowledge accelerates processes which are otherwise interrupted through information requests.

The findings concerning **H1** and **H2** were expected for this specific task, given they are a direct consequence of the system design. That is, we tried to make the task easier for the remote user, e.g. through task delegation and information short cuts. However, what is interesting is that the pointing gestures during the navigation phase were more effective when the robot completed them autonomously compared to manual execution by the remote user. There are several factors that we believe contribute to this observation. Firstly, once the robot knows which object to go towards, the robot is fast at readjusting the tip towards the object when the robot is moved around the scene. In comparison, it seems to be hard for the remote user to coordinate the robot's tip direction given the motion introduced by the local user and the visual feedback. Secondly, it seems that local users sometimes struggled to distinguish between intended tip gestures and tip motion that results from visual exploration. However, in autonomous mode, the robot's motion becomes distinctively faster and more precise, particularly concerning world-stabilisation. This makes it easier for local users to distinguish exploration from navigation. Additionally, the robot retreats into crouch position after a local job is completed. At the time of its design, this was intended to bring the end effector to a safe position before handing over its control back to the remote user. However, we observed that this also helped with the collaborative workflow as it signalled to the local user to prepare for a transition to another subtask.

As the introduction of task knowledge in assisted mode changes the distribution of knowledge between the three interacting agents, we expected to observe a change in relative verbal communication amounts. However, for the given task this part of **H2** is not supported by the results. Presumably, this is because in assisted mode the robot's task knowledge facilitates navigation, which cuts down required verbal commands by the remote user. At the same time, the robot executes local tasks autonomously, which reduces the number of instances where the local user has to report the system's current state. That way, the robot's assistance reduces the communication amount evenly so that the ratio won't change.

In terms of the robot's effect on collaboration quality, we argue that a collaborative robot should reduce workload and offer high usability for users. While this is supported by the overall TLX and SUS results for both users, improvement through the robot's assistance feature could only be found for remote users. This is less surprising, given that this work is focused on remote access and the assistive features were mainly designed to facilitate remote control of the robot. For example, the aiming feature takes away task load from the remote user whereas the difference is small for the local user whether the tip is controlled by the remote user or through the robot's assistance. Therefore, **H3** and **H4** are partially supported and we conclude that the assisted mode mainly benefits the work of the remote user.

The fact that the performance was independent of demographics and prior gaming experience, indicates that the system is suitable for people without technical expertise about the robot or the user interface. This makes the robot and the remote control system accessible for a broad range of applications, however, we suggest that different tasks require varying training approaches.

We note that the setup described here is a first attempt of collaborative telepresence with two humans in the loop. As such, our observations might be specific to the experimental task presented here. Further diverse examples are required to clarify to what extend our findings generalise to other tasks. Here, we combine remote guidance and telemanipulation which, on their own, are well established fields today. The proposed concept might not outperform the existing methods in their respective domains. For example, pure guidance might still be more effective using AR/VR and some remote tasks might be more efficiently solved using specialised robots (e.g. surgery). However, we believe that combining the two, as presented here, will make a difference when the task is suitable for task delegation to the robot. Beyond maintenance tasks, this could include elements from other industries, e.g. welding, farming or drilling. One might argue that some of these tasks could be completed more effectively by humans but a human with the right skills is not always at hand (e.g. a trained welder). We further suggest that an expert could use the proposed system to demonstrate how to solve a task through the robot to train a novice technician. They could then do it better or faster in the future if the task is indeed such that it is easier to do for a trained human than for the robot, though this is subject to future research.

Our study demonstrates that making the robot part of the team as an entity that makes decisions rather than just a remote controlled device, is of great benefit to facilitate efficient remote collaboration. The robot's autonomous features and decisions require task knowledge as a basis. In the presented example, the task was simple enough to implement essential information of all task-relevant objects and their current states. However, this might not be feasible for broader and more complex applications. In this case, we suggest that specific subtasks could be learnt from demonstration, e.g. how to identify an object and how to interact with it given a task's constraints. In that context, it would be interesting to explore whether there is more general task knowledge that could serve as a basis for the robot to quickly adapt to new tasks as well.

5.6 Chapter Conclusion

So far, remote collaboration and assistance between two humans has been limited to audio-visual instructions and feedback [21, 53, 54, 140]. At the same time, previous works on robot-based collaboration setups focused on specialised telemanipulation [16, 216, 242, 247] and remote controlled mobile platforms [205]. The work presented in this chapter explores a combination of both fields so that the resulting system benefits from the strengths of each. For remote assistance systems, that is the convenience of being able to rely on a human partner on the local work site while the telemanipulation aspect grants direct and immediate physical access to the helping expert.

In this chapter, we investigated the use of a handheld robot in our proposed collaborative remote assistance setup within a helper-worker scenario. We assessed the robot's capabilities through user studies, where a local user operates the handheld robot while being assisted by a remote user who gives verbal instructions and controls the robot's tip.

Regarding the research question **Q5** *"How does the handheld robot's autonomy and task knowledge affect performance and communication in a remote assistance setup?"*, our studies show that a handheld robot can mediate task information and physical interaction in collaborative assistance. Importantly, the robot's partial autonomy improves task performance with respect to time efficiency, workload and required communication bandwidth in the context of the specific experiment task presented here. Namely, the task was completed 37% faster, remote user's work load decreased by 25% and the required verbal communication by 38%. This means that when the robot was helping, the task was completed in a shorter time, while being easier to carry out for the remote user as less instructions were needed thanks to the delegation of subtasks to the robot.

From our observations, the use of the robot proved effective to facilitate diagnosis, guidance and to interact with task objects. Enabling the operator to control the angle of the detail view camera allows for dexterous aiming during inspection. Equally, spacial gestures supported effortless communication as it could replace wordy instructions and verbal information requests. Notably, aims concerning navigation and transitioning between subtasks were easier to interpret by local users when the arm was controlled by the robot. This knowledge is useful to inform the design of future remote assistance systems, especially for cases where the local user is confronted with a broad variety of choices such as in our specific case.

The concept of object selection for task delegation to the robot is key to reduce the operator's cognitive task load, which we believe generalises to other telepresence setups. It might also help with problems that are introduced by time-delays. As the robot is autonomous for short durations, a fast feedback response from the remote user might become less critical. We suggest that this could be investigated in future research of this domain.

This work is a first attempt to evaluate handheld collaborative robots in a remote assistance scenario. Our research delivers first indications that such a setup can overcome current robot constraints such as incomplete task knowledge and limited motion competences by using already well-established communication technologies.

Chapter 6

Conclusion and Further Work

Handheld robots are a new generation of tools that hold knowledge about their own purpose and the task at hand, which can be used to leverage task execution for both expert and novice users. In this thesis, we investigated the interaction between a handheld robot and one or several users with a focus on user prediction and remote collaboration. Previous work in this domain concentrates on surgery, free-hand fabrication and other specialised tools (Section 2.3). Recent research regarding more general purposed handheld robots focused on the hardware design, how the robot can indicate its internal states and how it can communicate feedback and instructions, e.g. through display information or rudimentary gestures (Section 2.2). While these works yield promising results in single-user applications, we believe that handheld robotics is still at an early stage. To unfold their full potential, handheld robots should be able to estimate their user's state, which is required for any collaboration style where decision making is a bilateral process. Furthermore, this work explores the idea of using a handheld robot as a mediator between two human collaborators in a helper-worker setup. The presented concepts are validated through experimental studies with a focus on collaborative performance and users' work experience.

6.1 Thesis Summary

This thesis explored new human-robot interaction strategies for handheld robots. Chapter 1 introduced the reader to the concept of handheld robots and placed it in the more general historic context of hand tools. Furthermore, it outlined possible applications and which problems need to be solved to realise those.

The background section (Chapter 2) presented related work in the domain of handheld robots, e.g. existing wearables and intelligent tools with a focus on the hardware designs that were used as a basis for our experimental setups. Furthermore, the chapter summarised transferable interaction concepts from other robotic domains concerning intention prediction, remote guidance and telemanipulation.

Chapter 3 addressed the problem of user-to-robot feedback following a vision-based approach that is based on the finding by Land et al. [111–113], which suggest that a human's eye gaze precedes the steps throughout a task. Therefore, this information can be used to inform the robot about the user's decisions. A remote eye tracker was integrated with the existing handheld robot hardware and coupled to a motion tracking system for 3D-eye gaze construction. In experimental studies, the gaze model was used as a proxy for user attention and this information was used to bias the robot's autonomous motion. A target reaching task was used to demonstrate that collaboration between the user and the robot is more smooth when the robot's autonomous decisions are biased by information about the user's gaze. Especially for higher speed demands, this mode exceeds the performance of doing the same job manually or when the robot is completely gaze steered, i.e. without the robot using its task knowledge. While it compares well with the fully autonomous mode, we found that the gaze-biased collaborative mode is more accepted among users as they felt more in charge rather than having to follow the robot's plan at all times. This suggests that in human-robot collaboration there can be a trade-off between task performance and user satisfaction. At the same time, we demonstrate that the estimation of users' attention during the task was able to help to prevent *automation surprise* when humans and robots work together. As the attention system is rather reactive, i.e. it responds to users' visual attention but does not infer intention, our findings are limited to tasks with low task complexity such as pick-and-place, or following a trajectory with only few alternatives.

Chapter 4 refined the idea of using users' eye gaze to determine their plans. The data of the attention model from the preceding chapter (Chapter 3) is used to make predictions about which object the user wants to interact with in the proximate future. This intention model is based on an SVM, a supervised learning model, that maps an object's attention profile over time to a probability for interaction. We used a block copy task to validate the intention model as it is characterised by a variety of ways to complete the task while some dependencies exist among task objects in a key-lock matching fashion. The results show that the model can reliably predict user actions one step ahead (1.5s) and in real-time and that its use significantly decreased user frustration in the specific block copy task. In accordance with the findings of the attention study, we show that the most productive (fully autonomous) mode is not always the most effective one concerning human-robot collaboration as users preferred to be in charge of the high-level planning part of the task.

In Chapter 5 we extended the exploration of human-robot interaction by introducing another user for a collaborative human-robot-human setup. The concept is built on the idea that the robot could serve as a mediator between a remotely located expert and a local novice user hosting the robot. Maintenance is a typical example of a scenario where a technician might need help from an expert, who might be located at a different continent. Thanks to the handheld robot, the remote user has visual and physical access to the work-site while the local user was able to help with carrying the device through uncontrolled environments and through providing local task state information. To adapt the robot to the requirements for remote collaboration, it was equipped with cameras, which are connected to a remote workstation that allows for visual feedback and controlling the robot. The proposed inspection and maintenance of a pipe system was used as an example task for the experimental proof-of-concept study. The core of this research is the integration of the robot in the control loop, i.e. the remote user is in charge of task planning but can delegate local tasks to the robot. The idea behind this is that this would free the expert's cognitive capacity to plan the subsequent step in the task rather than having to deal with the fine-grain motion control of the robot. Our results suggest that the use of the robot's assistive autonomous features improve overall teamwork speed by 37%, while the workload of the remote user is decreased by 25% in the given task.

6.2 Main Contributions

This research contributes to a better understanding of how humans can interact with a handheld robot for a more natural and effortless collaboration with the assisting robot as a teammate that is involved in decision processes. Starting points for the research presented in this thesis are five research questions that were formulated to define the new challenges in handheld robotics. In the following, contributions for the proposed research questions are summarised.

Q1 How can user attention be used to enhance cooperation with handheld robots? (Chapter 3)

- We demonstrate that eye gaze tracking can be used to estimate what the user directs their attention at when working with the handheld robot. Our proposed framework allows for gaze tracking in 3D space, based on a combination of a remote gaze tracker and a motion capturing system. The results of the accuracy and calibration experiments define the limits of the tracking setup and inform the boundaries of applications for the handheld robot.
- We introduce gaze-based user attention estimation and demonstrate how this information can be used to parametrise the robot's cooperative behaviour, namely, with the robot navigating to objects with visual focus.

Q2 How does the incorporation of attention affect task performance and the user's perceived task load? (Chapter 3)

- To assess the attention system, we propose a reaching task as basis for an experimental setup that allows measuring collaborative performance for validation of the model. As some parts of the experiment task are simulated, they can be parametrised easily, e.g. concerning speed demands and cognitive loads. This makes the proposed setup adaptable to future research experiments.

- The results of the attention experiment suggest that gaze tracking allows the robot to move more predictable with users feeling more in control of the task while maintaining the already high performance of the handheld robot in the fully autonomous mode.

Q3 How can user intention be modelled in the context of a handheld robot task? (Chapter 4)

- Beyond the attention model, that quantifies the level of user attention at a given time, we introduce an intention model. It is based on the user's gazing pattern over time and allows for online predictions of pick up actions with up to 87.94% accuracy, 500 ms ahead and dropping actions with an accuracy of 93.25%, 1500 ms ahead in real-time for the given task.

- We show that context information, such as the link between objects for a given task, can increase the model's accuracy, i.e. from 68.34% to 87.94%. We suggest that this property generalises to other models concerning user anticipation in human-robot collaboration too.

- The introduced version of a block copy task allows for a parametrisation of cognitive demands and to measure the collaboration performance as it demands timely decisions from the user, high-level planning and task coordination between the two parties. This could be used to benchmark future collaboration concepts.

Q4 To what extent does intention prediction of users affect the cooperation with a handheld robot? (Chapter 4)

- To assess the intention model in action, we use obedience and rebellion as two principal modes for the robot's anticipatory behaviour. In absence of universally accepted physiological metrics, this serves as a proxy to evaluate the intention model by comparing user frustration levels between the two conditions. We show that the proposed intention model allows the robot to align its plans to the user's intention, which leads to a decrease of frustration and an increase in productivity.

Q5 How do the handheld robot's autonomy and task knowledge affect perfor-

mance and communication in a remote assistance setup? (Chapter 5)

- We explore remote collaboration as a new application for handheld robots in a worker-helper setting. The helper is in the role of a remotely located expert that instructs and acts through the robot to assist a local user.

- The proposed pipe system maintenance task imposes common challenges in remote assistance, such as problem diagnosis, guidance, navigation through a 3D environment and collaborative problem-solving. While this was used to assess the handheld robot setup, we suggest that this could serve as a benchmark task to compare remote assistance systems. This is because it covers a broad variety of collaboration challenges, while completion time, SUS and TLX values serve as meaningful and generally applicable measures to compare different systems.

- Following the concept of the handheld robot as a partially autonomous member of the team, some parts of the task can be delegated to the robot on a local scale, e.g. aiming, turning open/shut a valve and placing a sensor for the detection of hidden cracks in the pipe system. Our results demonstrate that the introduction of the robot's autonomous features can reduce the cognitive load perceived by the remote user in the experiment task by 25% and increase collaborative performance by 37%.

- As per the results of the qualitative analysis, there are new collaborative behaviours that emerge between the three cooperating agents, i.e. the remote expert, the robot and the local user. We identified a repeated sequence of exploration, guidance, local task solving and retraction as a common high-level task solving strategy. This information is useful to guideline future designs of remote collaboration systems.

6.3 Discussion and Outlook

This work focused on the exploration of collaboration between a handheld robot and its user as well as the one between two humans with the robot as a platform for the transmission of instructions, information and actions. While this thesis covers theories around these concepts and demonstrates their effectiveness through case studies, we still see handheld robotics at an early stage with great potential for exciting future projects. There are still many areas to explore concerning means of human-robot feedback and robot-to-human communication. There is a vast variety of conceivable real-world applications for handheld robots that present many new research challenges. Furthermore, we discuss which obstacles need to be overcome towards this vision and how we believe this technology could change and impact our society.

6.3.1 Human Factors

One of the major challenges in exploratory research in the field of HRI is the sample size that experimenters rely on, which are often much smaller than data sets of other robotic domains. Like in this work, they usually involve a few 10s of participants as data collection is both time consuming and expensive. While we had a balanced gender equality in most of our experiments, the demographics were constrained by eligibility regulations, i.e. experiment participants had to be members of University of Bristol and willing to offer around 60 min volunteering time. This resulted in narrowing the cohort to an academic background and an average age of 27.1 years. All experiments were designed in a way that no prior expertise was required as everyone was trained to a level where they could handle the systems well before we started experimenting. We note that most participants came from technical backgrounds but the fact that there were no notable differences between performances of them compared to other participants suggests that the effect on the results were negligible. Moreover, we introduced questions to probe their technological familiarity, e.g. we asked them about their gaming habits, which did not correlate with performance in any of our experiments. However, beyond the scope of this work it would be interesting to investigate how people interact with the robot who are non-computer natives. For them, it might be harder to handle the robot but on the other hand, they might greatly benefit from the robot's capability to adapt to their attention.

Another limitation of our work is the fact that for all experiments, both the system and the task was new to the experiment participants, i.e. we did not investigate differences between levels of expertise. One could argue that, initially, the robot's assistive features are beneficial while the user is at a novice level but might at some point turns into a burden one they have a deeper understanding of the task. The question remains open when and in what situation the user should transition to completing the task without the robot or without interfering autonomy. We suggest that this transition could be smoothed through a gradual decrease of the robot's autonomy with increasing expertise. For example, the robot could stop deciding which object to interact next with but still assist in terms of positioning accuracy, e.g. in a grid-lock manner for pick and place tasks. Detecting the level of a person's expertise in a given task is subject to current research, e.g. [42], and could be applied to such HRI problems in the future.

Apart from demographics and expertise, a person's personality is an important human factor in HRI-related research. When we conducted the frustration-based intention validation experiments (Section 4.6), it was rather surprising to observe the diverging reactions of subjects to the rebelling robot, i.e. when it refused to follow users' intention. While some were confident enough to (correctly) blame the robot, others assumed that the robot's dominance was purposed to correct their faults. In this work, this phenomenon was not dominating to an extent that it would have compromised our results. However, it could be an interesting aspect to study, e.g. as part of a personality test in the pre-screening phase of similar experiments. This could help to gain more detailed results that take into account possible interaction between the robot's behaviour and a person's personality.

6.3.2 User Perception

The majority of this thesis addresses the perception of the user with a focus on their intention (Chapter 3 and 4). While gaze data proved to be a good choice for this specific purpose, other channels might be a better fit for other types of feedback. For example, the current stress level of the user might be an interesting property to detect, e.g. the robot could adapt its pace or even suggest the user to take a break from working. For an online, stress detection, physiologic signals could be key for immediate feedback. Examples are Brain-Computer Interfaces (BCI) [192], grasping force (e.g. at the handle) [212], or cardiovascular metrics such as Galvanic Skin Response (GSR) [91], Heart Rate (HR) and Heart Rate Variability (HRV) [50], i.e. the fluctuation of periods between heartbeats [134].

As HRV receives increasing attention in literature on stress detection [7, 8, 50, 70], we investigated its use for the handheld robot in a research excursion (see Appendix B). We found that, at least for optical HRV detection, this metric was less reliable than reported stress levels. For this reason, in our research, we relied on questionnaires to determine frustration levels (i.e., work in Section 4.6). Nonetheless, we encourage to follow the route of physiologic stress detection. Improved sensor technology might deliver more accurate HRV signals for mobile applications in the near future. In addition, fusing HRV sensors with other metrics, e.g. BCI and grasping force sensing might lead to much higher accuracy in stress prediction.

Stress feedback could be particularly useful in tasks where the robot helps the user to learn a new skill. One could think of a scenario where the user completes a task for the first time with the robot to then continue without it. While the effectiveness of this concept is subject to further research, we suggest that stress detection could help the robot to adapt to the user and their skill levels. However, this would require knowledge about how fast the detection system responds to a stress stimulus. We suggest that the work at hand presents a broad variety of research questions that would be exciting to explore.

For completeness, we also note that natural language processing is another conceivable channel for human-robot communication. This would enable the user to input commands without using the already occupied hands. However, to our opinion, this option is less preferable since it can make interactions rather awkward and slow, which can be in the way of an intuitive and fluent usage of a device.

6.3.3 Robot-to-Human Communication

With regard to a single-user setup, the robot-to-human communication could be further explored. This thesis focused on the human-robot side because possible feedback means for handheld robots have already been studied extensively by Gregg-Smith and Mayol-Cuevas [60] and so there was a higher demand for user perception. Nonetheless, we suggest that feedback could be provided beyond HMDs and mounted displays. For example, new generations of pico projectors and monochrome laser galvanometers could augment through scene context information similar to [93]. This could also be useful for the presented remote collaboration setup since the helper could use in-scene annotations as another way of instructing the local user. Equally, more meaningful arm gestures would be another possible channel with the advantage of this not requiring any additional hardware. Compared to AR feedback, these approaches have the advantage of being less invasive for the user. This suits the notion of a handheld self-contained tool rather than shifting the system to the field of wearables.

Furthermore, with the proposed intention model in place, the question arises how this information could be used. For example, the robot could feature alert sounds or handle vibrations when it detects that the user is about to do a mistake before it happens. This could be of particular practical interest for irreversible processes such as welding or glueing.

6.3.4 Remote Collaboration through Handheld Robots

Concerning handheld robots for remote collaboration, this work can be seen as a first attempt at using a mobile robot for human-robot-human collaboration (Chapter 5). While the current general-purpose design could be further specialised for this application, e.g. through the integration of the above-mentioned projectors, we suggest that there is more to explore on the remote user side as well. For example, the question remains open whether the robot's partial autonomy could help to compensate delay effects in telemanipulation.

Another interesting aspect to explore would be the incorporation of an intention system. Similar to the presented single-user case, this could be applied to the remote workstation, i.e. the robot anticipates the plans of the remote helper. In accordance with related studies in the field of intention prediction, our results suggest that different tasks require different prediction models. For example, we used two separate models to predict picking and placing actions for the best results. Therefore, we expect that the prediction of the decisions of the remote user would require a model tailored to this case too. This goes beyond the scope of this thesis but could serve as a starting point for future research.

6.3.5 *You Are Free*: Releasing the Robot from the Lab Environment

We point out that all our experiments were carried out in a controlled lab environment and the tasks were kept rather general, which was required to assess the proposed concepts and theories. Given the vast range of possible applications, e.g. assembly, weeds removal in organic agriculture, assisted welding, sculpting and remote assistance in various ways, there is a pressing demand to initialise field studies with real-world application. A fundamental step towards this is to replace the motion tracking system with an on-board localisation solution, e.g. SLAM for mobile devices such as [142]. Additionally, the robot would require a form of object detection, e.g. through Computer Vision (CV) [181], which is required for an informed decision concerning interactions.

Solving these problems would open up this technology to setups with multiple handheld robots, e.g. *crowd constructing* of buildings or *crowd farming*. Our vision is that an interconnected fleet of handheld robots would enable a group of novice users to work together on large-scale projects. This presents new challenges with respect to scalability of handheld robot technology. Apart from localisation challenges, a fleet of robots would require a shared knowledge base, e.g. a cloud system. This is necessary to store task information such as completion states, which is required to derive individual goals for the robots and their users. With regards to our findings of Chapter 5, we could imagine that while some robots of the fleet are autonomous, others might be controlled by remote experts to give instructions for processes that have not been automated yet. Another limitation of scalable solutions is the bottleneck of time-consuming programming of the robots. This could be solved through facilitating the programming process, e.g. by introducing learning from demonstration, which would significantly decrease the number of technical experts required to setup a robot fleet.

6.3.6 The Impact of Handheld Robots on Society

Robotics and AI are one of today's major disruptive technologies and as roboticists, an important part of our responsibility is to think about the impact of our research on society. Research in this field is guided by the vision of a future where heavy, dangerous or repetitive work is carried out by machines, which frees capacities for more creative tasks to be completed by humans. However, in many instances, the reality looks somewhat different. The introduction of robots in a capitalistic labour market comes with the cost of fear of job loss. This connotation has attracted considerable attention of the general public over the recent years [171]. This statement is backed up by an increasing number of scholarly articles, as summarised in a recent *MIT Technology Review* [234]. Furthermore, working conditions for workers that have not been substituted (yet), do not always improve. For example, workers at large and highly automated warehouses face exhaustion as they try to keep pace with robots while there is little interaction with other human workers [199]. To date, it is subject to speculation to what extent automation leads to a job gain [234] or whether mechanisms such as a universal basic income [88] or a tax on robots [246] could combat economic uncertainties and narrow wage gaps.

Concerning the developments in handheld robotics, we see this as a class of automation technologies that have the potential to complement human workers through a synergy effect, rather than replacing jobs. We believe that handheld robotics can enable human-robot teams to complete jobs that the respective party could not do without the other one. For example, locomotion and navigation through uncontrolled environments, i.e. outside factory lines, is still one of the major challenges in current robotic research, while most humans would find it trivial. Unlike humans, robots can process vast amounts of data in real-time, which can augment human work in various ways. For this reason, we believe that any form of collaborative HRI is an important contribution to a future society with a meaningful work culture.

References

[1] Handheld Robotics: handheldrobotics.org.

[2] Firas Abi-Farraj, Nicolo Pedemonte, and Paolo Robuffo Giordano. A visual-based shared control architecture for remote telemanipulation. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4266–4273. IEEE, November 2016.

[3] Firas Abi-Farraj, Takayuki Osa, Nicolo Pedemonte Jan Peters, Gerhard Neumann, and Paolo Robuffo Giordano. A learning-based shared control architecture for interactive task execution. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 329–335, May 2017.

[4] Reza Abiri, Soheil Borhani, Eric W Sellers, Yang Jiang, and Xiaopeng Zhao. A comprehensive review of EEG-based brain–computer interface paradigms. *Journal of Neural Engineering*, 16(1):011001–22, January 2019.

[5] George Adamides, Christos Katsanos, Yisrael Parmet, Georgios Christou, Michalis Xenos, Thanasis Hadzilacos, and Yael Edan. HRI usability evaluation of interaction modes for a teleoperated agricultural robotic sprayer. *Applied Ergonomics*, 62:237–246, July 2017.

[6] David W Aha and Alexandra Coman. The AI rebellion: Changing the narrative. In *AAAI-17 Workshop on AI, Ethics, and Society*, 2017.

[7] Hussein Al Osman, Mohamad Eid, and Abdulmotaleb El Saddik. U-biofeedback: a multimedia-based reference model for ubiquitous biofeedback systems. *Multimedia Tools and Applications*, 72(3):3143–3168, 2013.

[8] Hussein Al Osman, Haiwei Dong, and Abdulmotaleb El Saddik. Ubiquitous Biofeedback Serious Game for Stress Management. *IEEE Access*, 4:1274–1286, 2016.

[9] T Bailey and H Durrant-Whyte. Simultaneous localization and mapping (SLAM): part II. *IEEE Robotics & Automation Magazine*, 13(3):108–117, August 2000.

[10] G H Ballantyne. Robotic surgery, telerobotic surgery, telepresence, and telementoring. *Surgical Endoscopy*, 16(10): 1389–1402, October 2002.

[11] Garth H Ballantyne and Fred Moll. The da Vinci telerobotic surgical system: the virtual operative field and telepresence surgery. *The Surgical clinics of North America*, 83(6):1293–304– vii, December 2003.

[12] Dana H Ballard, Mary M Hayhoe, and Jeff B Pelz. Memory Representations in Natural Tasks. *Journal of Cognitive Neuroscience*, 7(1):66–80, 1995.

[13] Brian C Becker, Robert A MacLachlan, Louis A Lobes, and Cameron N Riviere. Semiautomated intraocular laser surgery using handheld instruments. *Lasers in Surgery and Medicine*, 42(3):264–273, March 2010.

[14] Brian C Becker, Sandrine Voros, Louis A Lobes, James T Handa, Gregory D Hager, and Cameron N Riviere. Retinal vessel cannulation with an image-guided handheld robot. In *2010 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2010)*, pages 5420–5423. IEEE, 2010.

[15] Brian C Becker, Robert A MacLachlan, Louis A Lobes, and Cameron N Riviere. Vision-based retinal membrane peeling with a handheld robot. In *2012 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1075–1080. IEEE, 2012.

[16] Moez Bellamine, Norihiro Abe, Kazuaki Tanaka, and Hirokazu Taki. Remote machinery maintenance system with the use of virtual reality. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)*, pages 38–43, 2002.

[17] Moez Bellamine, Norihiro Abe, Kazuaki Tanaka, Peng Chen, and Hirokazu Taki. A Virtual Reality Based System for Remote Maintenance of Rotating Machinery. In *Embedded and Ubiquitous Computing*, pages 164–173. Springer, Berlin, Heidelberg, August 2004.

[18] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*. Springer, 2009.

[19] David Beymer and Myron Flickner. Eye gaze tracking using an active stereo head. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)*, pages II–451, 2003.

[20] C D Binnie and P F Prior. Electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, 57(11): 1308–1319, November 1994.

[21] M Bordegoni, F Ferrise, E Carrabba, M Di Donato, M Fiorentino, and A E Uva. An application based on Augmented Reality and mobile technology to support remote maintenance. In *proceedings of 7th international Symposium on Artificial Intelligence, Robotics and Automation in Space, iSAIRAS*, pages 131–135, 2014.

[22] Sébastien Bottecchia, Jean-Marc Cieutat, and Jean-Pierre Jessel. T.A.C: Augmented Reality System for Collaborative Tele-Assistance in the Field of Maintenance through Internet. In *Proceedings of the 1st Augmented Human International Conference*, New York, New York, USA, 2010. ACM Press.

[23] D H Bourla, J P Hubschman, M Culjat, A Tsirbas, A Gupta, and S D Schwartz. Feasibility study of intraocular robotic surgery with the da Vinci surgical system . *RETINA*, 28(1):154–158, January 2008.

[24] Sébastien Briot and Wisama Khalil. Homogeneous Transformation Matrix. In *Dynamics of Parallel Robots*, pages 19–32. Springer International Publishing, Cham, June 2015.

[25] Gabriel Brisson, Takeo Kanade, Anthony DiGioia, and Branislav Jaramaz. Precision freehand sculpting of bone. In *In Medical Image Computing and Computer-Assisted Intervention*, pages 105–112. Springer, 2004.

[26] I N Bronshtein, K A Semendyayev, Gerhard Musiol, and Heiner Mühlig. *Handbook of Mathematics*. Springer Science & Business Media, 2013.

[27] John Brooke. *"SUS - A quick and dirty usability scale." Usability Evaluation In Industry*. CRC Press, June 1996.

[28] Jack O Burns, Benjamin Mellinkoff, Matthew Spydell, Terrence Fong, David A Kring, William D Pratt, Timothy Cichan, and Christine M Edwards. Science on the lunar surface facilitated by low latency telerobotics from a Lunar Orbital Platform - Gateway. *Acta Astronautica*, 154:195–203, January 2019.

[29] Clebeson Canuto, Plinio Moreno, Jorge Samatelo, Raquel Vassallo, and José Santos-Victor. Action Anticipation for Collaborative Environments: The Impact of Contextual Information and Uncertainty-Based Prediction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 1151–1160, September 2014.

[30] Michael Casper, Riddhit Mitra, Rahul Khare, Branislav Jaramaz, Brian Hamlin, Brian McGinley, David Mayman, Jeff Headrick, Kenneth Urish, Mark Gittins, Stephen Incavo, and Vivek Neginhal. Accuracy assessment of a novel image-free handheld robot for Total Knee Arthroplasty in a cadaveric study. *Computer Assisted Surgery*, 23(1): 14–20, October 2018.

[31] Yeonjoo Cha and Rohae Myung. Extended Fitts' law for 3D pointing tasks using 3D target arrangements. *International Journal of Industrial Ergonomics*, 43(4):350–355, July 2013.

[32] Dongjune Chang, Gwang Min Gu, and Jung Kim. Design of a novel tremor suppression device using a linear delta manipulator for micromanipulation. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2013)*, pages 413–418. IEEE, 2013.

[33] D Y Choi and C N Riviere. Flexure-based Manipulator for Active Handheld Microsurgical Instrument. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 2325–2328. IEEE, 2005.

[34] Alexandra Coman, Benjamin Johnson, Gordon Briggs, and David W Aha. Social attitudes of AI rebellion: a framework. In *AAAI-17 Workshop on AI, Ethics, and Society*, pages 93–104, 2017.

[35] P Dario, M C Carrozza, M Marcacci, S D'Attanasio, B Magnami, O Tonet, and G Megali. A novel mechatronic tool for computer-assisted arthroscopy. *Robotics and Autonomous Systems*, 4(1):15–29, March 2000.

[36] S D'Attanasio, O Tonet, G Megali, M C Carrozza, and P Dario. A semi-automatic handheld mechatronic endoscope with collision-avoidance capabilities. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation*, pages 1586–1591. IEEE, 2000.

[37] Andrew J Davison, Yolanda González Cid, and Nobuyuki Kita. Real-time 3D SLAM with wide-angle vision. *5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles*, 37(8):868–873, July 2004.

[38] Mathew DeDonato, Velin Dimitrov, Ruixiang Du, Ryan Giovacchini, Kevin Knoedler, Xianchao Long, Felipe Polido, Michael A Gennert, Taşkın Padır, Siyuan Feng, Hirotaka Moriguchi, Eric Whitman, X Xinjilefu, and Christopher G Atkeson. Human-in-the-loop Control of a Humanoid Robot for Disaster Response: A Report from the DARPA Robotics Challenge Trials. *Journal of Field Robotics*, 32(2):275–292, February 2015.

[39] Ye Ding, Myunghee Kim, Scott Kuindersma, and Conor J Walsh. Human-in-the-loop optimization of hip assistance with a soft exosuit during walking. *Science Robotics*, 3(15), 2018.

[40] Kris Doelling, Jeongsik Shin, and Dan O Popa. Service Robotics for the Home: A State of the Art Review. In *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*, New York, New York, USA, 2014. ACM Press.

[41] A Doshi and M M Trivedi. On the Roles of Eye Gaze and Head Dynamics in Predicting Driver's Intent to Change Lanes. *IEEE Transactions on Intelligent Transportation Systems*, 10(3):453–462, 2009.

[42] Hazel Doughty, Dima Damen, and Walterio Mayol-Cuevas. Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[43] Vibekananda Dutta and Teresa Zielinska. Predicting the Intention of Human Activities for Real-Time Human-Robot Interaction (HRI). In *Social Robotics*, pages 723–734. Springer, Cham, November 2016.

[44] Vibekananda Dutta and Teresa Zielinska. Action prediction based on physically grounded object affordances in human-object interactions. In *2017 11th International Workshop on Robot Motion and Control (RoMoCo)*, pages 47–52. IEEE, 2017.

[45] Vibekananda Dutta and Teresa Zielinska. Predicting Human Actions Taking into Account Object Affordances. *Journal of Intelligent & Robotic Systems*, 93(3-4):745–761, April 2018.

[46] Vibekananda Dutta and Teresa Zielinska. Prognosing Human Activity Using Actions Forecast and Structured Database. *IEEE Access*, 8:6098–6116, 2019.

[47] Florian Echtler, Fabian Sturm, Kay Kindermann, Gudrun Klinker, Joachim Stilla, Jorn Trilk, and Hesam Najafi. The Intelligent Welding Gun: Augmented Reality for Experimental Vehicle Construction. In *Virtual and Augmented Reality Applications in Manufacturing*, pages 333–360. Springer London, 2003.

[48] Joshua Elsdon and Yiannis Demiris. Assisted painting of 3D structures using shared control with a hand-held robot. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4891–4897, 2017.

[49] Joshua Elsdon and Yiannis Demiris. Augmented Reality for Feedback in a Shared Control Spraying Task. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1939–1946. IEEE, 2018.

[50] Chiranjeevi Kumar Endukuru and Sonali Tripathi. Evaluation of Cardiac Responses to Stress in Healthy Individuals - A Non Invasive Evaluation by Heart Rate Variability and Stroop Test . *International Journal of Scientific Research*, 5(7):286–289, July 2016.

[51] Paul M Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381–391, June 1954.

[52] J-P Friconneau, V Beaudoin, A Dammann, C Dremel, J P Martins, and C S Pitcher. ITER hot Cell—Remote handling system maintenance overview. *Fusion Engineering and Design*, 124:673–676, November 2017.

[53] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. In touch with the remote world: remote collaboration with augmented reality drawings and virtual navigation. In *VRST '14 Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, pages 197–205, New York, New York, USA, 2014. ACM Press.

[54] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, Tobias Höllerer, and experiment design. World-stabilized annotations and virtual scene navigation for remote collaboration. In *the 27th annual ACM symposium*, pages 449–459, New York, New York, USA, 2014. ACM Press.

[55] Berk Gonenc, Marcin A Balicki, James Handa, Peter Gehlbach, Cameron N Riviere, Russell H Taylor, and Iulian Iordachita. Preliminary evaluation of a micro-force sensing handheld robot for vitreoretinal surgery. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, pages 4125–4130. IEEE, 2012.

[56] R.A.R.C. Gopura, D.S.V. Bandara, K Kiguchi, and G K I Mann. Developments in hardware systems of active upper-limb exoskeleton robots: A review. *Robotics and Autonomous Systems*, 75:203–220, 2016.

[57] Austin Gregg-Smith. *Cooperative Handheld Robots*. PhD thesis, University of Bristol, 2016.

[58] Austin Gregg-Smith and Walterio W Mayol-Cuevas. The design and evaluation of a cooperative handheld robot. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1968–1975. IEEE, 2015.

[59] Austin Gregg-Smith and Walterio W Mayol-Cuevas. Inverse Kinematics and Design of a Novel 6-DoF Handheld Robot Arm. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2102–2109. IEEE, 2016.

[60] Austin Gregg-Smith and Walterio W Mayol-Cuevas. Investigating spatial guidance for a cooperative handheld robot. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3367–3374. IEEE, 2016.

[61] Pavel Gurevich, Joel Lanir, Benjamin Cohen, and Ran Stone. TeleAdvisor: A versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 619–622, New York, New York, USA, 2012. ACM Press.

[62] Eric J Hanly and Timothy J Broderick. Telerobotic Surgery. *Operative Techniques in General Surgery*, 7(4):170–181, December 2005.

[63] Sandra G Hart and Lowell E Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in psychology*, 52:139–183, January 1988.

[64] Mary M Hayhoe, Anurag Shrivastava, Ryan Mruczek, and Jeff B Pelz. Visual memory and motor planning in a natural task. *Journal of Vision*, 3(1):6–6, January 2003.

[65] G Clark Haynes, David Stager, Anthony Stentz, J Michael Vande Weghe, Brian Zajac, Herman Herman, Alonzo Kelly, Eric Meyhofer, Dean Anderson, Dane Bennington, Jordan Brindza, David Butterworth, Chris Dellin, Michael George, Jose Gonzalez-Mora, Morgan Jones, Prathamesh Kini, Michel Laverne, Nick Letwin, Eric Perko, Chris Pinkston, David Rice, Justin Scheifflee, Kyle Strabala, Mark Waldbaum, and Randy Warner. Developing a Robust Disaster Response Robot: CHIMP and the Robotics Challenge. *Journal of Field Robotics*, 34(2):281–304, February 2017.

[66] J A Healey and R W Picard. Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166, June 2005.

[67] M A Hearst. Mixed-initiative interaction. *IEEE Intelligent Systems*, 14(5):14–23, September 1999.

[68] M A Hearst, S T Dumais, E Osuna, J Platt, and B Scholkopf. Support Vector Machines. *IEEE Intelligent Systems and their Applications*, 13(4):18–28, 1998.

[69] Keita Higuch, Ryo Yonetani, and Yoichi Sato. Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5180–5190, New York, New York, USA, 2016. ACM Press.

[70] Nis Hjortskov, Dag Rissén, Anne Katrine Blangsted, Nils Fallentin, Ulf Lundberg, and Karen Søgaard. The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology*, 92(1-2):84–89, 2004.

[71] J R Hoare and L E Parker. Using on-line Conditional Random Fields to determine human intent for peer-to-peer human robot teaming. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010)*, pages 4914–4921. IEEE, October 2010.

[72] Guy Hoffman and Cynthia Breazeal. Effects of anticipatory action on human-robot teamwork efficiency, fluency, and perception of team. In *Proceeding of the ACM/IEEE International Conference on Human-Robot Interaction - HRI '07*, page 1, New York, New York, USA, 2007. ACM Press.

[73] Jin-Hyuk Hong, Youn-Suk Song, and Sung-Bae Cho. A Hierarchical Bayesian Network for Mixed-Initiative Human-Robot Interaction. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 3808–3813, January 2005.

[74] Chien-Ming Huang and Bilge Mutlu. Anticipatory robot control for efficient human-robot collaboration. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 83–90. IEEE, 2016.

[75] Chien-Ming Huang, Sean Andrist, Allison Sauppé, and Bilge Mutlu. Using gaze patterns to predict task intent in collaboration. *Frontiers in Psychology*, 6(1049):1–12, July 2015.

[76] I Hussain, L Meli, C Pacchierotti, and G Salvietti World. Vibrotactile haptic feedback for intuitive control of robotic extra fingers. In *World Haptics*, pages 394–399, June 2015.

[77] Irfan Hussain, Gionata Salvietti, Monica Malvezzi, and Domenico Prattichizzo. Design guidelines for a wearable robotic extra-finger. In *2015 IEEE 1st International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, pages 54–60. IEEE, 2015.

[78] Irfan Hussain, Gionata Salvietti, Leonardo Meli, Claudio Pacchierotti, David Cioncoloni, Simone Rossi, and Domenico Prattichizzo. Using the robotic sixth finger and vibrotactile feedback for grasp compensation in chronic stroke patients. In *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*, pages 67–72. IEEE, 2015.

[79] Irfan Hussain, Gionata Salvietti, and Domenico Prattichizzo. On Control Interfaces for the Robotic Sixth Finger. In *the 7th Augmented Human International Conference 2016*, New York, New York, USA, 2016. ACM Press.

[80] Irfan Hussain, Gionata Salvietti, Giovanni Spagnoletti, and Domenico Prattichizzo. The Soft-SixthFinger: a Wearable EMG Controlled Robotic Extra-Finger for Grasp Compensation in Chronic Stroke Patients. *IEEE Robotics and Automation Letters*, 1(2):1000–1006, July 2016.

[81] Irfan Hussain, Giovanni Spagnoletti, Gionata Salvietti, and Domenico Prattichizzo. An EMG Interface for the Control of Motion and Compliance of a Supernumerary Robotic Finger. *Frontiers in Neurorobotics*, 10:18, 2016.

[82] Irfan Hussain, Leonardo Meli, Claudio Pacchierotti, and Domenico Prattichizzo. A soft robotic supernumerary finger and a wearable cutaneous finger interface to compensate the missing grasping capabilities in chronic stroke patients. In *2017 IEEE World Haptics Conference (WHC)*, pages 183–188. IEEE, June 2017.

[83] Irfan Hussain, Gionata Salvietti, Giovanni Spagnoletti, David Cioncoloni, Simone Rossi, and Domenico Prattichizzo. A Soft Robotic Extra-Finger and Arm Support to Recover Grasp Capabilities in Chronic Stroke Patients. In *Wearable Robotics: Challenges and Trends*, pages 57–61. Springer, Cham, 2017.

[84] Irfan Hussain, Gionata Salvietti, Giovanni Spagnoletti, Monica Malvezzi, David Cioncoloni, Simone Rossi, and Domenico Prattichizzo. A soft supernumerary robotic finger and mobile arm support for grasping compensation and hemiparetic upper limb rehabilitation. *Robotics and Autonomous Systems*, 93:1–12, July 2017.

[85] Irfan Hussain, Giovanni Spagnoletti, Gionata Salvietti, and Domenico Prattichizzo. Toward wearable supernumerary robotic fingers to compensate missing grasping abilities in hemiparetic upper limb:. *The International Journal of Robotics Research*, 36(13-14):1414–1436, July 2017.

[86] G Illahi, M Siekkinen, and Enrico Masala. Foveated video streaming for cloud gaming. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)*, Luton, 2017.

[87] T Imaida, Y Yokokohji, T Doi, M Oda, and T Yoshikawa. Ground–Space Bilateral Teleoperation of ETS-VII Robot Arm by Direct Bilateral Coupling Under 7-s Time Delay Condition. *IEEE Transactions on Robotics and Automation*, 20(3):499–511, June 2004.

[88] Sohail Inayatullah. Teaching and learning in disruptive futures: Automation, universal basic income, and our jobless futures. *Knowledge Futures: Interdisciplinary Journal of Futures Studies*, 1(1):00–00, 2017.

[89] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*, volume 58 of *Book review*. Springer, New York, December 2015.

[90] Yun Jiang, Hema S Koppula, and Ashutosh Saxena. Modeling 3D Environments through Hidden Human Context. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(10):2040–2053, 2016.

[91] E Jovanov, A O'Donnell Lords, D Raskovic, P G Cox, R Adhami, and F Andrasik. Stress monitoring using a distributed wireless intelligent sensor system. *IEEE Engineering in Medicine and Biology Magazine*, 22(3):49–55, May 2003.

[92] Gavin Kane, Georg Eggers, Robert Boesecke, Jörg Raczkowsky, Heinz Wörn, Rüdiger Marmulla, and Joachim Mühling. System Design of a Hand-Held Mobile Robot for Craniotomy. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, pages 402–409. Springer, Berlin, Heidelberg, September 2009.

[93] Jari Kangas, Antti Sand, Tero Jokela, Petri Piippo, Peter Eskolin, Marja Salmimaa, and Roope Raisamo. Remote Expert for Assistance in a Physical Operational Task. In *Extended Abstracts of the 2018 CHI Conference*, New York, New York, USA, 2018. ACM Press.

[94] Shunichi Kasahara, Shohei Nagai, and Jun Rekimoto. JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera. *IEEE Transactions on Visualization and Computer Graphics*, 23(3):1222–1234, January 2017.

[95] Parham M Kebria, Hamid Abdi, Mohsen Moradi Dalvand, Abbas Khosravi, and Saeid Nahavandi. Control Methods for Internet-Based Teleoperation Systems: A Review. *IEEE Transactions on Human-Machine Systems*, 49(1):32–46, November 2018.

[96] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investigation*, 15(3):235–245, February 2018.

[97] Seungwon Kim, Mark Billinghurst, and Gun Lee. The Effect of Collaboration Styles and View Independence on Video-Mediated Remote Collaboration. *Computer Supported Cooperative Work (CSCW)*, 27(3):569–607, December 2018.

[98] Hendrickson Kirk and Ailawadi Kusum L. Six Lessons for In-Store Marketing from Six Years of Mobile Eye-Tracking Research. In *Review of Marketing Research*, pages 57–74. Emerald Group Publishing Limited, January 2014.

[99] Ron Kohavi. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *International Joint Conference on Artificial Intelligence IJCAI*, pages 1137–1145, 1995.

[100] Fatemeh Koochaki and Laleh Najafizadeh. Predicting Intention Through Eye Gaze Patterns. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2018.

[101] Fatemeh Koochaki and Laleh Najafizadeh. Eye Gaze-based Early Intent Prediction Utilizing CNN-LSTM. In *st Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC*, pages 1310–1313. IEEE, July 2019.

[102] I J Kopin, G Eisenhofer, and D Goldstein. Sympathoadrenal medullary system and stress. *Advances in Experimental Medicine and Biology*, 245:11–23, 1988.

[103] Hema S Koppula and Ashutosh Saxena. Anticipating Human Activities Using Object Affordances for Reactive Robotic Response. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):14–29, 2016.

[104] Hema Swetha Koppula, Rudhir Gupta, and Ashutosh Saxena. Learning human activities and object affordances from RGB-D videos. *The International Journal of Robotics Research*, 32(8):951–970, July 2013.

[105] David Kortenkamp, R Peter Bonasso, Dan Ryan, and Debbie Schreckenghost. Traded Control with Autonomous Robots as Mixed Initiative Interaction. In *AAAI Symposium on Mixed Initiative Interaction*, pages 89–94, 1997.

[106] Mareike Kritzler, Magdalena Murr, and Florian Michahelles. RemoteBob - Support of On-site Workers via a Telepresence Remote Expert System. In *the 6th international conference*, pages 7–14, New York, New York, USA, 2016. ACM Press.

[107] T Kurata, N Sakata, Masakatsu Kourogi, Hideaki Kuzuoka, and Mark Billinghurst. Remote collaboration using a shoulder-worn active camera/laser. In *Eighth International Symposium on Wearable Computers*, pages 62–69, 2004.

[108] H Kuzuoka and J Kosaka. GestureMan PS: Effect of a head and a pointing stick on robot mediated communication. In *Proceedings of HCII*, pages 1416–1420, 2003.

[109] H Kuzuoka, S Oyama, K Yamazaki, and Mamoru Mitsuishi. GestureMan: a mobile robot that embodies a remote instructor's actions. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pages 155–162, 2000.

[110] Stephanie Lackey, Daniel Barber, Lauren Reinerman, Norman I Badler, and Irwin Hudson. Defining Next-Generation Multi-Modal Communication in Human Robot Interaction:. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 55(1):461–464, September 2011.

[111] Michael Land and Benjamin Tatler. *Looking and ActingVision and eye movements in natural behaviour*. Oxford University Press, 2009.

[112] Michael Land, Neil Mennie, and Jennifer Rusted. The Roles of Vision and Eye Movements in the Control of Activities of Daily Living. *Perception*, 28(11):1311–1328, 1999.

[113] Michael F Land and Mary Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25-26):3559–3565, November 2001.

[114] Carlos M Lara-Barrios, Andrés Blanco-Ortega, Cesar H Guzmán-Valdivia, and Karla D Bustamante Valles. Literature review and current trends on transfemoral powered prosthetics. *Advanced Robotics*, 32(2):51–62, November 2017.

[115] Win Tun Latt, U-Xuan Tan, Cheng Yap Shee, Cameron N Riviere, and Wei Tech Ang. Compact Sensing Design of a Handheld Active Tremor Compensation Instrument. *IEEE Sensors Journal*, 9(12):1864–1871, December 2009.

[116] Hao Lee, Peter Walker Ferguson, and Jacob Rosen. Lower Limb Exoskeleton Systems—Overview. In *Wearable Robotics*, pages 207–229. Academic Press, January 2020.

[117] Teesid Leelasawassuk, Dima Damen, and Walterio W Mayol-Cuevas. Estimating Visual Attention from a Head Mounted IMU. In *Proceedings of the ACM International Symposium on Wearable Computers*, pages 147–150, New York, New York, USA, September 2015. ACM.

[118] C Lenz, A Sotzek, T Roder, H Radrich, A Knoll, M Huber, and S Glasauer. Human workflow analysis using 3D occupancy grid hand tracking in a human-robot collaboration scenario. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011)*, pages 3375–3380. IEEE, 2011.

[119] Christopher J Lepre, Hélène Roche, Dennis V Kent, Sonia Harmand, Rhonda L Quinn, Jean-Philippe Brugal, Pierre-Jean Texier, Arnaud Lenoble, and Craig S Feibel. An earlier origin for the Acheulian. *Nature*, 477(7362):82–85, August 2011.

[120] Na Li, Lei Yan, Hua Qian, Hongfang Wu, Jian Wu, and Sen Men. Review on Lower Extremity Exoskeleton Robot. *The Open Automation and Control Systems Journal*, 7(1), May 2015.

[121] C Linn, S Bender, J Prosser, Kevin Schmitt, and Dirk Werth. Virtual remote inspection—A new concept for virtual reality enhanced real-time maintenance. In *2017 23rd International Conference on Virtual System & Multimedia (VSMM)*, 2017.

[122] Tingting Liu, Jiaole Wang, and Max Q H Meng. Human robot cooperation based on human intention inference. In *2014 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 350–355. IEEE, 2014.

[123] Tingting Liu, Jiaole Wang, and Max Q H Meng. Evolving hidden Markov model based human intention learning and inference. In *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 206–211. IEEE, 2015.

[124] Yen-Chen Liu and Nikhil Chopra. Semi-autonomous teleoperation in task space with redundant slave robot under communication delays. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 679–684. IEEE, August 2011.

[125] Yen-Chen Liu and Nikhil Chopra. Control of semi-autonomous teleoperation system with time delays. *Automatica*, 49(6):1553–1565, June 2013.

[126] Yugang Liu and Goldie Nejat. Robotic Urban Search and Rescue: A Survey from the Control Perspective. *Journal of Intelligent & Robotic Systems*, 72(2):147–165, 2013.

[127] Baldin Llorens-Bonilla and H Harry Asada. A robot on the shoulder: Coordinated human-wearable robot control using Coloured Petri Nets and Partial Least Squares predictions. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 119–125. IEEE, 2014.

[128] Jess H Lonner, Julie R Smith, Frederic Picard, Brian Hamlin, Philip J Rowe, and Philip E Riches. High Degree of Accuracy of a Novel Image-free Handheld Robot for Unicondylar Knee Arthroplasty in a Cadaveric Study. *Clinical Orthopaedics and Related Research®*, 473(1):206–212, 2015.

[129] Dylan P Losey, Craig G McDonald, Edoardo Battaglia, and Marcia K O'Malley. A Review of Intent Detection, Arbitration, and Communication Aspects of Shared Control for Physical Human–Robot Interaction. *Applied Mechanics Reviews*, 70(1):203–19, February 2018.

[130] R A MacLachlan, R L Hollis, J N Martel, L A Lobes Jr., and C N Riviere. Toward Improved Electromagnetic Tracking for Handheld Robotics. In *Proceedings of the 3rd International Conference on Mechatronics and Robotics Engineering*, pages 75–80, New York, New York, USA, 2017. ACM Press.

[131] Robert A MacLachlan, Brian C Becker, Jaime Cuevas Tabarés, Gregg W Podnar, Louis A Lobes, and Cameron N Riviere. Micron: an Actively Stabilized Handheld Tool for Microsurgery. *IEEE Transactions on Robotics*, 28(1): 195–212, February 2012.

[132] Robert A MacLachlan, Nicholas Parody, Shohin Mukherjee, Ralph L Hollis, Cameron N Riviere, Joseph N Martel, and Louis A Lobes. Electromagnetic tracker for active handheld robotic systems. In *2016 IEEE SENSORS*. IEEE, 2016.

[133] Jim Mainprice, Rafi Hayne, and Dmitry Berenson. Predicting human reaching motion in collaborative tasks using Inverse Optimal Control and iterative re-planning. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 885–892. IEEE, 2015.

[134] A Malliani, F Lombardi, and M Pagani. Power spectrum analysis of heart rate variability: a tool to explore neural regulatory mechanisms. *Heart*, 71(1):1–2, January 1994.

[135] Monica Malvezzi, Zubair Iqbal, Maria Cristina Valigi, Maria Pozzi, Domenico Prattichizzo, and Gionata Salvietti. Design of Multiple Wearable Robotic Extra Fingers for Human Hand Augmentation. *Robotics*, 8(4):102, December 2019.

[136] Regan L Mandryk, Kori M Inkpen, and Thomas W Calvert. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & Information Technology*, 25(2):141–158, March 2011.

[137] Mohsen Mansouryar, Julian Steil, Yusuke Sugano, and Andreas Bulling. 3D Gaze Estimation from 2D Pupil Positions on Monocular Head-Mounted Eye Trackers . In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research Applications*, pages 197–200. ACM, March 2016.

[138] Lauren E Marsh, Patricia Kanngiesser, and Bruce Hood. When and how does labour lead to love? The ontogeny and mechanisms of the IKEA effect. *Cognition*, 170:245–253, January 2018.

[139] Naresh Marturi, Alireza Rastegarpanah, Chie Takahashi, Maxime Adjigble, Rustam Stolkin, Sebastian Zurek, Marek Kopicki, Mohammed Talha, Jeffrey A Kuo, and Yasemin Bekiroglu. Towards advanced robotic manipulation for nuclear decommissioning: A pilot study on tele-operation and autonomy. In *2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA)*. IEEE, 2016.

[140] Riccardo Masoni, Francesco Ferrise, Monica Bordegoni, Michele Gattullo, Antonio E Uva, Michele Fiorentino, Ernesto Carrabba, and Michele Di Donato. Supporting Remote Maintenance in Industry 4.0 through Augmented Reality. *Procedia Manufacturing*, 11:1296–1302, 2017.

[141] Akira Matsuda and Jun Rekimoto. ScalableBody : A Telepresence Robot Supporting Socially Acceptable Interactions and Human Augmentation through Vertical Actuation. In *the 29th Annual Symposium*, pages 103–105, New York, New York, USA, 2016. ACM Press.

[142] Walterio W Mayol, Andrew J Davison, Ben J Tordoff, and David W Murray. Applying Active Vision and SLAM to Wearables. In *Robotics Research. The Eleventh International Symposium*, pages 325–334. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[143] J C McCall, D P Wipf, M M Trivedi, and B D Rao. Lane Change Intent Analysis Using Robust Operators and Sparse Bayesian Learning. *IEEE Transactions on Intelligent Transportation Systems*, 8(3):431–440, 2007.

[144] David P McMullen, Guy Hotson, Kapil D Katyal, Brock A Wester, Matthew S Fifer, Timothy G McGee, Andrew Harris, Matthew S Johannes, R Jacob Vogelstein, Alan D Ravitz, William S Anderson, Nitish V Thakor, and Nathan E Crone. Demonstration of a Semi-Autonomous Hybrid Brain-Machine Interface Using Human Intracranial EEG, Eye Tracking, and Computer Vision to Control a Robotic Upper Limb Prosthetic. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(4):784–796, July 2014.

[145] Neil Mennie, Mary Hayhoe, and Brian Sullivan. Look-ahead fixations: anticipatory eye movements in natural tasks. *Experimental Brain Research*, 179(3):427–442, December 2006.

[146] Leslie Mertz. Taking on Essential Tremor: New Tools and Approaches Offer patients Increased Treatment Options. *IEEE Pulse*, 7(3):20–25, 2016.

[147] Joseph C Miller, Joelle C Ruthig, April R Bradley, Richard A Wise, Heather A Pedersen, and Jo M Ellison. Learning effects in the block design task: A stimulus parameter-based approach. *Psychological Assessment*, 21(4):570–577, 2009.

[148] Svjetlana Miocinovic, Ali Shoeb, Sarah Wang, Nicole Swann, Anupam Pathak, and Jill Ostrem. Tremor Severity Estimation Using Liftware Instrumented Eating Utensil (I8.003). *Neurology*, 86(16 Supplement), April 2016.

[149] Giovanni Mirabella and Mikhail A Lebedev. Interfacing to the brain's motor decisions. *Journal of Neurophysiology*, 117(3):1305–1319, March 2017.

[150] H Moravec. Mind children. Harvard University Press, Cambridge, MA, 1988.

[151] D Mourtzis, V Zogopoulos, and E Vlachou. Augmented Reality Application to Support Remote Maintenance as a Service in the Robotics Industry. *Procedia CIRP*, 63:46–51, 2017.

[152] Fiona Mulvey and Michael Heubner. Eye Movements and Attention. In *Gaze Interaction and Applications of Eye Tracking*, pages 129–152. IGI Global, 2012.

[153] Riccardo Muradore and Paolo Fiorini. A review of bilateral teleoperation algorithms. *Acta Polytechnica Hungarica*, 13(1):191–208, 2016.

[154] Takashi Nagamatsu, Ryuichi Sugano, Yukina Iwamoto, Junzo Kamahara, and Naoki Tanaka. *User-calibration-free gaze tracking with estimation of the horizontal angles between the visual and the optical axes of both eyes*. ACM, March 2010.

[155] K Naruse, S Kawai, H Yokoi, and Y Kakazu. Development of wearable exoskeleton power assist system for lower back support. In *2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3630–3635. IEEE, 2003.

[156] C W Nielsen, D J Bruemmer, and D A Few. Framing and evaluating human-robot interactions. In *Proceedings of the Workshop on Metrics for Human-Robot Interaction*, pages 29–36, 2008.

[157] Michael I Norton, Daniel Mochon, and Dan Ariely. The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, 22(3):453–460, July 2012.

[158] I R Nourbakhsh, K Sycara, M Koes, M Yong, M Lewis, and S Burion. Human-robot teaming for search and rescue. *IEEE Pervasive Computing*, 4(1):72–79, February 2005.

[159] S S Nudehi, R Mukherjee, and M Ghodoussi. A shared-control approach to haptic interface design for minimally invasive telesurgical training. *IEEE Transactions on Control Systems Technology*, 13(4):588–592, June 2005.

[160] Johanna Renny Octavia, Karin Coninx, and Kris Luyten. Squeeze me and i'll change: An exploration of frustration-triggered adaptation for multimodal interaction. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 79–86. IEEE, 2011.

[161] Jean-Marc Odobez and Sileye Ba. A Cognitive and Unsupervised Map Adaptation Approach to the Recognition of the Focus of Attention from Head Pose. In *2007 IEEE International Conference on Multimedia and Expo*, pages 1379–1382. IEEE, 2007.

[162] Jiazhi Ou, Susan R Fussell, Xilin Chen, Leslie D Setlock, and Jie Yang. Gestural Communication over Video Stream: Supporting Multimodal Interaction for Remote Collaborative Physical Tasks. In *ICMI '03: Proceedings of the 5th International Conference on Multimodal Interfaces*, pages 242–249, 2003.

[163] Parker Owan, Joseph Garbini, and Santosh Devasia. Uncertainty-based Arbitration of Human-Machine Shared Control. *arXiv preprint*, 1511.05996, November 2015.

[164] Alexandra Papoutsaki, Patsorn Sangkloy, James Laskey, Nediyana Daskalova, Jeff Huang, and James Hays. WebGazer: Scalable Webcam Eye Tracking Using User Interactions. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence - IJCAI 2016*, January 2016.

[165] Federico Parietti and H Harry Asada. Supernumerary Robotic Limbs for aircraft fuselage assembly: Body stabilization and guidance by bracing. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1176–1183. IEEE, 2014.

[166] Federico Parietti and Harry H Asada. Dynamic analysis and state estimation for wearable robotic limbs subject to human-induced disturbances. In *2013 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3880–3887. IEEE, May 2013.

[167] Federico Parietti, Kameron Chan, and H Harry Asada. Bracing the human body with supernumerary Robotic Limbs for physical assistance and load reduction. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 141–148. IEEE, 2014.

[168] Federico Parietti, Kameron C Chan, Banks Hunter, and H Harry Asada. Design and control of Supernumerary Robotic Limbs for balance augmentation. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5010–5017. IEEE, 2015.

[169] Jeff Pelz, Mary Hayhoe, and Russ Loeber. The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, 139(3):266–277, 2001.

[170] Jeff B Pelz and Roxanne Canosa. Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, 41(25-26):3587–3596, 2001.

[171] Q C Pham, R Madhavan, L Righetti, W Smart, and R Chatila. The Impact of Robotics and Automation on Working Conditions and Employment [Ethical, Legal, and Societal Issues]. *IEEE Robotics & Automation Magazine*, 25(2): 126–128, June 2018.

[172] Peter Paul Pott, Achim Wagner, Andreas Köpfle, Essameddin Badreddin, Reinhard Männer, Hans-Peter Weiser, Hanns-Peter Scharf, and Markus Schwarz. A handheld surgical manipulator: ITD - Design and first results. In *CARS - computer assisted radiology and surgery proceedings of the th International Congress and Exhibition*, Chicago, USA, 2004.

[173] Domenico Prattichizzo, Monica Malvezzi, Irfan Hussain, and Gionata Salvietti. The Sixth-Finger: A modular extra-finger to enhance human hand capabilities. In *2014 RO-MAN: The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 993–998. IEEE, August 2014.

[174] Domenico Prattichizzo, Gionata Salvietti, Francesco Chinello, and Monica Malvezzi. An object-based mapping algorithm to control wearable robotic extra-fingers. In *2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 1563–1568. IEEE, July 2014.

[175] Domenico Prattichizzo, Monica Malvezzi, and Gionata Salvietti. Supernumerary Robotic Fingers to Compensate and Augment Human Manipulation Abilities. In *Inclusive Robotics for a Better Society*, pages 188–194. Springer, Cham, October 2018.

[176] Romain Prévost, Alec Jacobson, Wojciech Jarosz, and Olga Sorkine-Hornung. Large-scale painting of photographs by interactive optimization. *Computers & Graphics*, 55:108–117, April 2016.

[177] Abhishek Ranjan, Jeremy P Birnholtz, and Ravin Balakrishnan. Dynamic shared visual spaces: Experimenting with automatic camera control in a remote repair task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1177–1186, San Jose, California, USA, 2007. ACM Press.

[178] H C Ravichandar and A Dani. Intention Inference for Human-Robot Collaboration in Assistive Robotics. In *Human Modelling for Bio-Inspired Robotics*, pages 217–249. Elsevier, 2017.

[179] Harish chaandar Ravichandar and Ashwin Dani. Human intention inference through interacting multiple model filtering. In *2015 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pages 220–225, San Diego, CA, USA, 2015.

[180] Harish chaandar Ravichandar, Avnish Kumar, and Ashwin Dani. Bayesian Human Intention Inference Through Multiple Model Filtering with Gaze-based Priors. In *th International Conference on Information Fusion FUSION*, pages 2296–2302, June 2016.

[181] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788, 2016.

[182] Ryan Rifkin and Aldebaro Klautau. In Defense of One-Vs-All Classification. *Journal of Machine Learning Research*, pages 101–141, June 2004.

[183] Alec Rivers, Ilan E Moyer, and Frédo Durand. Position-Correcting Tools for 2D Digital Fabrication. *ACM Transactions on Graphics (TOG)*, 31(4):88–7, August 2012.

[184] C N Riviere, R S Rader, and P K Khosla. Characteristics of hand motion of eye surgeons. In *19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering'*, pages 1690–1693. IEEE, 1997.

[185] C N Riviere, Wei Tech Ang, and P K Khosla. Toward active tremor canceling in handheld microsurgical instruments. *IEEE Transactions on Robotics and Automation*, 19(5):793–800, October 2003.

[186] Kostia Robert, Dingyun Zhu, Weidong Huang, and Leila Alem. MobileHelper: Remote Guiding Using Smart Mobile Devices, Hand Gestures and Augmented Reality. In *SIGGRAPH Asia Symposium on Mobile Graphics and Interactive Applications. ACM*, August 2013.

[187] Michele Rucci, Paul V McGraw, and Richard J Krauzlis. Fixational eye movements and perception. *Vision Research*, 118(C):1–4, January 2016.

[188] John Rushby. Using model checking to help discover mode confusions and other automation surprises. *Reliability Engineering & System Safety*, 75(2):167–177, February 2002.

[189] Joyce Sabari, Dimitre G Stefanov, Judy Chan, Lorraine Goed, and Joyce Starr. Adapted Feeding Utensils for People With Parkinson's-Related or Essential Tremor. *The American Journal of Occupational Therapy : Official Publication of the American Occupational Therapy Association*, 73(2):7302205120, March 2019.

[190] K Sainani. The Importance of Accounting for Correlated Observations. *PM&R*, 2(9):858–861, September 2010.

[191] Timothy A Salthouse and Cecil L Ellis. Determinants of Eye-Fixation Duration. *The American Journal of Psychology*, 93(2):207, June 1980.

[192] Mine Sarac, Ela Koyas, Ahmetcan Erdogan, Mujdat Cetin, and Volkan Patoglu. Brain Computer Interface based robotic rehabilitation with online modification of task speed. In *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR 2013)*. IEEE, May 2013.

[193] MHD Yamen Saraiji, Tomoya Sasaki, Reo Matsumura, Kouta Minamizawa, and Masahiko Inami. Fusion: Full Body Surrogacy for Collaborative Communication . In *SIGGRAPH '18: Special Interest Group on Computer Graphics and Interactive Techniques Conference*, New York, NY, USA, August 2018. ACM.

[194] Jocelyn Scheirer, Raul Fernandez, Jonathan Klein, and Rosalind W Picard. Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers*, 14(2):93–118, February 2002.

[195] André Schiele and Gianfranco Visentin. The ESA human arm exoskeleton for space robotics telepresence. In *proceedings of 7th international Symposium on Artificial Intelligence, Robotics and Automation in Space, iSAIRAS*, pages 19–23, Nara, Japan, 2003.

[196] Randall S Schuler. Definition and conceptualization of stress in organizations. *Organizational Behavior and Human Performance*, 25(2):184–215, April 1980.

[197] M L Schwarz, A Wagner, A El-Shenawy, R Gundling, A Köpfle, H Handel, E Badreddin, R Männer, H P Scharf, M Götz, M Schill, and P P Pott. A Handheld Robot for Orthopedic Surgery - ITD. In *World Congress on Medical Physics and Biomedical Engineering: September 7-12, 2009, Munich, Germany*, pages 99–102. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[198] Max Schwarz, Tobias Rodehutskors, David Droeschel, Marius Beul, Michael Schreiber, Nikita Araslanov, Ivan Ivanov, Christian Lenz, Jan Razlaw, Sebastian Schüller, David Schwarz, Angeliki Topalidou-Kyniazopoulou, and Sven Behnke. NimbRo Rescue: Solving Disaster-response Tasks with the Mobile Manipulation Robot Momaro. *Journal of Field Robotics*, 34(2):400–425, November 2016.

[199] Alan Selby. Undercover at Amazon: Exhausted humans are inefficient so robots are taking over. *Mirror Online*, 2017.

[200] Fred Shaffer and J P Ginsberg. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, 5(258):1–17, 2017.

[201] Yang Shen, Peter Walker Ferguson, and Jacob Rosen. Upper Limb Exoskeleton Systems—Overview. In *Wearable Robotics*, pages 1–22. Academic Press, January 2020.

[202] Roy Shilkrot, Pattie Maes, Joseph A Paradiso, and Amit Zoran. Augmented Airbrush for Computer Aided Painting (CAP). *ACM Transactions on Graphics (TOG)*, March 2015.

[203] Daniel J Simons and Christopher F Chabris. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, 28(9):1059–1074, 1999.

[204] Robert Skilton, Nathan Hamilton, Robert Howell, Chris Lamb, and Jorge Rodriguez. MASCOT 6: Achieving high dexterity tele-manipulation with a modern architectural design for fusion remote maintenance. *Fusion Engineering and Design*, 136:575–578, November 2018.

[205] João Soares, Alberto Vale, and Rodrigo Ventura. A Multi-purpose Rescue Vehicle and a human–robot interface architecture for remote assistance in ITER. *Fusion Engineering and Design*, 98-99:1656–1659, October 2015.

[206] R S Sodhi, B R Jones, D Forsyth, Brian P Bailey, and Guiliano Maciocci. BeThere: 3D mobile collaboration with spatial input. In *CHI Changing Perspectives*, pages 179–188, Paris, France, 2013.

[207] R William Soukoreff and I Scott MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, 61(6):751–789, December 2004.

[208] R Stiefelhagen, C Fogen, P Gieselmann, H Holzapfel, K Nickel, and A Waibel. Natural Human-Robot Interaction using Speech, Head Pose and Gestures . In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, pages 2422–2427. IEEE, 2004.

[209] Janis Stolzenwald and Walterio Mayol-Cuevas. I Can See Your Aim: Estimating User Attention From Gaze For Handheld Robot Collaboration. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3897–3904, October 2018.

[210] Janis Stolzenwald and Walterio W Mayol-Cuevas. Rebellion and Obedience: The Effects of Intention Prediction in Cooperative Handheld Robots. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3012–3019, Macau, China, November 2019.

[211] Janis Stolzenwald and Walterio W Mayol-Cuevas. Reach Out and Help: Assisted Remote Collaboration through a Handheld Robot . *arXiv preprint*, pages 1–8, 2020.

[212] Jonathan Sykes and Simon Brown. *Affective gaming: measuring emotion through the gamepad*. measuring emotion through the gamepad. ACM, New York, New York, USA, April 2003.

[213] Matthew Tait and Mark Billinghurst. The Effect of View Independence in a Collaborative AR System. *Computer Supported Cooperative Work (CSCW)*, 24(6):563–589, August 2015.

[214] M Talha, E A M Ghalamzan, C Takahashi, J Kuo, W Ingamells, and R Stolkin. Towards robotic decommissioning of legacy nuclear plant: Results of human-factors experiments with tele-robotic manipulation, and a discussion of challenges and approaches for decommissioning. *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 166–173, October 2016.

[215] A Tang, O Fakourfar, C Neustaedter, and S Bateman. Collaboration in 360 videochat: Challenges and opportunities. In *Proceedings of the Conference on Designing Interactive Systems DIS*, pages 1327–1339, 2017.

[216] Aksel A Transeth, Oystein Skotheim, Henrik Schumann-Olsen, Gorm Johansen, Jens Thielemann, and Erik Kyrkjebo. A robotic concept for remote maintenance operations: A robust 3D object detection and pose estimation method and a novel robot tool. *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5099–5106, July 2010.

[217] Pete Trautman. Generalized Shared Control versus Classical Shared Control: Illustrative Examples. *arXiv preprint*, November 2016.

[218] Jochen Triesch, Dana H Ballard, Mary M Hayhoe, and Brian T Sullivan. What you see is what you need. *Journal of Vision*, 3(1):86–94, 2003.

[219] Jocelyne Troccaz and Yves Delnondedieu. Semi-active guiding systems in surgery. A two-dof prototype of the passive arm with dynamic constraints (PADyC). *Mechatronics*, 6(4):399–421, June 1996.

[220] Panagiota Tsarouchi, Sotiris Makris, and George Chryssolouris. Human–robot interaction review and challenges on task planning and programming. *International Journal of Computer Integrated Manufacturing*, 29(8):916–931, February 2016.

[221] Vighnesh Vatsal and Guy Hoffman. Wearing Your Arm on Your Sleeve: Studying Usage Contexts for a Wearable Robotic Forearm. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 974–980. IEEE, 2017.

[222] Vighnesh Vatsal and Guy Hoffman. At Arm's Length: Challenges in Building a Wearable Robotic Forearm for Human-Robot Collaboration . In *Companion of the 2018 ACM/IEEE International Conference*, pages 271–272, New York, New York, USA, 2018. ACM Press.

[223] Vighnesh Vatsal and Guy Hoffman. A Wearable Robotic Forearm for Human-Robot Collaboration. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, pages 329–330, New York, New York, USA, 2018. ACM Press.

[224] Vighnesh Vatsal and Guy Hoffman. Design and Analysis of a Wearable Robotic Forearm. In *IEEE International Conference on Robotics and Automation ICRA*, pages 5489–5496, April 2018.

[225] Vighnesh Vatsal and Guy Hoffman. End-Effector Stabilization of a Wearable Robotic Arm Using Time Series Modeling of Human Disturbances. In *Proceedings of the ASME Dynamic Systems and Control Conference DSCC*, pages 1–10, Park City, Utah, USA, June 2019.

[226] Catherine Veronneau, Jeff Denis, Louis-Philippe Lebel, Marc Denninger, Vincent Blanchard, Alexandre Girard, and Jean-Sebastien Plante. Multifunctional Remotely Actuated 3-DOF Supernumerary Robotic Arm Based on Magnetorheological Clutches and Hydrostatic Transmission Lines. *IEEE Robotics and Automation Letters*, 5(2): 2546–2553, February 2020.

[227] Adriana N Vest, Giulia Da Poian, Qiao Li, Chengyu Liu, Shamim Nemati, Amit J Shah, and Gari D Clifford. An open source benchmarked toolbox for cardiovascular waveform and interval analysis. *Physiological Measurement*, 39 (10):105004, October 2018.

[228] Ksenia Volkova, Mikhail A Lebedev, Alexander Kaplan, and Alexei Ossadtchi. Decoding Movement From Electrocorticographic Activity: A Review. *Frontiers in Neuroinformatics*, 13:74, 2019.

[229] Wagner, A, Pott, P P, Schwarz, MLR, Scharf, H P, and Weiser, P. Control of a Handheld Robot for Orthopaedic Surgery. *5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles*, 37(14):477–482, September 2004.

[230] Haofei Wang and Bertram E Shi. Gaze awareness improves collaboration efficiency in a collaborative assembly task. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, New York, New York, USA, 2019. ACM Press.

[231] Junfeng Wang, Yaqing Feng, Cheng Zeng, and Shiqi Li. An augmented reality based system for remote collaborative maintenance instruction of complex products. In *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 309–314, Taipei, July 2014. IEEE.

[232] Armin Wedler, Martina Wildee, Josef Reilla, Martin J Schustera, Mallikarjuna Vayugundlaa, Sebastian G Brunnera, Kristin Bussmanna, Andreas Dömela, Martin Drauschkea, Heinrich Gmeinera, Hannah Lehnera, Peter Lehnera, Marcus G Müllera, Wolfgang Stürzla, Rudolph Triebela, Bernhard Vodermayera, Anko Börnerb, Rainer Krennc, Armin Dammannd, Uwe-Carsten Fiebigd, Emanuel Staudingerd, Frank Wenzhöfere, Sascha Flögelf, Stefan Sommerf, Tamim Asfourg, Michael Fladh, Sören Hohmannh, Martin Brandaueri, and Alin Olimpiu Albu-Schäffera. From single autonomous robots toward cooperative robotic interactions for future planetary exploration missions . In *Proceedings of the 69th International Astronautical Congress IAC*, Bremen, September 2018.

[233] Timothy G Wilson. Advancement of Technology and Its Impact on Urologists: Release of the da Vinci Xi, A New Surgical Robot. *European Urology*, 66(5):793–794, November 2014.

[234] Erin Winick. Every study we could find on what automation will do to jobs, in one chart. *MIT Technology Review*, February 2018.

[235] Wilkey Wong, Mike Bartels, and Nina Chrobot. Practical Eye Tracking of the Ecommerce Website User Experience. In *Universal Access in Human-Computer Interaction. Design for All and Accessibility Practice*, pages 109–118. Springer, Cham, Cham, June 2014.

[236] Faye Wu and Harry Asada. Bio-Artificial Synergies for Grasp Posture Control of Supernumerary Robotic Fingers. In *Robotics: Science and Systems 2014*. Robotics: Science and Systems Foundation, July 2014.

[237] Faye Y Wu and H Harry Asada. "Hold-and-Manipulate" with a Single Hand Being Assisted by Wearable Extra

Fingers . In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6205–6212. IEEE, May 2015.

[238] X Xue, J Tang, N Sammes, and Y Du. Dynamic modeling of single tubular SOFC combining heat/mass transfer and electrochemical reaction effects. *Journal of Power Sources*, 142(1-2):211–222, March 2005.

[239] Takuya Yamamoto, Mai Otsuki, Hideaki Kuzuoka, and Yusuke Suzuki. Tele-Guidance System to Support Anticipation during Communication. *Multimodal Technologies and Interaction*, 2(3):55, September 2018.

[240] Mayu M Yamashita, Junichi Yamaoka, and Yasuaki Kakehi. enchanted scissors: A Scissor Interface for Support in Cutting and Interactive Fabrication. In *ACM SIGGRAPH 2013 Posters*, New York, New York, USA, 2013. ACM Press.

[241] Sungwook Yang, Robert A MacLachlan, Joseph N Martel, Louis A Lobes, and Cameron N Riviere. Comparative Evaluation of Handheld Robot-Aided Intraocular Laser Surgery. *IEEE Transactions on Robotics*, 32(1):246–251, 2016.

[242] A W W Yew, S K Ong, and A Y C Nee. Immersive Augmented Reality Environment for the Teleoperation of Maintenance Robots. *Procedia CIRP*, 61:305–310, 2017.

[243] Weilie Yi and Dana Ballard. Recognizing Behavior in Hand-Eye Coordination Patterns. *International Journal of Humanoid Robotics*, 06(03):337–359, November 2011.

[244] Shelten G Yuen, Michael C Yip, Nikolay V Vasilyev, Douglas P Perrin, Pedro J del Nido, and Robert D Howe. Robotic Force Stabilization for Beating Heart Intracardiac Surgery. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, pages 26–33. Springer, Berlin, Heidelberg, Berlin, Heidelberg, September 2009.

[245] Shelten G Yuen, Douglas P Perrin, Nikolay V Vasilyev, Pedro J del Nido, and Robert D Howe. Force tracking with feed-forward motion estimation for beating heart surgery. *IEEE Transactions on Robotics*, 26(5):888–896, 2010.

[246] Pengqing Zhang. Automation, wage inequality and implications of a robot tax. *International Review of Economics and Finance*, 59:500–509, January 2019.

[247] Qiang Zhang, Ling Zhou, and Zengfu Wang. Design and implementation of wormlike creeping mobile robot for EAST remote maintenance system. *Fusion Engineering and Design*, 118:81–97, May 2017.

[248] Tianci Zhang, Lun Gong, Shuxin Wang, and Siyang Zuo. Hand-Held Instrument with Integrated Parallel Mechanism for Active Tremor Compensation During Microsurgery. *Annals of Biomedical Engineering*, 48(1):413–425, January 2020.

[249] Amit Zoran and Joseph A Paradiso. FreeD - A Freehand Digital Sculpting Tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2613–2616, New York, New York, USA, 2013. ACM Press.

[250] Amit Zoran, Roy Shilkrot, and Joseph Paradiso. Human-computer interaction for hybrid carving. In *the 26th Annual ACM Symposium*, pages 433–440, New York, New York, USA, 2013. ACM Press.

[251] Amit Zoran, Roy Shilkrot, Suranga Nanyakkara, and Joseph Paradiso. The Hybrid Artisans: A Case Study in Smart Tools . *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(3):1–29, June 2014.

[252] Andrea Zunino, Jacopo Cavazza, Atesh Koul, Andrea Cavallo, Cristina Becchio, and Vittorio Murino. What Will I Do Next? The Intention from Motion Experiment. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–8. IEEE, July 2017.

[253] Andrea Zunino, Jacopo Cavazza, Atesh Koul, Andrea Cavallo, Cristina Becchio, and Vittorio Murino. Predicting Human Intentions from Motion Cues Only. In *the 2017 ACM*, pages 591–599, New York, New York, USA, 2017. ACM Press.

Appendix A

Publication Abstracts

Janis Stolzenwald and Walterio Mayol-Cuevas. I Can See Your Aim: Estimating User Attention From Gaze For Handheld Robot Collaboration. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3897–3904, October 2018.

Abstract — This paper explores the estimation of user attention in the setting of a cooperative handheld robot a robot designed to behave as a handheld tool but that has levels of task knowledge. We use a tool-mounted gaze tracking system, which, after modelling via a pilot study, we use as a proxy for estimating the attention of the user. This information is then used for cooperation with users in a task of selecting and engaging with objects on a dynamic screen. Via a video game setup, we test various degrees of robot autonomy from fully autonomous, where the robot knows what it has to do and acts, to no autonomy where the user is in full control of the task. Our results measure performance and subjective metrics and show how the attention model benefits the interaction and preference of users.

Janis Stolzenwald and Walterio W Mayol-Cuevas. Rebellion and Obedience: The Effects of Intention Prediction in Cooperative Handheld Robots. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3012–3019, Macau, China, November 2019.

Abstract — Within this work, we explore intention inference for user actions in the context of a handheld robot setup. Handheld robots share the shape and properties of handheld tools while being able to process task information and aid manipulation. Here, we propose an intention prediction model to enhance cooperative task solving. The model derives intention

from the user's gaze pattern which is captured using a robot-mounted remote eye tracker. The proposed model yields real-time capabilities and reliable accuracy up to 1.5s prior to predicted actions being executed. We assess the model in an assisted pick and place task and show how the robot's intention obedience or rebellion affects the cooperation with the robot.

Janis Stolzenwald and Walterio W Mayol-Cuevas. Reach Out and Help: Assisted Remote Collaboration through a Handheld Robot . *arXiv preprint*, pages 1–8, 2020. (in submission)

Abstract — We explore a remote collaboration setup, which involves three parties: a local worker, a remote helper and a handheld robot carried by the local worker. We propose a system that allows a remote user to assist the local user through diagnosis, guidance and physical interaction as a novel aspect with the handheld robot providing task knowledge and enhanced motion and accuracy capabilities. Through experimental studies, we assess the proposed system in two different configurations: with and without the robots assistance in terms of object interactions and task knowledge. We show that the handheld robot can mediate the helpers instructions and remote object interactions while the robots semi-autonomous features improve task performance by 24%, reduce the workload for the remote user and decrease required communication bandwidth between both users. This study is a first attempt to evaluate how this new type of collaborative robot works in a remote assistance scenario, a setup that we believe is important to leverage current robot constraints and existing communication technologies.

Appendix B

An Attempt to Detect Frustration of Handheld Robot Users

The material in this appendix was generated in the context of the intention validation study presented in Chapter 4 (Section 4.6) as part of our search for a feasible metric to quantify user frustration. It justifies, why we chose subjective metrics over physiologic metrics for our experiments, however, the material is non-essential for the overall argument of Chapter 4. Therefore, we decided to exclude this part from the main body of the thesis, but append it as this could be useful to inform future research routes in this field.

B.1 Introduction

In Chapter 4 we introduced an attention model for handheld robots. The core idea of the validation method was to measure users' frustration levels for two main robot conditions, i.e. rebellion and obedience. This new way of model validation was based on the assumption, that the robot could only inflict frustration in the rebel mode, if the model predicts correctly. The reason is that the robot avoids steps that are predicted to be intended by the user when it is in the rebel mode. Hence, an increased frustration level (compared to the obedience mode) indicates that the predictions were correct. The caveat of this method is that we had to rely on questionnaire results (TLX), i.e. subjective metrics to determine frustration levels. While this was sufficient for our studies, the question arises whether there exists an objective metric that offers a higher precision due to the absence of subjective bias. Furthermore, objective detection would allow for a real-time estimation of stress levels, e.g. as required for the robot to adapt to user's preferences.

Recent research reveals that physiological metrics such as the cardiovascular response are more sensitive and accurate to detect stress-based emotions [136]. These means cannot directly measure the complex phenomenon of frustration but detect resulting stress responses such as arousal [194]. Therefore, we use physiological stress detection as a proxy for determining a person's frustration level. While several methods were explored in recent works [8, 50], it is unclear whether these are suited for the complex constraints and requirements we are facing within a handheld robot setup.

In this research excursion, we look at methods for physiological stress detection and how they compare to questionnaire results. This study is based on user experiments in two conditions, with and without frustration triggers during trials. We compare subjective metrics with cardiovascular data from a handle-integrated sensor. The results informed decisions made in the intention validation study in Chapter 4.

B.2 A Review of Physiologic Stress Detection

There are varying definitions for stress [196] depending on the specific context they are used for. However, most share the common ground that psychological stress can be defined as a reaction which changes an organism from a calm state to an excited state. This form of arousal can be distinguished between *eustress*, such as joy, or *distress* which is often referred to as a *fight or flight* response [102]. As we are interested in arousal that stems from frustration, in this work we will refer to *distress* when *stress* is mentioned.

While to date there is no standardised form of physiologic stress evaluation that would be universally recognised [96], in the past, physiological data such as cardiovascular or GSR, EMG and respiration have been used for an estimate of humans' stress levels. These means were often used to detect the level of cognitive workload for example in a driving scenario [66] or in office environments [7, 70].

In their real-world driving study, Healey and Picard [66] monitored participants' physiological data to determine a driver's relative stress level. Measurements from Electrocardiography (ECG), GSR, EMG and respiration responses were continuously recorded during the trials. They found that GSR and ECG-derived features such as HR and heart rate variability (HRV) correlate with stress levels.

In the aforementioned works on stress detection in office environments, subjects' HRV signals were recorded during in-lab computer exercises which were furthermore followed by a questionnaire as part of the subjective stress assessment [7, 70]. In both studies, each participant completed the same task in two sessions, a stress session and a control session, where the stress session was marked by increased stress stimuli. While Al Osman et al. [7] introduced an increased temporal demand for these sessions, Hjortskov et al.

[70] used experimenters' unfriendly attitude as a stressor. In both studies, HRV data was used to successfully detect which kind of session it stems from. Interestingly, this was not true for the results of the subjective stress results hence why Hjortskov et al. suggest that physiological data might be a more sensitive measure of mental stress.

Another domain for the use of biofeedback, which has increasingly caught researchers' attention over recent years, is the application for video games. That way physiological data such as button pressure and GSR [160, 194] or HRV data [8] helps detecting frustration-based stress which could help controlling it in real time through adapting the game in a closed-loop setup.

As part of their search for an objective stress metrics, Kim et al. [96] identified HRV data as the most promising candidate among a set of physiological metrics. In particular, compared to other physiological metrics, HRV is less affected by positive arousal, therefore, it can be considered as a robust feature for stress detection. Their suggestion is supported by Endukuru and Tripathi [50] who explored a broad range of HRV metrics in their work about human's cardiac response to stress. Within their experimental studies, participants were exposed to the Stroop test while ECG data was recorded. As part of their analysis of the HRV components, they found that the power ratio of Low Frequency over High Frequency (LF/HF) was most sensitive to the individual's stress.

We note, that to date the literature does not suggest an absolute interpretation of HRV data, i.e. there is no universal stress index [96]. However, the aforementioned works demonstrate that HRV, and in particular the LF/HF component, can serve as a tool to identify the *relative* stress levels of compared scenarios.

B.3 Heart Rate Variability and Stress Levels

Human heart activity is non-voluntarily controlled by the brain via the Autonomic Nervous System (ANS) which consists of two main branches, the Sympathetic Nervous System (SNS) and the Parasympathetic Nervous System (PNS) [50]. While sympatic activity has an exiting effect, e.g. in a challenging situation, parasympatic activity has a prohibitory effect, e.g. lowering the HR during sleep. Their constant interaction is reflected in HRV [134], which therefore indicates the balance of the ANS and may thus be considered as a stress indicator.

HRV is defined as the variation in time intervals between heartbeats, commonly referred to as RR-intervals since the signal was historically derived from the R-peaks of an ECG curve (Figure B.1). The components of the HRV signal can be divided into two domains, the time domain and the frequency domain where the frequency domain plays a major

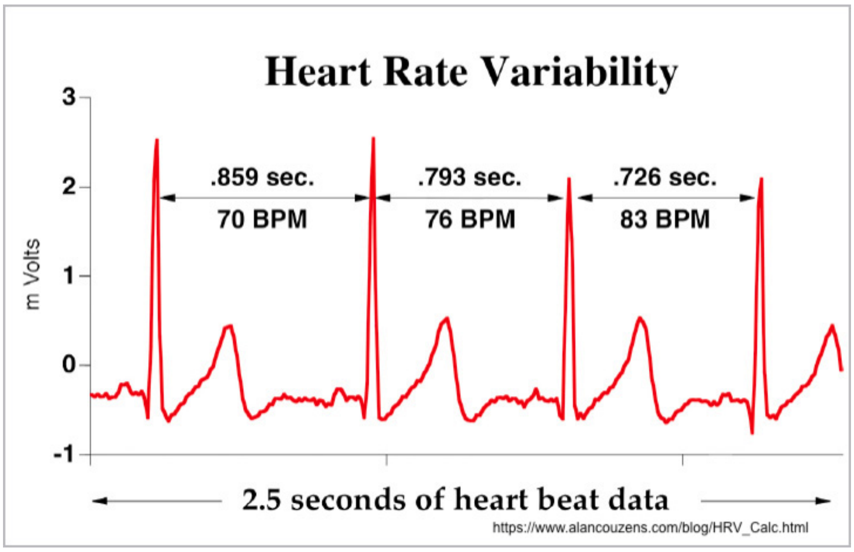


Figure B.1: This illustrates how RR-intervals are derived from an ECG signal, which serve as a basis to calculate heart rate variability.

roll in stress-related research [7, 8, 50, 70] as the power of some frequency bands have been directly associated with ANS activity. Using Fast Fourier Transformation [200], this can be extracted from the RR-based HRV signal. The following major bands have been established in research:

- Low Frequency (LF) within 0.003 to 0.04 Hz
- High Frequency (HF) within 0.15 to 0.4 Hz

While the LF component is influenced by both SNS and PNS, HF has been closely linked to parasympatic activity. the ratio LF/HF is therefore associated with the balance of SNS and PNS activity in the ANS and aforementioned studies found a significant increase in LF/HF when stress was induced [7, 8, 50, 70].

B.4 Method of Frustration Detection Study

Recall that this study aims to detect stress levels from physiological data in the context of handheld robot applications. In a first step, we integrated a HRV sensor that allows for real time detection of RR-intervals. There are a variety of sensors available in the current market. There are models where electrodes are placed on the body, e.g. through sticky patches or strapped around the chest[1]. However, this is a rather invasive option and does not go in line with the notion *pick up and go*. Therefore, we decided to use the optical finger clip sensor *CoreSense*[2], which is based on measuring the periodic fluctuation of

[1] e.g. Garmin Heart Rate Monitor Chest Strap
[2] https://elitehrv.com/corsense

vessel volume. The advantage is, that this sensor can be integrated in the robot's handle (Figure B.2/B.3). The device transmits the RR-data to the lab computer via Bluetooth, which processes the signal to obtain HR and LF/HF using an open source toolbox for cardiovascular waveform analysis [227].

Figure B.2: CoreSense: a high resolution RR-sensor that was used to derive HRV information from experiment trials.

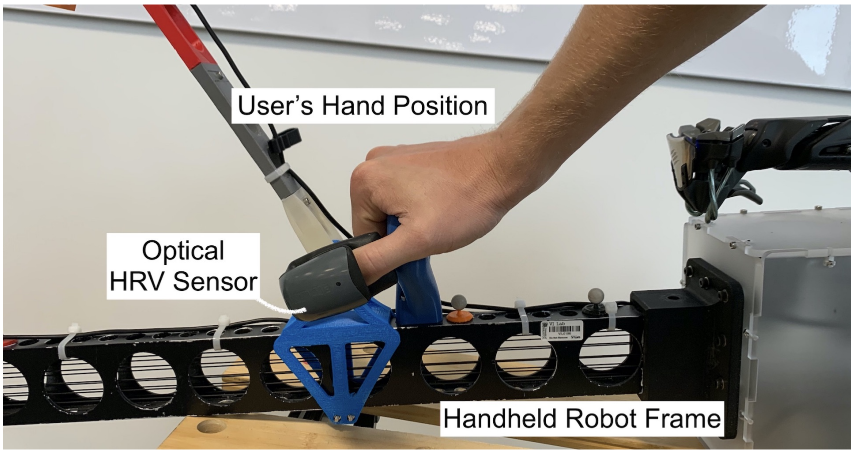


Figure B.3: This picture shows the integration of the CoreSense sensor in the handheld robot setup. It is placed close to the handle so that the participant's thumb can rest in it while holding the robot.

To assess the HRV setup, we ran a series of experiments with a task that can be parametrised with regards to frustration stimuli. We used the same block copy task, which we introduced in Section 4.2.1 for intention modelling. Put short, the participant uses the handheld robot to pick blocks from a stock area and place them on matching locations in a pattern. Recall that the block elements are simulated in a screen, which facilitates the introduction of frustration triggers. We used two psychological tricks to generate what would be perceived as user mistakes. The first one is based on the principle, that humans struggle to identify detail changes in an environment that remains the same otherwise, which is commonly known as *change blindness* [203, 218]. At the time, when the user picks a block from the stock area, the associated target element in the pattern got swapped out with another (non-matching) one. This leaves the participant with the impression that they picked the wrong piece in the first place. The second trigger was an occasional dropping of the piece during placing, as though they placed it inaccurately. In that case, the block needed to be picked from the stock area again.

We recruited 16 participants for our studies (females = 7, $m_{age} = 27, SD = 5.8$), all volunteers from our campus. There was no compensation for participating in the experiments. Each completed to experiment trials, one with the frustration triggers enabled (frustration condition) and one without (control condition). The order was counterbalanced to cancel training effects. The robot remained passive leaving the user in charge of any action. For each trial the participant had 5 min and was asked to continue the task until the time was up and to try to complete as many pieces as possible. During task execution, their HRV data was recorded through the aforementioned sensor setup. Each trial was followed by a TLX questionnaire of which the frustration component was later used for the analysis.

B.5 Results

We performed a series of pairwise t-tests to analyse the differences in means of the control condition compared to the frustration condition for each of the recorded metrics. Recall these are the TLX frustration component as a subjective rating and HRV (in LF/HF) as proposed objective metric. We also apply this analysis to the HR data, since it is available from the HRV preprocessing step anyway, so we might as well include it. 3 individuals were removed from the analysis of physiological metrics, due to malfunctioning of the sensor during trials. The results are summarised in Table B.1 and a plot can be seen in Figure B.4.

The t-test results show that the frustration component of the TLX questionnaire is rated significantly ($p = .008$) higher in frustration condition compared to control condition. Concerning heart rate variability, the mean LF/HF increased with induced frustration, however, the difference is non-significant ($p = .804$). Regarding recorded heart rate, no significant difference in Beats Per Minute (BPM) means could be determined.

	Control		**Frustration Triggered**		***t*-test Results**	
	Mean	SD	Mean	SD	*t*	*p*
TLX Frustration	24.6	21.8	46.8	22.9	t(15) = -2.803	p = .008 ***
HRV [LF/HF]	2.5	2.2	2.3	1.4	t(12) = 0.251	p = .804
Heart Rate [BPM]	97.3	23.6	92.4	23.4	t(12) = 0.542	p = .592

Table B.1: This shows the means and SDs of the frustration metrics for the two experiment conditions, i.e. control and frustration. To the right are the associated t-test results for the mean differences. The starred value is significant while non-starred values yield no significance. A plot of the data can be seen in Figure B.4

.

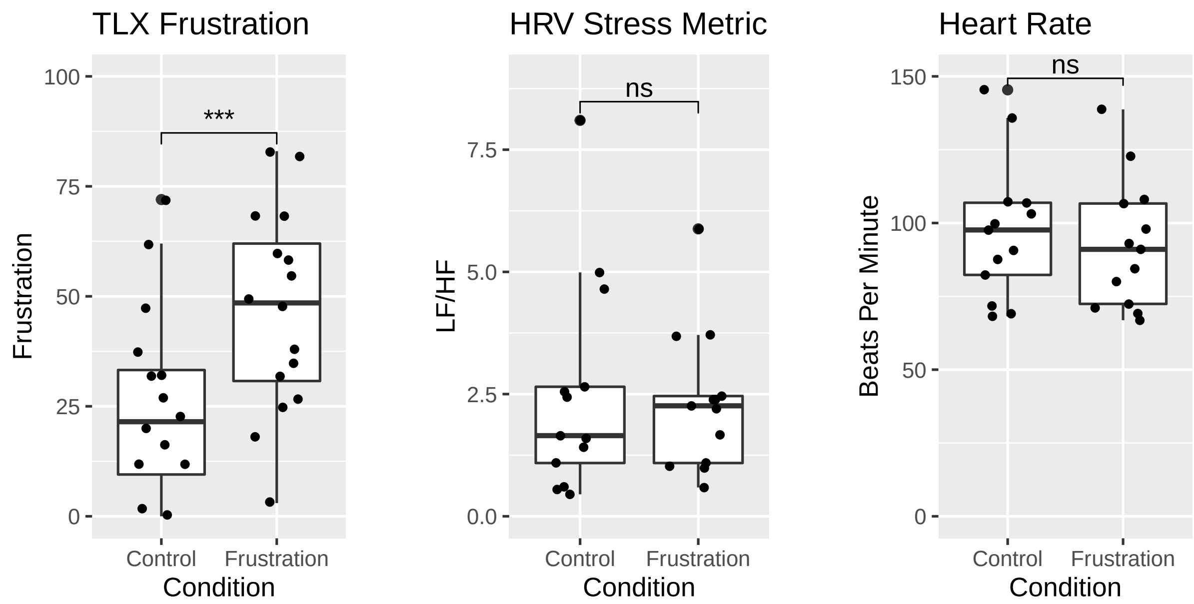


*Figure B.4: Comparison of results from physiological and subjective metrics for stress/frustration. For each metric, the frustration induced trials get compared to the control condition. The Pairwise t-tests (see Table B.1) yield no significance (ns) for the physiological metrics and high significance ($p = .008$ ***).*

B.6 Discussion and Conclusion

In this study we aimed to compare possible stress metrics as an indicator of frustration with the constraint of applicability to handheld robots, i.e. non-invasive methods. We proposed an experimental setup that allows for subtle and controlled frustration triggers in a handheld robot task. Disabling and enabling these triggers serve as a parameter to switch between the control condition and the frustration condition, respectively. The fact that a significant difference could be observed in self reported frustration levels is evidence that the proposed triggers are effective which justifies the analysis of their effect on the physiological metrics.

Concerning HR measures, the given results were expected. As discussed in the background (Section B.2), HR does not allow to distinguish between negative and positive arousal, i.e. distress and eustress. In this aspect, our results go in line with previous research.

Based on the presented literature, we expected the LF/HF to rise with increased frustration. Our results point towards that direction, however, the mean difference is not significant. Although we had to exclude data points for the HRV analysis, it is surprising that reported frustration is more reliable than the physiologic response. One reason might be that the sensor was not built for during-motion applications. Although, the hand is resting relative to the handle and the sensor, the overall body motion during the task might introduce too much interference for a reliable reading. Another reason for these results could be, that the frustration triggers were not stressful enough to stimulate sympathetic activity in the ANS, i.e. unlike in the car driving example [66], a mistake in the proposed task setup is not threatening and so the stress response is presumably smaller. We also note that the mean reported frustration amount for the condition with triggers enabled is below 50% of the possible scale. This supports the assumption that the induced frustration was moderate and not enough for a physiological response. One might argue, that the same was true in the studies by [50], where stress of office workers was detected from HRV data. However, their control condition is resting rather than a non-stressful task, which might lead to a much bigger difference since the body is inactive in that state.

We conclude that non-invasive HRV sensors are still at an early stage of development and to date insufficient for an application in handheld robots. However, this study also shows that the TLX questionnaire is sufficient as a proxy to perceived frustration. This study is a first attempt to bring physiologic frustration detection to mobile robotic platforms. We suggest that research could continue in that direction in the future, as HRV sensors continue to improve.

Appendix C

Bill of Materials

General Materials (used in all chapters)

- Handheld Robot hardware introduced in [59], CAD models available at www.handheldrobotics.org
- Desktop Computer: Intel Core i7 CPU 3.4 GHz, 16 GB RAM, NVIDIA Quadro M2000 Windows 7 (required for the XNA game studio framework)
- Motion Capturing: OptiTrack (Natural Point) with 7 cameras (model Flex 3), used with the associated software Motive v1.8.
- Microsoft XNA Game Studio (requires Windows 7)
- Display for experimental content: TV screen (LCD display, 42")

Materials Chapter 3 and 4

- Tobii Eye Tracker 4C
- Tobii X SDK
- Handheld Robot Supplements (see list below)

Materials Chapter 5

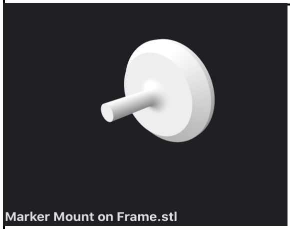
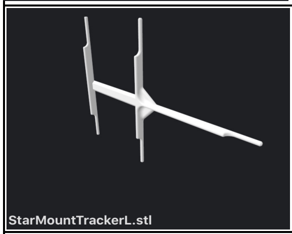
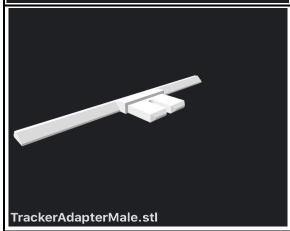
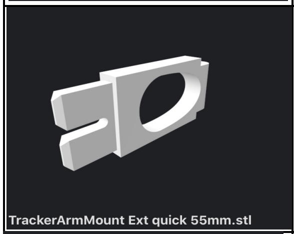
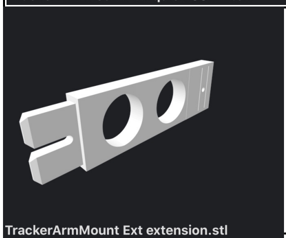
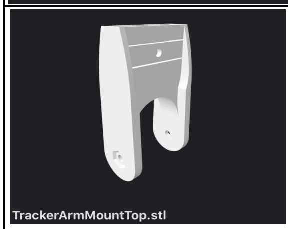
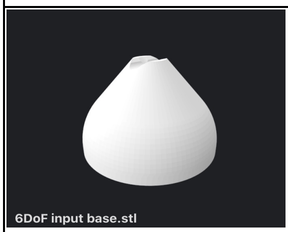
- Microsoft HD LifeCam Studio (2x)
- Handheld Robot Supplements (see list below)

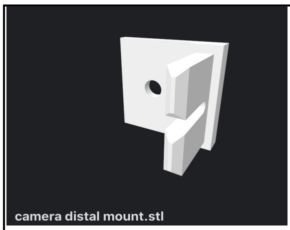
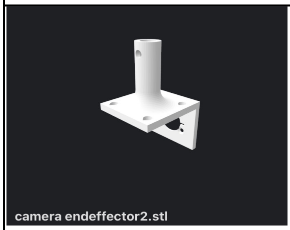
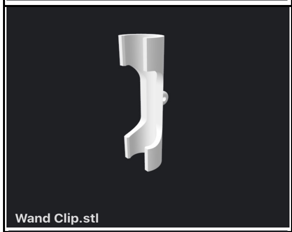
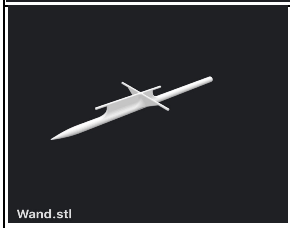
Material for Appendix B

- Elite HRV CoreSense (HRV sensor)

List of Handheld Robot Supplements

All CAD models are available at www.handheldrobotics.org (per listing below)

Tumbnail	Use in Chapter	Quantity	File Name	Comment
	3	6	Marker Mount on Frame.ipt	Used to attach markers to the frame of the robot
	3	2	MarkerMountTracker.ipt	Mounting frame to attach markers to the eye gaze tracker
	3	1	TrackerAdapterMale.ipt	Adapter to attach the tracker to the mounting
	3	1	TrackerArmMount Ext quick 55mm.ipt	Top piece of tracker mounting
	3	1	TrackerArmMount Ext extension.ipt	Middle piece of tracker mounting
	3	1	TrackerArmMountBase V2.ipt	Mounting connection to the robot's frame
	3	1	TrackerArmMountTop.ipt	Base link of tracker mounting
	5	1	6DoF input base.ipt	5-DoF input base

	5	1	camera distal mount.ipt	Camera adapter (distal worksite view)
	5	1	camera endeffector2.ipt	Camera adapter (close up tip view)
	5	1	Wand Clip.ipt	Attaches the wand to the base
	5	1	Wand.ipt	Used as 5-DoF user input with attached markers

Appendix D

Questionnaires

Figure 8.6

Other comments: (after filling the form)

NASA Task Load Index (Shooter Game)

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

ID	Trial	mode

Mental Demand — How mentally demanding was the task?

Very Low — Very High

(e.g. thinking, deciding, estimating, looking, searching, etc. Was the task simple or complex, exacting or forgiving?)

Physical Demand — How physically demanding was the task?

Very Low — Very High

(e.g. pushing, pulling, turning, controlling activating, eye movement etc.)? Was the task easy or demanding, slow or brisk, slack or tiring, restful or laborious?

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

How successful do you think you were in accomplishing the goal of the task set by the experiment (or yourself)? How satisfied were you with your performance in accomplishing the goals?

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

How hard did you have to work (mentally and physically) to accomplish your level of performance?

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

When thinking about the robot's behaviour, please rate how much you agree or disagree with the following statements:

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
The robot helped me with the task	☐	☐	☐	☐	☐
The robot obstructed me during the task	☐	☐	☐	☐	☐

System Usability Scale

ID: Mode:

Please indicate to what extent you agree with the following statements:

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. I think that I would like to use this robot frequently.	☐	☐	☐	☐	☐
2. I found the system unnecessarily complex.	☐	☐	☐	☐	☐
3. I thought the robot was easy to use.	☐	☐	☐	☐	☐
4. I think that I would need the support of a technical person to be able to use this robot.	☐	☐	☐	☐	☐
5. I found the various functions in this robot were well integrated.	☐	☐	☐	☐	☐
6. I thought there was too much inconsistency in this system.	☐	☐	☐	☐	☐
7. would imagine that most people would learn to use this robot very quickly.	☐	☐	☐	☐	☐
8. I found the robot very cumbersome (awkward) to use.	☐	☐	☐	☐	☐
9. I felt very confident using the robot.	☐	☐	☐	☐	☐
10. I needed to learn a lot of things before I could get going with this robot.	☐	☐	☐	☐	☐