
# FROM REGISTRATION TO EVALUATION: TOWARDS EFFICIENT HUMAN-ROBOT INTERACTION

DISSERTATION

zur Erlangung des akademischen Grades

**Doktoringenieurin (Dr.-Ing.)**

angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg

von **M.Sc. Tonia Mielke**

geb. am 08.08.1997 in Wittingen

Gutachterinnen/Gutachter
Prof. Dr. Christian Hansen
Prof. Dr. Marta Kersten-Oertel
Prof. Dr. Stefanie Speidel

Magdeburg, den 15.12.2025

Tonia Mielke: *From Registration to Evaluation: Towards Efficient Human-Robot Interaction,* © 2025

# ABSTRACT

Human-Robot Interaction (HRI) enables the integration of human expertise into robot control, allowing systems to adapt to dynamic and complex environments. This requires intuitive and efficient user interfaces, with hand gesture interaction emerging as a promising approach. Additionally, the integration of Extended Reality (XR) has shown potential to support HRI. However, several challenges remain that must be addressed to ensure efficient HRI.

First, accurate interaction and the integration of XR into robotic workspaces require a spatial registration between the robot and the corresponding interaction or virtual coordinate frames. Therefore, the first part of this thesis explores XR-to-robot registration techniques, demonstrating that point-based methods enable efficient registration. Further analysis revealed that registration accuracy can be significantly improved by carefully selecting the tracking method and point characteristics. Registration accuracy can be further enhanced by employing a proposed refinement approach, which integrates points recorded between registration points.

The second part addresses interaction-specific challenges. Hand gestures inherently lack haptic feedback, which can complicate contact-intensive tasks. Two approaches are explored to mitigate this: partial automation and sensory substitution. Results indicate that both strategies improve performance while partial automation also reduces operator workload. Another challenge arising is fatigue caused by the sustained hand posture. To address this, different control modes and sensor placements are explored. It is found that position control outperforms rate control, and that while robot-mounted sensors enable precise control, environment-mounted sensors can reduce fatigue.

Human-centered investigations require extensive user studies, which are often constrained by the cost and limited availability of physical robotic setups. XR environments offer a promising alternative by simulating such setups. To evaluate the validity of results from user studies conducted in XR environments, this thesis compares two degrees of virtualization across three tasks. The findings indicate that while performance in precision tasks is setup-dependent, XR-based studies can produce valid results for broader tasks. A follow-up study on force assistance techniques in both XR and physical setups extends these findings, showing that even when absolute performance differs, relative trends remain consistent across environments.

By addressing three key factors - registration, interaction design, and evaluation setup - this work is an important step towards efficient HRI. The findings provide evidence-based recommendations for registration techniques, force assistance strategies, interaction design, and the use of XR in user studies, contributing to more intuitive and accessible HRI.

# ZUSAMMENFASSUNG

Die Mensch-Roboter-Interaktion (HRI) ermöglicht die Integration menschlicher Expertise in die Robotersteuerung, wodurch sich Robotersysteme an dynamische und komplexe Umgebungen anpassen können. Ein vielversprechender Ansatz für die dafür erforderliche intuitive und effiziente Steuerung ist die Interaktion mittels Handgesten. Darüber hinaus hat sich die Integration von Extended Reality (XR) als unterstützend für die HRI erwiesen. Dennoch bestehen weiterhin Herausforderungen, deren Bewältigung eine Voraussetzung für eine effiziente Steuerung darstellt.

Zunächst erfordern die präzise Interaktion und Integration von XR in Arbeitsbereiche von Robotern eine räumliche Registrierung zwischen dem Roboter und den entsprechenden Interaktions- oder virtuellen Koordinatensystemen. Daher untersucht der erste Teil dieser Arbeit XR-zu-Roboter-Registrierungstechniken und zeigt, dass punktbasierte Methoden eine effiziente Registrierung ermöglichen. Weitere Analysen verdeutlichen, dass die Registrierungsgenauigkeit durch die gezielte Auswahl der Trackingmethode und der Eigenschaften der verwendeten Punkte signifikant verbessert werden kann. Die Registrierungsgenauigkeit kann weiter erhöht werden, indem ein Verfeinerungsansatz angewendet wird, der Punkte einbezieht, die zwischen den Registrierungspunkten erfasst wurden.

Der zweite Teil der Arbeit befasst sich mit interaktionsspezifischen Herausforderungen. Freihandgesten bieten definitionsgemäß kein haptisches Feedback, wodurch kontaktintensive Aufgaben erschwert werden. Um dies zu kompensieren, werden zwei Ansätze untersucht: Teilautomatisierung und sensorische Substitution. Die Ergebnisse zeigen, dass beide Strategien die Leistung verbessern können, wobei die Teilautomatisierung zusätzlich die Arbeitsbelastung der Benutzer*innen reduziert. Eine weitere Herausforderung ist die Muskelermüdung durch die andauernde Handhaltung. Um dem entgegenzuwirken, werden verschiedene Steuerungsmodi und Sensorplatzierungen evaluiert. Dabei zeigte sich, dass die Positionssteuerung der Geschwindigkeitssteuerung überlegen ist und dass am Roboter befestigte Sensoren zwar eine präzise Steuerung ermöglichen, in der Roboterumgebung platzierte Sensoren, jedoch die Ermüdung verringern können.

Da menschenzentrierte Untersuchungen umfangreiche Nutzerstudien erfordern, diese jedoch häufig durch Kosten und begrenzten Zugang zu physischer Hardware eingeschränkt sind, bieten XR-Umgebungen eine vielversprechende Alternative, indem Testumgebungen simuliert werden. Zur Bewertung ihrer Validität vergleicht diese Arbeit zwei Virtualisierungsgrade anhand dreier Aufgaben. Die Ergebnisse zeigen, dass die Leistung bei Präzisionsaufgaben zwar von der Studienumgebung abhängt, XR-basierte Studien für gröbere Aufgaben jedoch valide Ergebnisse liefern können. Eine ergänzende Studie zu Kraftunterstützungstechniken in XR- und physischen Umgebungen erwei-

tert diese Ergebnisse und zeigt, dass selbst bei unterschiedlichen absolute Ergebnissen, relative Trends über die Umgebungen hinweg vergleichbar sind.

Durch die Analyse der drei zentralen Einflussfaktoren - Registrierung, Interaktionsdesign und Evaluierungsumgebung - leistet die vorliegende Arbeit einen wichtigen Beitrag zur Entwicklung effizienter HRI. Aus den Ergebnissen lassen sich evidenzbasierte Empfehlungen für Registrierungstechniken, Strategien zur Kraftunterstützung, das Interaktionsdesign sowie den Einsatz von XR in Nutzerstudien ableiten. Diese tragen zu einer intuitiveren und zugänglicheren HRI bei.

## ACKNOWLEDGMENTS

The research reported in this thesis was conducted at Otto von Guericke University Magdeburg and the Research Campus *STIMULATE* in Magdeburg, Germany. It was partially funded by the Federal Ministry of Research, Technology and Space under grant number 13GW0473A.

# CONTENTS

# LIST OF FIGURES

## LIST OF TABLES

# ACRONYMS

| | |
|---|---|
| AAD | Average Absolute Deviation |
| AMax | Average Maximal Applied Force |
| AR | Augmented Reality |
| AV | Augmented Virtuality |
| AW | Acoustic Window Task |
| BF | Bayes Factor |
| CT | Computed Tomography |
| DOF | Degree of Freedom |
| FA | Force Automation |
| FLE | Fiducial Localization Error |
| FOV | Field of View |
| FRE | Fiducial Registration Error |
| $H_0$ | Null Hypothesis |
| $H_1$ | Alternative Hypothesis |
| HCI | Human-Computer Interaction |
| HG | Hand-Guiding |
| HI | Hand Gesture Interaction |
| HRI | Human-Robot Interaction |
| HRC | Human-Robot Collaboration |
| IQR | Interquartile Range |
| IMU | Inertial Measurement Unit |
| HMD | Head-Mounted Display |
| ICP | Iterative Closest Point Algorithm |
| LORA | Levels of Robot Autonomy |
| M | Mean |
| MR | Mixed Reality |
| NUI | Natural User Interface |
| NT | Needle Tip Task |
| RANSAC | Random Sample Consensus |
| RE | Real Environment |
| RQ | Research Question |
| SA | Simulated Annealing |
| SD | Standard Deviation |

| | |
|---|---|
| SEQ | Single Ease Question |
| TCP | Tool-Center Point |
| TCT | Task Completion Time |
| TLX | Task Load Index |
| TRE | Target Registration Error |
| UI | User Interface |
| US | Ultrasound |
| UWP | Universal Windows Platform |
| VR | Virtual Reality |
| VE | Virtual Environment |
| XR | Extended Reality |
| XRP | Extended Reality Prototyping |

# 1 INTRODUCTION

## 1.1 MOTIVATION

Robotics is emerging as a prominent field not only in industrial settings but also across diverse domains such as medicine, service, and assistive technologies [230]. While robot manipulators have traditionally been used to automate tasks that are *dirty, dull, or dangerous* [313], modern applications often involve unstructured environments, complex decision-making, and dynamic interactions that exceed the current capabilities of fully autonomous systems [91]. As a result, Human-Robot Collaboration (HRC) is gaining importance, enabling the integration of operators' perceptual abilities and expertise into robotic workflows [301, 360]. However, this collaboration leads to the performance of robotic systems being dependent on the capability of the operator to comprehend and operate the robot, highlighting the need for efficient user interfaces allowing accurate and fast control while minimizing the user's workload [136, 270].

Human-Robot Interaction (HRI) is commonly facilitated through hand-guiding, where the operator physically manipulates the robot, or through external controllers such as joysticks or control panels. However, these methods can be associated with high physical and cognitive workloads [183]. To overcome these limitations, hand gestures have emerged as a natural and intuitive interaction modality [356, 373]. Leveraging the inherent human ability to communicate and manipulate through gesticulation, mid-air hand gesture control offers an intuitive, direct, and flexible means of conveying commands without the need for physical devices [331]. To further enhance this interaction, Extended Reality (XR) technologies offer novel opportunities enabling more intuitive interaction and providing effective feedback [311]. XR has demonstrated potential to simplify robot programming [113], real-time control [357], visualize task-relevant information [175], communicate robot intent [202], and improve workspace awareness [240].

Despite these advances, critical challenges remain that require further research to enable efficient robot control [270]. Both hand gesture interaction and XR integration depend on precise spatial alignment between the coordinate reference frames of the robot and those of the interaction or augmentation. This alignment, commonly referred to as registration, is essential for correctly mapping hand gestures to robot motion and for placing virtual content in the intended spatial context. However, registration is often overlooked, highlighting the need for further research into how to efficiently and accurately align these coordinate systems. Moreover, utilizing hand gestures for robot control presents several challenges. First, the interaction is typically unilateral: it transfers the hand's pose to the robot but lacks feedback channels to convey critical information, such as haptic feedback [235, 8, 373].

Second, maintaining the hand in a mid-air position for extended periods can be physically demanding and lead to fatigue [140, 373].

To address these challenges, research through user studies is essential. However, conducting user studies in physical robotic workspaces tends to be costly, complex, and often inaccessible, as they involve expensive hardware and safety protocols [252, 338]. Extended Reality Prototyping (XRP) offers a promising alternative to physical evaluation environments by partially simulating user study setups using XR. Through interactive virtual representations of the robot or test setup, prototypes can be evaluated without the need for physical hardware. While XRP has the potential to reduce experimental costs and accelerate research and development [120, 251], the transferability of results obtained through XR-based studies still requires thorough investigation.

Motivated by these challenges, this thesis investigates key factors essential for the development of gesture-based HRI with robot manipulators, with the goal of enhancing overall HRI efficiency. It focuses on designing interactive registration methods to improve XR-robot alignment, developing hand gesture interaction techniques that enable efficient control, and evaluating the validity of XR-based prototyping approaches for HRI.

## 1.2 RESEARCH FOCUS

Building on the identified challenges and goals, this thesis focuses on three key research areas essential for the development of hand gesture-based HRI to investigate the main Research Question (RQ): *How can efficiency be improved in HRI?*

### 1.2.1 *XR-to-Robot Registration*

Adequate registration is required to align the interaction and the robotic space for efficient HRI. This is also essential for integrating XR components into the robotic workspace. While most current research focuses on the design of interaction concepts and XR elements, the registration process itself is underexplored. Therefore, the first part of this thesis focuses on investigating registration techniques for XR-to-robot alignment. This leads to the first RQ:

> **RQ1** | How can efficient and accurate XR-to-robot registration be achieved?

### 1.2.2 *Hand Gesture-Based HRI*

Efficient HRI further relies on well-designed interaction concepts. Previous work has highlighted that hand gestures are a promising interaction technique, enabling natural [373], intuitive [183], and flexible [62] interactions. However, despite these advantages, previous studies have also identified remaining challenges, such as missing visual or sensory feedback [183, 8] and fatigue [44, 282, 373]. Therefore, the second part of this thesis explores

how to address these challenges and thereby enhance performance in HRI. This leads to the following RQ:

**RQ2** | How can efficient mid-air gesture-based HRI be designed?

To address this, two specific areas of focus are considered. The first concerns force perception. Due to the indirect interaction with the environment in HRI, users often have limited awareness of contact forces at the robotic end effector. Therefore, the first sub-question investigates whether force assistance approaches, which are designed to address issues caused by the lack of haptic feedback, can support users in contact-intensive tasks where force perception is essential.

**RQ2.1** | Can force assistance improve efficiency in HRI?

The second sub-question focuses on the design of the hand gesture interaction itself. The efficiency of this interaction approach is influenced by design parameters such as the control mode and sensor placement. A specific challenge that arises during hand gesture interaction is fatigue, caused by prolonged hand postures. To address this, the second sub-question investigates how hand gesture interaction techniques can be designed to support efficient robot control while minimizing fatigue.

**RQ2.2** | How can fatigue be reduced in hand gesture-based HRI?

### 1.2.3 *XR Prototyping for HRI Evaluation*

To address RQ2, human-centered research is required, typically implemented through user studies. However, conducting such studies with physical robotic arms can be costly, and access to the necessary hardware may be limited. As a result, the use of simulated environments in XR has become a common approach in the literature. Despite the growing popularity of XRP for HRI, few studies have investigated whether results obtained in simulated environments are valid, specifically, whether their results can be reliably transferred to real-world scenarios. This leads to the final RQ of this thesis:

**RQ3** | Can XRP produce transferable results for HRI research?

### 1.2.4 *Use Case: Robotic Ultrasound*

For addressing RQ2 and RQ3, a representative use case is required to investigate HRI approaches and the applicability of XRP systematically. RQ2.1 investigates strategies to mitigate challenges arising from limited haptic perception, which is especially relevant in robotic tasks that involve direct physical interaction with the environment. These contact-intensive tasks

include industrial applications like grinding [187, 371] and polishing [82], as well as medical tasks such as suturing [314], palpation [24, 122], and ultrasound imaging [234, 308].

In robotic ultrasound, accurate force application is particularly critical. A consistent contact force between the probe and the tissue is essential to ensure effective acoustic coupling. Variations in this force can deform tissue, compromise image quality, and reduce measurement repeatability [53, 117, 350]. Studies have shown that maintaining a stable contact force improves imaging reliability and reduces measurement errors, resulting in more accurate and reproducible outcomes [165]. Given the importance of controlled force application, robotic ultrasound is a meaningful use case for contact-intensive robotic tasks. It will thus serve as an exemplary use case for addressing RQ2 and RQ3.

## 1.3 CONTRIBUTION

By investigating registration, interaction, and evaluation, this thesis contributes to enhancing the overall efficiency of HRI. It is based on seven experiments, whose findings are presented in seven peer-reviewed publications [1–7]. The results of these studies informed the formulation of empirically based guidelines for the design and implementation of more efficient HRI systems.

The scientific contribution of this work, including the conceptualization, development of methodology, and implementation, was carried out by the author of this thesis. In addition, the author of this thesis conducted the literature review, designed and executed the experiments, and led the writing of the original manuscripts. Coauthors contributed through discussions, feedback, assistance with statistical analysis, and support in photographic documentation. For two publications [1, 2], one coauthor explicitly assisted with the implementation of the ultrasound simulation.

## 1.4 THESIS STRUCTURE

The structure of this thesis is outlined as follows. Additionally, figure 1.1 provides an overview.

**Part I: XR-to-Robot Registration**
The first part focuses on aligning the XR coordinate space, which is used for content visualization and hand gesture tracking, with the robotic workspace. Related work is presented, and two experiments are described:

*Chapter 4: Investigation of Registration Methods*
This chapter compares three registration approaches: a manual interactive method and two point-based techniques. The point-based methods rely either on physically defined registration points or on directly tracking the robotic end effector to establish mid-air points. A comparative user study and its results are presented.

*Chapter 5: Investigation of Registration Parameters*
To improve the accuracy of point-based methods, this chapter investigates several influencing factors, including tracking method, marker size, registration point characteristics, and XR device placement. Additionally, a surface-based refinement approach is explored that uses points captured between the main registration points to improve accuracy.

**Part II: Hand Gesture-Based Human-Robot Interaction**
The second part addresses the efficiency of hand gesture-based HRI. By presenting related work, two key challenges associated with this interaction method are identified: the missing haptic feedback and fatigue. These challenges are then addressed through three experiments:

*Chapter 8: Force Assistance*
This chapter investigates two assistance approaches to compensate for missing haptic feedback. First, related work is described, followed by two experiments. In Section 8.2, contact force automation is presented as a method that autonomously regulates movement in one Degree of Freedom (DOF) to maintain the desired contact force. An experiment evaluates this approach on two practical tasks, comparing it to a state-of-the-art hand-guiding technique. In Section 8.3, sensory substitution is explored by mapping force feedback onto visual and vibrotactile modalities. A user study compares these feedback types and investigates factors affecting visual force feedback, such as design and placement.

*Chapter 9: Sensor Placement and Control Modes*
This chapter addresses fatigue caused by hand gesture interaction by first presenting related work, followed by a user study. In this experiment, different sensor placements and control modes are investigated.

**Part III: XR Prototyping for HRI Evaluations**
The third part explores the potential of XR to simulate user study setups for HRI research. To this end, two experiments are presented.

*Chapter 12: Investigation of XRP Virtualization Levels*
This chapter describes the development of different XRP prototypes, including different levels of virtualization, such as using a physical robot with a simulated task or a fully simulated scenario. An experiment is presented that evaluates how results transfer across these levels of virtualization in three practical tasks.

*Chapter 13: Investigation of XRP for Force Assistance*
This chapter presents the second experiment on XRP for HRI, focused on contact-intensive tasks. A user study is described that evaluates the force assistance approaches from Chapter 8 in both a physical and a simulated XRP setup.

**Chapter 15: Closing**
The final chapter summarizes the key findings and addresses the research questions introduced at the beginning of the thesis. Additionally, limitations are discussed, and directions for future work are proposed.

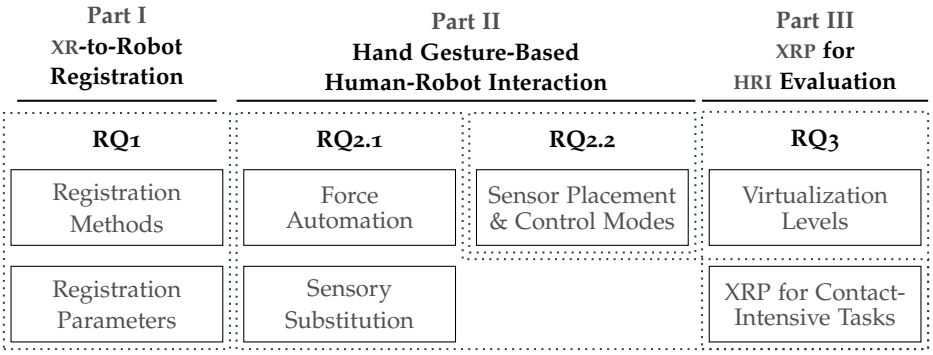



Figure 1.1: Schematic overview of thesis structure.

# 2 PRELIMINARIES

## 2.1 TECHNICAL BACKGROUND

This section provides an overview of the technical foundations relevant to this thesis, with a focus on robotics, gesture-based interaction, XR, registration, and the use case of robotic ultrasound.

### 2.1.1 *Robotics*

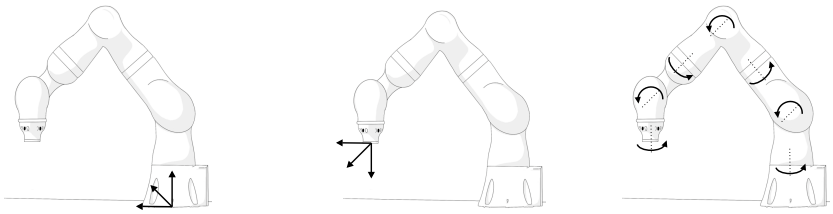
According to the Robotics Institute of America (RIA), "A robot is a reprogrammable multifunctional manipulator designed to move material, parts, tools, or specialized devices through variable programmed motions for the performance of a variety of tasks" [71]. Robots can generally be classified into two categories: those with a fixed base (*robot manipulators*) and those with a mobile base (*mobile robots*) [304, p. 1]. This work focuses on robot manipulators, which typically consist of the manipulator (an arm composed of links and joints that enable movement) and an end effector (attached to the last joint) [154, pp. 2-8]. The end effector can be equipped with different tools, enabling interaction with the environment to fulfill different tasks. The area that can be reached by the end effector is called *robotic workspace*.

This section introduces key concepts in robot control, including kinematics and control modes, followed by an overview of autonomy levels and principles of HRI relevant to this thesis.

#### 2.1.1.1 *Robot Control*

KINEMATICS For robot control, both the robot's posture and control commands can be described using different coordinate frames [230] (see Figure 2.1). The *Robot Reference Frame* (also referred to as the World Reference Frame) defines positions in Cartesian space and provides a fixed global reference for specifying the position and orientation of the end effector. In contrast, the *Tool Reference Frame* is attached to the end effector, and its orientation depends on the end effector's pose, allowing for the specification of positions relative to the end effector itself. Finally, the *Joint Reference Frame* describes the robot's configuration in terms of the individual angles of each joint, rather than Cartesian positions.

These coordinate frames are fundamental because desired motions are typically specified as trajectories in Cartesian space, while the robot's control system requires commands in joint space [99]. The analytical relationship between joint angles and the end effector's position and orientation is described by *kinematics* [304, p. 30]. *Forward kinematics* derives the Cartesian pose of the end effector in the robot reference frame from given joint angles. Conversely,



(a) Robot Reference Frame (b) Tool Reference Frame (c) Joint Reference Frame

Figure 2.1: Overview of the robot's reference frames.

*inverse kinematics* calculates the joint angles required to achieve a desired end effector pose, enabling control in Cartesian space [303, p. 27].

In redundant robotic systems, the number of joints exceeds the DOFs required to define end effector poses. Therefore, a single Cartesian end effector position can correspond to multiple joint configurations. Movements of the joint angles that do not affect the end effector's position or orientation are referred to as *nullspace motion* [304, p. 122].

MOTION CONTROL Motion control of the robot can be realized through *position control*, where trajectories are defined by specifying either joint angles or Cartesian end effector positions [200]. The resulting end effector movement is achieved by minimizing the error between the current and desired joint angles. However, some tasks require not only movement along a trajectory, but also physical interaction between the end effector and the environment. Such interactions result in contact forces, which, if excessive, can cause stress or damage to both the robot and its surroundings [304, p. 363].

In position control, avoiding excessive contact forces requires accurate modeling of both the manipulator and the environment. This can be particularly challenging, as detailed models of the environment are often difficult to obtain [304, p. 364]. As an alternative, *force control* incorporates measured contact forces directly into the control loop [200]. Force control approaches can be broadly categorized into indirect and direct methods [303, pp. 161-164].

In indirect force control, such as *impedance control*, the robot is modeled as a virtual mass-spring-damper system with adjustable stiffness and damping parameters. The contact force arises from deviations between the actual and desired end effector motions, modulated by these parameters, enabling compliant behavior during interaction. In contrast, direct force control requires explicit specification of both the desired motion and the desired contact force. This is often implemented through *hybrid position/force control*, which decomposes the task space into two mutually independent subspaces. By allowing position control in some DOFs while controlling force in others, it enables the coordinated control of motion and interaction forces.

#### 2.1.1.2 *Levels of Autonomy*

Robot manipulators can be controlled under different Levels of Robot Autonomy (LORA), ranging from teleoperation to fully autonomous systems. To determine a robot's LORA, Beer et al. [30] proposed a framework that

classifies autonomy based on the functional allocation of sensing, planning, and acting (see Table 2.1). In manual teleoperation, the human is responsible for all tasks, including sensing the environment, planning the control actions, and executing the robot's movements. As the LORA increases, these responsibilities are progressively delegated to the robot, shifting the human's role from direct control to supervision. The autonomy level of a robotic system directly influences the nature and extent of HRI required [147]. While higher levels of LORA may reduce the frequency of interaction, they require more sophisticated or higher-level forms of HRI.

Table 2.1: LORA and associated responsibilities for sensing, planning, and acting (Robot: 🤖, Human: 🧍), adapted from [30].

| LORA | Sense | Plan | Act |
|---|---|---|---|
| Manual Teleoperation | 🧍 | 🧍 | 🧍 |
| Action Support | 🧍/🤖 | 🧍 | 🧍/🤖 |
| Assisted Teleoperation | 🧍/🤖 | 🧍 | 🧍/🤖 |
| Batch Processing | 🧍/🤖 | 🧍 | 🤖 |
| Decision Support | 🧍/🤖 | 🧍/🤖 | 🤖 |
| Shared Control Human Initiative | 🧍/🤖 | 🧍/🤖 | 🤖 |
| Shared Control Robot Initiative | 🧍/🤖 | 🧍/🤖 | 🤖 |
| Supervisory Control | 🧍/🤖 | 🤖 | 🤖 |
| Executive Control | 🤖 | (🧍)/🤖 | 🤖 |
| Full Autonomy | 🤖 | 🤖 | 🤖 |

While LORA is a useful framework for classifying the autonomy of robotic systems, it is important to note that the level of autonomy may vary depending on the task, environment, or system state [298]. Therefore, LORA should be understood as a spectrum rather than a set of discrete levels. Different subtasks within a system may be associated with different levels of autonomy, which can be combined as needed.

This thesis focuses on the lower levels of LORA, where the human operator's perception is used to sense the environment, plan motion, and directly control the robot. However, the insights are also relevant to more general systems in which some subtasks may be performed at higher levels of autonomy, using sensor data to plan and control the robot autonomously, while manual control remains essential for others.

#### 2.1.1.3 *Human-Robot Interaction*

HRI can be classified into three categories: coexistence, cooperation, and collaboration [136]. In *Human-Robot Coexistence*, robots and humans share a dynamic workspace without working on a common task, where the main objective is mutual avoidance. In *Human-Robot Cooperation*, humans and robots work towards the same overall goal, but perform separate, clearly divided subtasks. In Human-Robot Collaboration (HRC), there is direct interaction between humans and robots to solve a complex task collaboratively, enabling the strengths of both human operators and robots to be effectively combined.

The human operator assumes a specific role in these different categories, which can be classified as supervisor, operator, collaborator, cooperator, or bystander [237]. The *supervisor* monitors the robot and provides instructions, a role that is especially relevant at higher LORA. In contrast, the *operator* corresponds to lower LORA, where the human directly controls the robot and maintains a higher level of authority. Aligning with the interaction categories described above, the *collaborator* works closely with the robot to achieve a shared goal, while the *cooperator* pursues the same goal independently, without requiring direct coordination. Finally, the *bystander* does not interact with the robot but shares the same workspace.

This thesis focuses on the human role of the operator, where interaction involves directly controlling the robot. This approach enables the integration of human expertise into robot control [266]. Such interaction is particularly valuable for teaching robots new tasks, for handling tasks that cannot be fully automated due to complex or unstructured environments, or in situations where the risk of harm is high [266, 301, 361]. As robots are increasingly deployed in diverse domains, operators may not always possess robotics expertise. Therefore, intuitive interfaces that can be used by untrained users are essential [136, 266]. The HRI can be characterized by interaction interfaces and interaction paradigms.

INTERACTION INTERFACES Robot control can be realized either through direct physical contact or via indirect methods [136]. In physical interaction, force/torque sensors are used to detect external forces, allowing the robot to be guided compliantly by the user. In contrast, indirect interaction can be achieved through user interfaces or sensory input devices [176]. User interfaces include teach pendants (handheld devices including a screen and buttons), joysticks, haptic input devices, game controllers, or applications on smartphones and desktop computers. Sensory input employs external sensors to enable control through, for example, hand gestures, speech, or brain-machine interfaces [176].

CONTROL PARADIGMS When user input is used to continuously control a robot's pose, different control paradigms can be applied. The two most common approaches are position control and rate control [169]. It is worth noting that the term *position control* is used in the literature to refer to both the control paradigm and the motion control principle (see Section 2.1.1.1), which are distinct and unrelated concepts. In the *position control* paradigm, changes in user input, such as movement of a controller or gesture, are directly mapped to Cartesian position changes of the robotic end effector [365]. The input may be applied either scaled or unscaled, and typically across multiple DOFs. In contrast, *rate control* interprets the user's positional or orientational input as a command for velocity, mapping it to continuous motion of the robotic end effector at a corresponding speed and direction [169]. The end effector continues to move as long as the input is held, stopping only when the input returns to a neutral position.

### 2.1.2 *Gesture-Based Interaction*

As described in Section 2.1.1.3, different interfaces can be used to enable HRI. One approach explored in the broader field of human-machine interaction is the use of Natural User Interfaces (NUIs), which provide intuitive and direct interfaces that draw on natural, everyday behavior [256, p. 472]. The goal is to create interactions that feel natural to both novice and expert users by using context-based metaphors [346, p. 13]. Gesture-based interaction is a core element of NUIs, offering an intuitive way to control systems through physical movement [256, p. 507]. Gestures are defined as "expressive, meaningful body motions involving physical movements of the fingers, hands, arms, head, face, or body" [222]. In HRI, hand gestures are of particular interest as they are commonly used to manipulate and point at objects, making them a natural and intuitive means of controlling robots in 3D space [335].

To provide a structured understanding of hand gesture interaction in the context of HRI, the following subsections cover common gesture styles, hand tracking technologies, and the integration of gestures with other input modalities through multimodal interaction.

#### 2.1.2.1 *Gesture Styles*

Gestures can be categorized into deictic, manipulative, semaphoric, gesticulation, and language gestures [163]. These gestures enable different interaction tasks such as selection and manipulation, navigation, system control, and symbolic input [47, p. 135]. In HRI, key tasks include manipulation for robot control, selection, and system control. Therefore, deictic, manipulative, and semaphoric gestures are particularly relevant (see Figure 2.2).

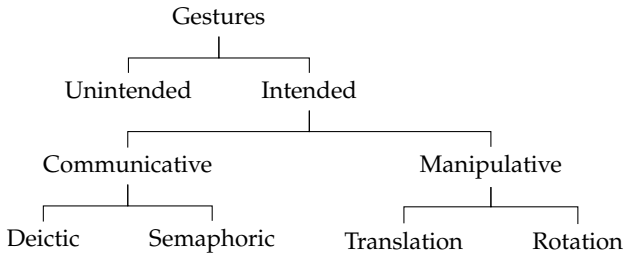



Figure 2.2: Taxonomy of gestures for HRI (adapted from [245, 332]).

- *Deictic Gestures* involve pointing to indicate or select spatial locations [163]. A well-known example is Bolt's "Put-That-There" system [42], in which pointing gestures are used to select and manipulate objects. In robot control, deictic gestures can specify local reference points to specify objects to interact with or define target positions [335].
- *Semaphoric Gestures*, also referred to as symbolic gestures, rely on a predefined set of static or dynamic gestures mapped to discrete commands [256, p. 499]. This enables gesture-based activation of specific functions. In robot control, such gestures can be employed to allow for explicit or implicit commands, e.g., for system control [332].

- *Manipulative Gestures* aim to directly control an object through a close mapping between user motion and system response [261]. In robot control, these gestures allow continuous mapping of hand movements to robotic end effector motion [332]. The degree of control can vary across one or more DOFs, including separate manipulation of translational or rotational components, or full 6-DOF control [369].

#### 2.1.2.2 *Hand Tracking*

Accurate hand tracking is essential for gesture-based HRI, as it determines how reliably user intent can be recognized and translated into robot behavior. Two main approaches are commonly used to detect hand motion: device-based tracking and optical tracking [256, pp. 530-533].

*Device-based approaches* involve holding sensors in the hand or attaching them to the body. To track hand motion, handheld controllers [334] or inertial sensors attached to the wrist or arm [192] can be employed. However, these approaches are only capable of tracking the general position and orientation of the hands, rather than the position of individual fingers. This constrains communicative gestures and therefore limits the naturalness of the interaction. This limitation can be addressed by using data gloves, which enable finger tracking through different types of sensors, such as strain gauges [256, p. 297]. While they enable accurate tracking, the setup process can be time-consuming and may lead to reduced comfort [256, pp. 530-531]. Another approach involves myoelectric sensors that detect hand gestures based on biopotential changes in muscle tissue [192]. However, these sensors are primarily suited for recognizing static and dynamic hand gestures, rather than capturing detailed manipulative hand movements [153].

*Optical tracking methods* rely on cameras or other vision-based sensors to detect and interpret hand movements using computer vision methods. The employed cameras include RGB, depth, time-of-flight, thermal, infrared, and stereo cameras [242]. These systems enable "come as you are" interaction, as they do not require users to wear or hold any devices [256, p. 524].

#### 2.1.2.3 *Multimodality*

Interaction can be facilitated through multimodal techniques, which combine multiple input modalities to enhance expressiveness and flexibility [47, p. 268]. The key idea of multimodality is to overcome the weaknesses of individual input methods by combining them with the strengths of others [286, p. 305]. This is especially important in complex environments, requiring interaction across different tasks [243]. Hand gestures are inherently spatial and thus suited for spatial referencing. However, they might be less suited for complex system control. While semaphoric gestures can be used to create a "gesture alphabet", assigning different functions to distinct, characteristic gestures [256, p. 498], only a limited number of gestures should be used to ensure they are easy to remember and robustly detectable [256, p. 524]. Therefore, it can be useful to supplement gesture input with additional modalities. For example, voice, head, or foot gestures, or button inputs can be employed

to handle more descriptive tasks, such as parameter settings or issuing specific commands [286, pp. 304-306]. Furthermore, secondary modalities can function as *clutching mechanisms* to activate or deactivate gesture recognition, helping to prevent unintended gestures from triggering actions [76].

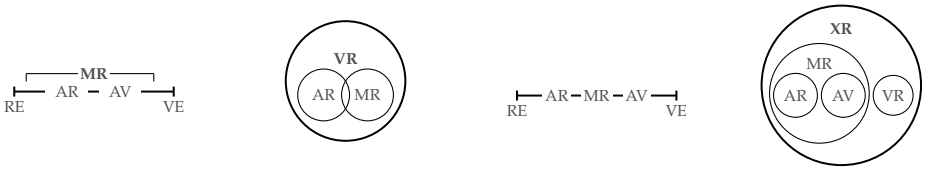
### 2.1.3 *Extended Reality*

As previously mentioned, XR holds potential by providing additional information in the robotic workspace and supporting rapid HRI prototyping [311]. To provide a technical foundation for its application in this context, this section introduces key aspects of XR, beginning with an overview of XR taxonomy, followed by common display techniques.

#### 2.1.3.1 *Taxonomy*

HRI can be extended beyond interactions in purely physical environments by incorporating virtual elements or by creating fully virtual replicas of the environment [311, 351]. Systems that modify or extend reality are commonly categorized using the terms Augmented Reality (AR), Augmented Virtuality (AV), Virtual Reality (VR), Mixed Reality (MR), and XR. Following Rauschnabel et al. [269], there are four main approaches to defining and grouping these terms (see Figure 2.3):

- *MR-dominant view:* The widely adopted taxonomy by Milgram et al. [219] introduces a continuum between Real Environments (REs) ("consisting solely of real objects") and Virtual Environments (VEs) ("consisting solely of virtual objects"). Any environment combining both real and virtual content is considered MR, including AR and AV [307].
- *VR-dominant view:* This perspective considers VR as the overarching category, with AR and MR viewed as subsets or intermediate forms [25].
- *MR-centered view:* In contrast to the MR-dominant view, this approach positions MR between AR and AV, emphasizing the conceptual spectrum between what is "real" and what is "possible" [101].
- *XR-view:* In this view, XR is introduced as the umbrella term encompassing AR and VR [80, pp. 5-6]. It may also incorporate the MR-dominant perspective, in which case XR includes VR and MR, with MR itself comprising both AR and AV [88, p. 21].




(a) MR-dominant (b) VR-dominant (c) MR-centered (d) XR-view

Figure 2.3: Schematic overview of different views on reality formats (adapted from [88, p. 21, 269]).

This thesis adopts the view proposed by Dörner et al. [88, p. 21], which combines the *XR-view* and the *MR-dominant view* (see Figure 2.3d). Accordingly, XR is used as an umbrella term encompassing VR, MR, AR, and AV. In VR, the user's perception of the real environment is entirely *replaced* by a virtual one, whereas in MR, the real environment is *enriched* with virtual content [88, pp. 10-21]. MR itself is considered a continuum; depending on the relative proportions of real and virtual elements, it can be further distinguished as AR (predominantly real) or AV (predominantly virtual) [220].

#### 2.1.3.2 *Display Techniques*

Since human perception is inherently multimodal, XR output devices can engage different sensory modalities such as visual, auditory, haptic, olfactory, and gustatory senses [47, p. 250]. However, this thesis focuses on visual perception. For visual output, displays are used to present the augmented or virtual world [88, p. 151]. MR displays, which combine real and virtual content, can be categorized into three types [87, pp. 248-249]:

- *Video see-through*: The RE is captured by a video camera, digitally superimposed with virtual elements, and then displayed to the user.
- *Optical see-through*: The user perceives the RE directly through a semi-transparent display, with optically overlaid virtual content.
- *Spatial projection*: The virtual content is projected directly onto surfaces in the RE, augmenting them without the need for explicit screens.

Based on these principles, different display technologies can be employed. These can be classified according to their distance from the eye, including head-attached, handheld, stationary, and projected displays [286, p. 59].

HEAD-ATTACHED DISPLAYS There are three main types of displays worn on the user's head: retinal displays, Head-Mounted Displays (HMDs), and head-mounted projectors. Retinal displays project images directly onto the retina of the eye. HMDs use screens placed in front of the user's eyes, while head-mounted projectors use miniature projectors to cast images onto surfaces in the RE [28, pp. 72-73]. The most common head-attached displays are HMDs [286, p. 59]. HMDs can be either semi-transparent, enabling optical see-through, or fully enclosed, as in VR headsets. By integrating cameras whose images are shown on the displays of fully enclosed HMDs, video see-through can be enabled [319, p. 23].

HANDHELD DISPLAYS Video see-through can also be enabled by handheld displays such as tablet computers and smartphones. By utilizing their rear-facing cameras, the RE can be captured and displayed. These cameras can also be used for optical tracking, which, in combination with integrated sensors, allows the position of the device to be determined [87, p. 271]. However, due to the different viewpoints of the user and the camera, the displayed image is not perspectively correct for the user [88, p. 307], and the displays are relatively small [57].

STATIONARY DISPLAYS Stationary displays consist of one or more large monitors or projection screens that can be oriented vertically or horizontally [88, p. 173]. The user's head is usually tracked to enable perspectively correct rendering. Stationary VR displays may consist of screens or projections arranged in different configurations, such as L-shapes, curved screens, or CAVE systems [78], to display the VE [87, pp. 132-133]. In MR, stationary displays are either combined with cameras to enable video see-through or virtual mirrors, or they consist of semi-transparent displays, such as Fogscreens, for optical see-through [286, pp. 72-77].

PROJECTED DISPLAYS Projectors can be used to enable MR without explicit displays by projecting directly onto the RE [28, p. 87]. The goal is to illuminate real objects to alter their perceived appearance [88, pp. 308-309]. This requires knowledge of the position and orientation of the projector relative to the projection surface, as well as models of the projected objects to ensure accurate alignment. Projected displays can be combined with active shutter glasses to enable 3D virtual objects [286, p. 79].

### 2.1.4 *Tracking, Calibration, and Registration*

To enable HRI through hand gestures and to integrate XR into robotic workspaces, three key concepts are essential: tracking, calibration, and registration [286, pp. 87-88]. *Tracking* refers to the continuous, dynamic sensing of objects within the workspace to determine their pose. In the context of HRI, tracking is essential to determine the position and orientation of the user, their hands for interaction, their heads for correct XR display, and the robotic end effector for robot control. *Calibration* involves the offline adjustment of measurements by comparing readings from two different devices [286, p. 87]. Calibration is required to ensure that tracking systems deliver measurements on a known scale. For example, when different tracking systems with distinct reference points are used to track the same object or the robot itself, calibration is necessary to determine the required offsets [37]. *Registration* describes the "alignment of coordinate systems" [286, p. 87]. In the context of HRI, this alignment is required to establish a common reference frame between sensors used for interaction, the XR coordinate frame, and the robotic workspace.

The following subsections will detail tracking, calibration, and registration approaches relevant to this thesis. For more comprehensive discussions on additional tracking methods or calibration techniques like camera calibration, refer to the works of Schmalstieg and Höllerer [286, pp. 85-120], Tönnis [319, pp. 43-95], or Doerner et al. [88, pp. 113-135].

#### 2.1.4.1 *Tracking*

In the context of HRI and XR, tracking refers to the three-dimensional tracking of objects, including all six DOFs, which consist of an object's position and orientation. Different tracking systems exist to determine an object's pose, each based on different physical principles [319, p. 44].

OPTICAL TRACKING In optical tracking, a camera is used to determine the relative pose of an object with respect to the camera [87, p. 104]. These methods can be divided into *marker-based* approaches, which use easily recognizable markers in the camera image, and *markerless* approaches. Additionally, tracking systems can be categorized as *outside-in*, where cameras are positioned externally and observe the object, or *inside-out*, where the camera is mounted on the object itself and captures the surrounding environment.

*Marker-based* tracking introduces artificial landmarks, also known as fiducials, in the environment [312]. These approaches primarily rely on either RGB or infrared cameras [87, pp. 104–106]. When using RGB cameras, so-called *image markers* are employed. These are two-dimensional patterns without rotational symmetry [319, pp. 47–48]. The position and orientation of these typically rectangular markers can be derived by extracting edges from the camera image through image processing. When using infrared cameras, the markers are either passive retro-reflective markers or active markers that emit light. These markers appear as bright points in the video stream. To determine the pose of a marker, triangulation is performed using two or more infrared cameras [87, p. 106]. While a single marker is sufficient to determine the 3D position, a so-called *marker shield*, consisting of at least three markers, is required to estimate the full 6-DOF pose.

*Markerless* approaches omit the use of artificial markers by relying on natural features, such as edges and interest points, extracted from the environment [286, p. 112]. These features can be extracted from RGB images, or RGB-D cameras can be employed, projecting infrared patterns or using time-of-flight technology for depth estimation [87, p. 107].

MECHANICAL TRACKING Mechanical tracking determines the pose through rigid mechanical connections attached to the tracked object. Using forward kinematics (see Section 2.1.1.1). The position at the end of an articulated arm can be derived from the known lengths of the limbs and angular measurements [286, p. 96]. The required angles can be measured using gears, potentiometers, and bend sensors [38, p. 47]. This approach is not only used in robotics, but also for interaction using mechanical input devices [87, p. 110] or to track body movements through wearable sensors [38, p. 47].

INERTIAL TRACKING In contrast to optical and mechanical tracking, where an object's pose is measured relative to a fixed origin, inertial tracking measures relative poses with respect to a starting position [319, pp. 52–53]. This is achieved using inertial sensors that measure linear and angular acceleration, deriving the position and orientation through a twofold integration [88, pp. 114–115]. By incorporating three linear and three angular acceleration sensors arranged orthogonally in so-called Inertial Measurement Units (IMUs), the six DOFs pose can be tracked [38, p. 35].

#### 2.1.4.2 *Calibration*

The aforementioned tracking techniques enable the estimation of the pose of objects or markers. For this tracking to be spatially accurate, several components must be calibrated. These include calibration of the tracking system, the displays, and the tracked objects [286, p. 194].

In optical tracking, this involves calibrating the camera by determining intrinsic parameters and correcting lens distortions [81, pp. 668-675]. For mechanical tracking, calibration involves aligning sensor readings with actual end effector positions [274]. In the case of inertial tracking, calibration assesses systematic errors such as bias, scale factor, and misalignment [253].

Additionally, to ensure correct spatial rendering of the tracking data, the displays must be calibrated, including eye and head calibration [286, pp. 183-187]. Finally, the tracked objects themselves often require calibration. This is especially important when the tracking system, for example through markers, is at a static offset from the actual point of interest, such as a tool tip.

TIP CALIBRATION Tip calibration determines the position of a rigidly attached point relative to a tracking target [226, pp. 213-214]. While this method is commonly used for calibrating surgical or industrial tools, it is also relevant in HRI, as it enables accurate tracking of a robotic end effector.

When using optical tracking, markers can be attached to the end effector. However, placing them directly at the Tool-Center Point (TCP) is often not feasible. Since the TCP serves as the reference point for robot control, its accurate localization is crucial. Therefore, the static offset between the tracking marker and the TCP must be determined. This can be achieved using a method known as *pivot calibration*. In pivot calibration, a marker is rotated around a fixed point, resulting in a spherical trajectory [319, p. 64] (see Figure 2.4). The offset $t^p = [x_p, y_p, z_p]^T$ between the marker and the tip can then be estimated using approaches based on the least-squares method [363]. Common techniques include sphere fitting [13], an algebraic one-step method [203], and an algebraic two-step formulation [166].

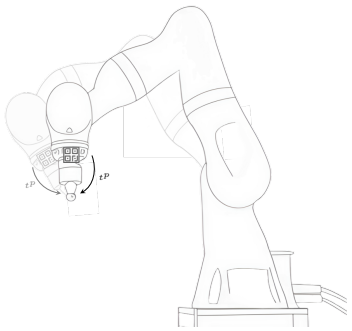


Figure 2.4: Visualization of tip calibration, with a marker attached to the robot. The robot rotates around the TCP, and the marker's position is recorded during the motion.

#### 2.1.4.3 *Registration*

Calibrated tracking enables the determination of the pose of physical objects within their own local coordinate frame. However, tracking information is often required relative to other coordinate systems as well. For example, to map hand motion accurately to robot movement, the relative transformation between the hand tracking sensor and the robot coordinate frame must be known. Similarly, when XR is used to display information directly within the robotic workspace, registration between XR and the robotic workspace is necessary to ensure the correct placement of virtual content [286, p. 87].

The transformation between different coordinate systems can be derived through *registration*. Following Bernhardt et al. [32], there are four main approaches: manual, point-based, surface-based, and volume-based registration. Since volume-based registration requires volumetric data, it is primarily employed in medical image registration using image voxel intensities [276] and will therefore not be considered further. This thesis will instead focus on the three most common registration methods for XR registration: manual, point-based, and surface-based registration [46, p. 173].

MANUAL REGISTRATION In manual registration, the alignment between the physical and virtual coordinate spaces is achieved interactively by manually adjusting up to nine DOFs [32]. The number of DOFs that need to be adjusted can be reduced by combining manual registration with point-based approaches, for example, obtaining the translation through an interactively selected landmark and then only adjusting the rotation manually [255].

POINT-BASED REGISTRATION In point-based registration, also known as *paired-point registration*, a set of corresponding 3D points $p_i$ and $p'_i$ in both coordinate systems is used to determine the transformation between the coordinate spaces [362, pp. 161-163]. By recording three or more non-collinear points, the translation $t$ and rotation $R$ between the coordinate frames can be calculated. To this end, the least-squares method can be employed, which minimizes the square error $\Sigma^2$ described in Equation 2.1 [23].

$$\Sigma^2 = \sum_{i=1}^{N} \left\| p'_i - (Rp_i + t) \right\|^2 \tag{2.1}$$

SURFACE-BASED REGISTRATION Instead of utilizing corresponding points for alignment, surface-based registration relies on surfaces, which are often represented as point clouds [276, p. 141]. This type of registration is therefore often referred to as *point cloud registration*. Similar to point-based registration, the transformation is calculated by minimizing the distance between the point clouds. However, since point correspondences are not explicitly provided, they must be estimated as part of the registration process.

One commonly employed approach is the Iterative Closest Point Algorithm (ICP) introduced by Besl and McKay [33], which is an iterative, two-stage algorithm. In the first stage, for each point in one point set, the closest

point in the other set is identified to form corresponding pairs. The second stage computes the transformation that minimizes the distance between those points, for example, using the point-based technique discussed earlier. In the next iteration, the point correspondences are then updated, and a new transformation is calculated. This process is repeated until convergence [139]. Since ICP converges only locally, an appropriate initial transformation is necessary to prevent convergence to suboptimal local minima [362, p. 166]. This initial estimate can be obtained through manual or point-based registration methods.

#### 2.1.4.4 *Noise Filtering and Outlier Rejection*

During tracking, calibration, and registration, measurement errors can lead to noisy or inconsistent tracking data, which can negatively affect the accuracy and stability of pose estimation. To address this, methods that reduce the influence of noise and reject erroneous data points can be applied. This section presents two widely used approaches for this purpose: the *Kalman filter* for recursive state estimation, and the Random Sample Consensus (RANSAC) algorithm for robust model fitting and outlier rejection.

KALMAN FILTER To address jitter in sensor data, filtering techniques can be applied to smooth the signal. One commonly used method is the Kalman filter [162]. This filter operates under the assumption that the error follows a normal distribution and that the error can be minimized by computing a linear combination of the current state and the measurement [286, p. 193]. The Kalman filter consists of two main steps: a time update and a measurement update [340]. In the time update, the current state estimate and its associated uncertainty are used to predict the state at the next time step. In the measurement update, new sensor measurements are incorporated to correct the prediction, thereby improving the overall estimate.

RANSAC The RANSAC algorithm, proposed by Fischler and Bolles [105], addresses the challenge of large numbers of outliers in data by generating candidate solutions using the smallest possible subset of data points required to estimate the model parameters. In each iteration, RANSAC randomly selects a minimal subset of data points, computes a model based on them, and then evaluates how many points from the full dataset are consistent with this model. These inliers form the *consensus set*. If the size of the consensus set exceeds a predefined threshold, the model parameters are re-estimated using all inliers; otherwise, the process is repeated with a new random subset [84].

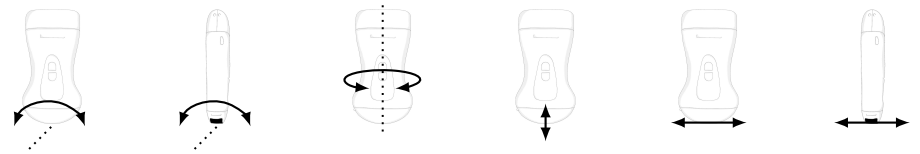
RANSAC can be effectively applied to problems such as *pivot calibration* or *point-based registration*. Both are based on least-squares estimation and are therefore highly sensitive to erroneous data points. Using RANSAC in these contexts can improve robustness by reducing the influence of outliers [363].

### 2.1.5 *Use Case: Robotic Ultrasound*

As outlined in Section 1.2.4, robotic ultrasound represents a meaningful use case for investigating HRI approaches, particularly due to its contact-intensive nature. This section, therefore, provides the necessary foundations on ultrasound imaging and its integration with robotic systems.

#### 2.1.5.1 *Ultrasound Imaging*

Medical Ultrasound (US) imaging, also referred to as *sonography* or *echography*, uses the physical properties of sound waves in tissue for diagnostic and therapeutic imaging purposes [123, p. 97]. By generating and emitting US waves in different directions and receiving the returning echoes, 2D slice images of internal structures can be obtained [10]. Since different tissues have characteristic acoustic impedances (i.e., resistance to the propagation of sound), interactions such as absorption, reflection, and dispersion result in distinct attenuations of the US waves [260]. While US provides good soft tissue contrast, it is limited in regions where bone or air obstructs wave propagation [123, p. 97]. Areas that allow imaging of the region of interest, as there is no obstruction through bones or gas, are termed *acoustic window*. Effective US imaging requires different probe movements, including rocking (in-plane rotation), fanning (out-of-plane rotation), rotating, compressing, sliding, and sweeping [51, pp. 38-41](see Figure 2.5).



(a) Rocking (b) Fanning (c) Rotating (d) Pressing (e) Sliding (f) Sweeping

Figure 2.5: Overview of different motions required for US imaging.

US imaging is widely used in clinical practice as it relies on non-ionizing radiation, is less expensive than other imaging modalities such as Computed Tomography (CT) or Magnetic Resonance Imaging, provides real-time imaging, and is portable [302]. Its applications span different medical disciplines, including abdominal, breast, gynecological, and cardiovascular imaging [260].

However, high-quality US imaging depends heavily on the skill and experience of the operator, as image quality is strongly influenced by correct probe placement [296] and the application of appropriate contact force [53, 117]. This results in variability in image quality, even among experienced practitioners [172]. Additionally, maintaining steady probe pressure while interpreting images and adjusting settings imposes not only a high cognitive workload but also a high physical workload, contributing to musculoskeletal disorders [98, 129].

#### 2.1.5.2 *Robotic Ultrasound*

To address these challenges, robotic US has emerged as a promising solution, where a US probe is attached to the end effector of a robot manipulator [156]. Robotic US has been explored for applications such as diagnostic imaging, needle-based interventions, and intraoperative imaging [257]. These systems can be categorized based on their LORA.

Robotic US systems at low LORA are often referred to as *teleoperation*, where the probe is controlled manually by a remote or co-located practitioner [133]. This approach enables remote imaging in areas lacking trained experts and reduces the physical strain associated with holding the probe in manual US.

*Collaborative assistance*, corresponding to LORA levels 3 and 4, aims to support clinicians by enabling faster, more precise, and reproducible imaging through shared autonomy [133]. This can include partial automation features such as motion compensation or force regulation [188, 297].

At higher LORA, *autonomous* robotic US systems are capable of autonomous task planning and execution. Key enabling technologies include force control, orientation optimization, and path planning, supporting features such as motion awareness, deformation compensation, and visual servoing [156].

As described in Section 2.1.1.2, this thesis focuses on lower LORA, where the operator directly controls the robot. While progress toward autonomous robotic US systems is being made [133], manual control remains essential to initiate such systems [64, 196, 297, 368], and real-time adjustments are required to incorporate the operator's expertise during interventions, enhancing adaptability and safety [179, 358].

## 2.2 METHODOLOGY PRELIMINARIES

In HRI research, two main methodological approaches are commonly distinguished: *human-centered* and *robot-centered* [279]. The human-centered approach emphasizes user studies to evaluate the design and usability (effectiveness, efficiency, and satisfaction) of HRI interfaces. In contrast, the robot-centered approach focuses on developing control methods and implementation technologies to enhance robot performance [168]. While the term *human-centered* is often associated with Human-Centered Design, which ISO 9241-210:2019 defines as an "approach to systems design and development that aims to make interactive systems more usable by focusing on the use of the system and applying human factors, ergonomics, and usability knowledge and techniques" [150], its use in this thesis refers to a broader research orientation rather than a specific design process. Although technical development plays a key role in this work, it is driven by the goal of optimizing interfaces for users, reducing their workload, rather than evaluating isolated robot performance metrics. As Rea et al. observe, HRI research often prioritizes technical solutions, yet human-centered methods remain crucial for enabling operators to "take full advantage of the hardware and algorithms being developed" [270].

To support this perspective, the following section outlines the methodological preliminaries of this work. It introduces the evaluation framework used to assess human-centered efficiency and explains the statistical methods employed to analyze the significance and strength of observed effects.

### 2.2.1 *Evaluation Measures*

As outlined in the RQs formulated in Section 1.2, the goal of this thesis is to enable *efficient* HRI through human-centered investigations. To achieve this, the concept of efficiency must first be clearly defined. Adamides et al. [11] state that "In HRI, efficiency is measured in terms of the time required to complete a task; effectiveness is measured in terms of how well a task is completed". However, this definition reduces efficiency solely to one measure, which is insufficient for capturing the broader implications of the HRI methods under investigation [34]. According to the ISO 9000:2005 standard, efficiency is defined as the "relationship between the result achieved and the resources used" [149]. This more comprehensive definition implies that multiple measures must be considered to assess both the *results* and the *resources* involved. The results can be quantified using objective metrics, such as binary measures of success (e.g., task success) or more nuanced indicators like accuracy [318, p. 65]. The resources used can include both objective measures, such as task completion time, and subjective measures, such as perceived mental or physical workload. To operationalize this definition of efficiency, the following subsections will describe relevant objective and subjective indicators that can be used to assess the *results achieved* and the *resources used* in HRI.

#### 2.2.1.1 *Objective Measures*

Objective measures provide qualitative or quantitative data that can be used to assess the *performance* of methods in HRI [20]. Since performance is influenced by the human factors of the system, such measures are essential for a human-centered evaluation. This section presents two commonly used objective quantitative indicators: task completion time and accuracy. In addition, as this thesis investigates registration methods, measures relevant for assessing registration accuracy are also introduced.

DURATION Task Completion Time (TCT), also referred to as *time on task*, is typically measured as the time elapsed between the start and completion of a task [318, pp. 74-75]. It is one of the most commonly evaluated factors for assessing performance in HRI [20]. Other temporal measures include *time in mode*, which quantifies the duration of specific functions, actions, or pauses, and *time until event*, which measures durations such as reaction time [144].

ACCURACY Accuracy is a measure used to quantify the error with which users complete a task [144]. This can include error rates, spatial accuracy, and precision. *Error rates* quantify the number of incorrect actions during task

completion and are applicable when user actions can be clearly categorized as correct or incorrect. *Spatial accuracy* is used to measure the distance from a predefined target, including translation errors (e.g., Euclidean distance) and rotation errors (e.g., the smallest angle between the actual and target orientation) [210]. In HRI, such measures are particularly relevant for steering or positioning tasks. *Precision* describes the ratio of correct information to all retrieved information and is, therefore, more relevant in HRI for information retrieval tasks (e.g., sensor data) than as a direct measure of user performance.

Error rates, spatial accuracy, and precision primarily quantify the outcome or final result of a task. However, in many HRI scenarios, it is also important to monitor relevant data continuously. For example, tracking the distance to a desired trajectory [20] or observing contact force [82] offers insights into user behavior and interaction quality beyond final task outcomes. These continuous measures can be quantified by sampling data at regular intervals and reporting average values or maximum deviations [27, 82].

One important factor when considering both time and accuracy is the *time-accuracy trade-off* (also referred to as the speed-accuracy trade-off), which describes how these two measures influence each other [173, p.24]. In practice, this means that accuracy can often be improved by spending more time on a task, while prioritizing speed may lead to decreased accuracy. To account for this trade-off, it is important to include both measures to gain a more comprehensive understanding of user performance [201].

REGISTRATION MEASURES To assess registration quality, different accuracy metrics have been introduced, including the Fiducial Localization Error (FLE), Target Registration Error (TRE), and Fiducial Registration Error (FRE) [124, pp. 120-124]. The FLE represents the distance between the true positions and the measured positions at the registration points (also referred to as *fiducials*). While the FLE, caused by sensor noise, calibration errors, or tracking inaccuracies, contributes to registration errors, a more relevant metric is the TRE. The TRE utilizes *target points* that were not used for registration, quantifying the distance between corresponding points after registration. It is defined as the distance between a point $p$ after applying the transformation $T$ and its corresponding point $p'$:

$$\text{TRE} = ||T(p) - p'|| \tag{2.2}$$

In contrast, the FRE describes the distance between corresponding *registration points* after registration. While the FRE is easy to measure, it has been shown to be uncorrelated with the TRE and therefore is not a reliable predictor of registration accuracy [108].

#### 2.2.1.2 *Subjective Measures*

Subjective measures can be described as measures "that rely on human experience, judgment, perception, or cognition" [345, p. 33]. They may take the form of qualitative data, such as descriptions or quotes, or quantitative

data, such as numerical ratings [300, p. 327]. This subsection provides an overview of subjective measures relevant to this thesis.

INTERVIEWS According to Fontana and Frey [109], there are four main types of interviews: open-ended, structured, semi-structured, and group interviews. While all the first three can be employed for user studies with individuals, the interviewing approach selected for a study depends on the purpose of the interview. Unstructured interviews are exploratory, allowing for the generation of rich data, while structured interviews enable quicker assessment of specific questions [300, pp. 285-286]. Semi-structured interviews are the most common approach in HRI studies [326]. Especially in hypothesis-driven studies, semi-structured interviews are employed to provide explanatory data that supports the recorded quantitative results. Therefore, in this thesis, semi-structured interviews are employed, combining closed and open questions.

QUESTIONNAIRES Similar to interviews, questionnaires can include both closed and open questions [300, p. 194-197]. While they can be used to acquire data such as participant demographics, rating scales can be employed to compare specific answers across respondents. Likert items are commonly used here, providing a gradual scale with verbal anchors that allow for capturing respondents' attitudes towards a topic.

PERCEIVED WORKLOAD The NASA Task Load Index (TLX) is a questionnaire designed to assess workload across six dimensions, as presented by Hart and Staveland [131]. These dimensions encompass mental demand, physical demand, temporal demand, performance, effort, and frustration. The original questionnaire consists of two parts: the first involves rating the different dimensions on a scale from 0 to 20, while the second applies a weighting scheme. However, the most common modification eliminates the weighting part, making the questionnaire simpler to apply [130]. This modified version, known as the raw TLX, will be used in this thesis, as it allows for an initial assessment of perceived workload.

PHYSICAL EXERTION To assess physical exertion, the Borg CR10 scale [43] is commonly used. This scale measures perceived exertion through values ranging from 0 to 10, with 0 representing "Nothing at All" and 10 indicating "Very, Very Hard (Maximal)". The Borg CR10 has demonstrated strong correlations with EMG-based metrics [247, 317]. Additionally, it has been shown that Borg CR10 scores can be more reliable at low levels of physical exertion, where EMG measurements may become unreliable [233].

PERCEIVED TASK EASE The Single Ease Question (SEQ) assesses how difficult or easy a task is perceived by asking participants to rate the statement "Overall, this task was:" on a single seven-point scale ranging from "Very Easy" to "Very Difficult" [284, p. 186]. Compared to other single-item questionnaires, it has been shown to be more sensitive than the ratio-scaled Usability Magni-

tude Estimation and easier to administer than the interval-scaled Subjective Mental Effort Questionnaire [283].

HUMAN-ROBOT TRUST Schaefer [285] introduced the *Trust Perception Scale-HRI*, a 40-item scale using 10 equidistant percentage-based response options to assess how often a robot is perceived to exhibit certain characteristics (e.g., "What percentage of the time will this robot (be)..."). To enable quicker assessment, especially when comparing multiple conditions, a 14-item short form was proposed. Eilers et al. [92] modified the questionnaire by adopting a 5-point Likert item to rate the items, making it easier to administer than the original percentage-based format. This thesis applies the 14-item subscale with responses on a 5-point Likert item to enable efficient assessment.

### 2.2.2 *Statistical Measures of Evidence*

To evaluate the quantitative measures described above, different statistical measures of evidence can be employed. These measures support the assessment of the effects of *independent variables*, which define the experimental *factors* under investigation, on *dependent variables*, which correspond to the measured outcomes. To this end, statistical hypotheses are formulated. In a two-sided test, the Null Hypothesis ($H_0$) assumes no effect, while the Alternative Hypothesis ($H_1$) assumes the presence of an effect [104, p. 28].

To assess the strength and nature of the evidence collected, several statistical indicators are commonly used. Among the most important are the *p-value*, which evaluates the likelihood of the observed data under the null hypothesis; the *effect size*, which quantifies the magnitude of an observed effect; and the *Bayes factor*, which offers a probabilistic comparison of competing hypotheses. The following subsections describe each of these measures in detail.

#### 2.2.2.1 *p-Value*

To determine whether observed effects in user studies are statistically significant, *null hypothesis significance testing* is employed. As statistical methods cannot directly prove $H_1$, the focus lies on testing whether $H_0$ can be rejected [104, p. 28]. If the data provide sufficient evidence to reject $H_0$, support is inferred for $H_1$. Central to this approach is the *p-value*, which indicates the probability of obtaining a result as extreme as the observed one, assuming $H_0$ is true [342]. A *level of significance* (typically 0.05 [106, p. 44]) serves as threshold, with p-values lower suggesting statistically significant effects [342].

The choice of statistical method for obtaining p-values depends on the experimental design, the types of variables involved, and whether certain assumptions are met. A key distinction lies in the study design: in within-subjects designs, repeated measurements are acquired from the same participants under multiple conditions. In contrast, between-subjects designs use different participants for each condition [79, pp. 114-115]. As this thesis focuses on within-subjects evaluations, tests for repeated measures data are primarily relevant, with further distinctions illustrated in Figure 2.6.

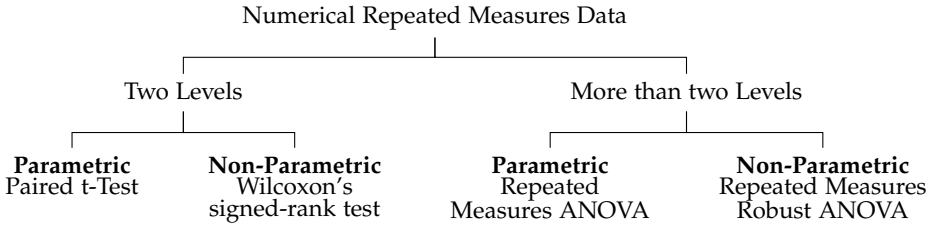



Figure 2.6: Overview of statistical tests for null hypothesis significance testing for numerical repeated measures data (adapted from [134]).

One such distinction concerns the number of factor levels, as comparing two levels differs from comparing more than two levels. A final consideration is whether the assumptions for the tests are satisfied. The main assumptions are normality (e.g., assessed using the Shapiro-Wilk test) and homogeneity of variances (e.g., tested with Levene's test) [134]. When these are met, parametric tests are used; otherwise, non-parametric alternatives are applied.

#### 2.2.2.2 *Effect Size*

While the p-value indicates whether an effect is statistically significant, it does not convey whether the effect is practically meaningful [104, p. 57]. This limitation can be addressed by reporting the effect size, which quantifies the magnitude of an effect and measures the deviation from the data or distribution that would be expected under the assumption of $H_0$ [342]. The type of effect size reported depends on the statistical test employed: for t-tests, it typically is Cohen's $d$; for ANOVAs, $\eta^2$; for robust ANOVAs, $\delta_t$; for the Friedman test, Kendall's $W$; and for Wilcoxon tests, the effect size $r$. Verbal anchors for interpreting these effect sizes were introduced in the behavioral sciences [70]. The corresponding thresholds are summarized in Table 2.2.

Table 2.2: Evidence categories for effect sizes $d$, $\eta^2$, $\delta_t$, $W$, and $r$ [16, 70].

| | d | $\eta^2$ | $\delta_t$ | $W$ | $r$ |
|---|---|---|---|---|---|
| Small | 0.2-0.5 | 0.01-0.06 | \|0.2\| - \|0.5\| | 0.1-0.3 | 0.1-0.3 |
| Medium | 0.5-0.8 | 0.06-0.14 | \|0.5\| - \|0.8\| | 0.3-0.5 | 0.3-0.5 |
| Large | >0.8 | >0.14 | > \|0.8\| | >0.5 | >0.5 |

#### 2.2.2.3 *Bayes Factor*

One further limitation of null hypothesis significance testing is that it does not permit strong conclusions in favor of the null hypothesis [69]. This limitation can be overcome by the *Bayes Factor Analysis*, which quantifies the evidence supporting both the alternative and the null hypothesis [287]. The Bayes Factor (BF) represents the ratio of the marginal likelihoods of two hypotheses, with the marginal likelihood describing how well the data support each hypothesis [142, p. 60]. For the observed data $y$ and the marginal likelihood $m(y|H_t)$, the BF is thus defined as:

$$BF_{10} = \frac{m(y|H_1)}{m(y|H_0)} \tag{2.3}$$

BF Analysis typically yields values ranging from 0.01 to 100, which are commonly interpreted using descriptive categories [184, p. 105] (see Figure 2.7).

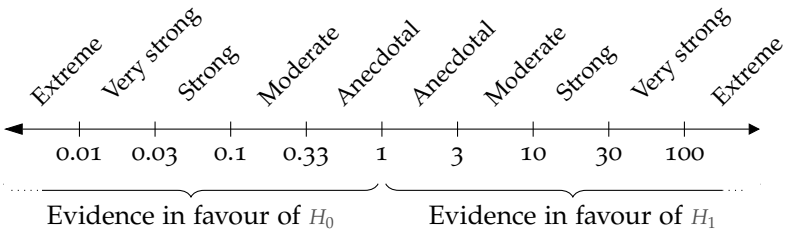



Figure 2.7: Descriptive classification for interpreting BF based on Lee and Wagenmaker [184, p. 105] (adapted from [262]).

## 2.3 IMPLEMENTATION DETAILS

This section outlines the approaches for robot control, gesture-based interaction, XR, registration, and robotic US implemented in this thesis. Since the underlying implementation framework is consistent across the projects, it is treated as background information. However, some studies specifically examine aspects of this basic setup, introducing variations. When such variations occur, they will be explicitly noted in the corresponding sections.

### 2.3.1 *Robot Control*

For the investigations, a robot manipulator, namely the *KUKA LBR iiwa* (KUKA AG, Germany), was employed. This is a lightweight robot manipulator with seven axes, equipped with integrated force-torque sensors, and specifically designed as a collaborative robot (also commonly referred to as a *cobot*). The KUKA LBR iiwa was chosen as a representative robot, as it is commonly used in industrial [213] and medical applications [156]. It was mounted on an elevated frame, with a desk positioned in front of it to provide a working surface (see Figure 2.8). The robot control was implemented using *KUKA Sunrise.OS* and the *KUKA Sunrise.SmartServo* library. Movement commands were issued as relative Cartesian positions, with inverse kinematics and trajectory planning handled by *KUKA Sunrise.OS*. The motion was executed using impedance control with a stiffness of 800 N/m, enabling precise robot movement when no external forces are present. For force measurements, the robot's force/torque sensors are utilized.

To increase safety during interaction, a maximum safety threshold of $F_{\text{max}} = 30\,\text{N}$ was applied. When the external forces exceeded $F_{\text{max}}$, the interaction was stopped, and the end effector was moved 10 mm in the opposite direction of the force. Additionally, to avoid singularities and prevent the robot from reaching its joint limits, Cartesian boundaries were applied to

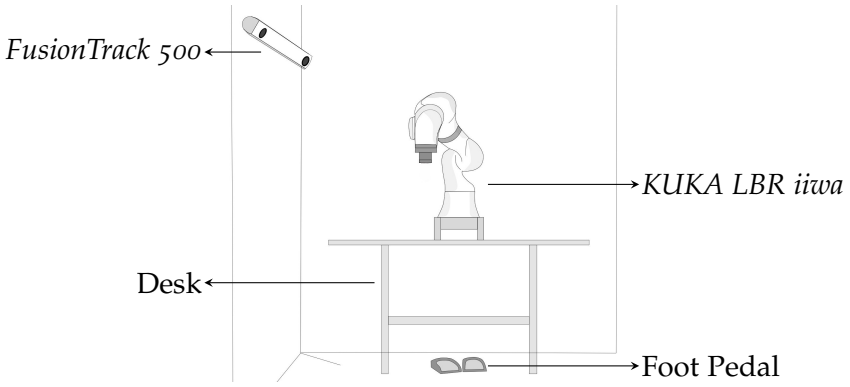



Figure 2.8: Setup of robot for the experiments conducted in this thesis.

reduce the effective workspace. Furthermore, a maximum linear velocity of 150 mm/s and an angular velocity of 0.75 rad/s were set to increase safety.

The robotic end effector was equipped with the *Media Flange Touch Electrical* (KUKA AG, Germany), which includes two programmable buttons and an LED strip. The LED strip was used to indicate the control status of the robot: green light signaled active interaction, red light indicated that $F_{\text{max}}$ was exceeded, and blue light marked the reaching of workspace boundaries.

### 2.3.2 *Gesture-Based Interaction*

In this thesis, hand gestures are employed for robot control, as they have been shown to be a natural [373], intuitive [183, 199, 356], flexible [61], and fast [183, 282] interaction method. For robot control using hand gestures, different types of input are required (see Figure 2.9). In this work, the position and orientation of the end effector are controlled through manipulative gestures. Using position control, relative hand movements are mapped onto corresponding relative end effector motions. In the initial studies of this thesis, an unscaled mapping was used, based on previous work [183, 9]. Following an evaluation of different mapping strategies, later studies adopted a 1:1 mapping for translation and a 2:1 scaled mapping for rotation.

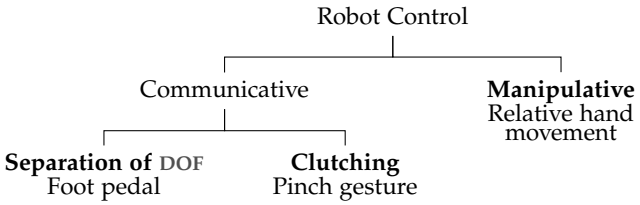



Figure 2.9: Employed interaction modalities for robot control.

As semaphoric gestures are well-suited for clutching when using manipulative hand gestures [8], a pinch gesture is implemented to pause interaction and prevent unintended input. Following prior findings that holding a gesture is preferable to toggling the interaction state for clutching [294], the system activates interaction while the pinch gesture is held.

Furthermore, prior research has demonstrated that separating the DOFs manipulated during interaction can improve performance by reducing unintended transformations [216]. Therefore, the system supports manipulation of rotation and translation either independently or simultaneously. A foot pedal (*Docooler* (Tomptop Technology Co., China)) is used to switch between rotation, translation, and full 6-DOF control. The foot pedal was chosen based on previous work indicating that semaphoric gestures may be perceived as too similar, and that voice commands can reduce perceived responsiveness [9]. Feedback about the active mode is provided in AR by highlighting a button in the upper center of the Field of View (FOV) (see Figure 2.10)

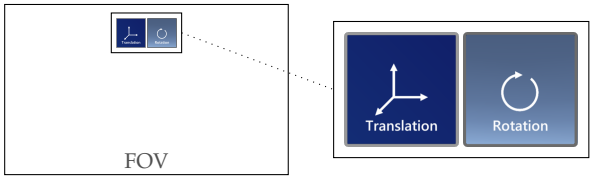



Figure 2.10: Visual feedback on active DOFs during interaction. Active DOFs have a dark blue background, while inactive DOFs appear transparent.

Optical hand tracking has been shown to offer greater usability and more natural, unconstrained interaction compared to wearable sensors such as data gloves or controllers [83, 354]. Additionally, optical systems typically provide lower cost, faster setup, and higher three-dimensional tracking accuracy than glove-based approaches [75]. Because optical tracking supports natural interaction by enabling both communicative and manipulative gesture recognition, this approach is employed in this thesis. In the initial projects of this thesis, hand tracking is performed using the *HoloLens 2* (Microsoft Corporation, USA), which utilizes a depth camera based on time-of-flight technology [321]. This device was selected in the early stages, as it supports both hand tracking and XR capabilities, thereby simplifying prototyping and development. In later studies, the *Leap Motion Controller 2* (Ultraleap, USA) is used. This sensor, using two infrared cameras to derive hand pose from raw sensor data [121], was selected for its flexibility in sensor placement.

### 2.3.3 *Extended Reality*

In robotics, XR is commonly implemented through HMDs, handheld devices, spatial screens, displays mounted on the robot, or projectors. These can be head-worn, hand-held, stationary, or robot-mounted [311]. While earlier XR systems primarily relied on standard monitors, the current trend favors the use of HMDs as they support natural interaction and integrate sensors that enable hand and marker tracking [205]. Among these devices, the *HoloLens 2* (Microsoft Corporation, USA) is a widely adopted optical see-through headset [72, 112]. Therefore, the *HoloLens 2* is used for XR in this thesis. Applications for the HoloLens are developed in *Unity* (Unity Software Inc., USA) using the *Mixed Reality Toolkit* (Microsoft Corporation, USA).

### 2.3.4 *Tracking, Calibration, and Registration*

TRACKING For implementing the different prototypes in this thesis, tracking is required for both component registration and the localization of the end effector and task-relevant elements. Specifically, it enables the registration process by allowing the robotic end effector to be tracked in both the robot and the XR coordinate frames. The tracking in the robot coordinate frame is achieved via mechanical tracking and forward kinematics. For tracking in the XR coordinate frame, an image marker is attached to the end effector and tracked using the *Vuforia SDK* (PTC Inc., USA). Since an HMD is used as the XR device, accurate world pose estimation of the marker requires not only marker tracking but also continuous tracking of the HMD itself. On the HoloLens 2, this is accomplished through inside-out tracking using four cameras combined with an IMU [88, pp. 132-133]. To track task elements, for example, to assess successful task completion or to provide ground truth for registration accuracy, an *Atracsys FusionTrack 500* (Atracsys LLM, Switzerland) infrared tracking system is employed.

CALIBRATION For pivot calibration of the marker attached to the robot, the end effector is automatically pivoted around its TCP, and the corresponding marker positions are recorded. To ensure variation in the recorded poses and reduce redundancy, a new pose is only stored when the current marker pose differs from the previous one by more than a predefined threshold. To compute the offset between the marker and the TCP, the algebraic two-step method is employed, as algebraic methods have been shown to provide higher accuracy than sphere fitting [363]. This method relies on the relationship shown in Equation 2.4, which holds for any pair of marker positions $t$ and orientations $R$ during pivoting. The offset $t^p$ between marker and TCP can then be computed by solving the overdetermined linear system described in Equation 2.5.

$$R_i t^p + t_i = R_j t^p + t_j \tag{2.4}$$

$$\begin{bmatrix} R_1 - R_2 \\ \vdots \\ R_{m-1} - R_m \end{bmatrix} \mathbf{t}^p = \begin{bmatrix} \mathbf{t}_2 - \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_m - \mathbf{t}_{m-1} \end{bmatrix} \tag{2.5}$$

To reduce the impact of outliers on the calibration results, the RANSAC algorithm is employed, resampling the recorded points to calculate the calibration.

REGISTRATION Registration is conducted by measuring corresponding points in both the robot and the XR coordinate frames, as described above. However, since the points recorded in the XR coordinate frame follow a *left-handed coordinate* system (as defined by Unity), and the robot points follow a *right-handed coordinate* system (as defined by KUKA Sunrise.OS), the XR points must be transformed prior to registration. This is done by negating the $z$-values to ensure consistency.

The resulting point sets are then aligned using the Kabsch algorithm [23], minimizing the squared error between corresponding points. This algorithm solves for rotation and translation separately and consists of three main steps:

1. Computing the centroids $p$ and $p'$ of both point sets (see Equation 2.6), and recentering them at the origin

$$p = \frac{1}{N}\sum_{i=1}^{N} p_i \quad \text{and} \quad p' = \frac{1}{N}\sum_{i=1}^{N} p'_i \tag{2.6}$$

2. Finding the optimal rotation $R$ that minimizes Equation 2.7, e.g., through Singular Value Decomposition

$$\Sigma^2 = \sum_{i=1}^{N} \left\| p'_i - (Rp_i) \right\|^2 \tag{2.7}$$

3. Determining the translation $t$ by solving Equation 2.8

$$t = p' - Rp \tag{2.8}$$

While solving the registration provides the transformation between coordinate frames, reliable orientation conversion requires accounting for the differing rotation conventions of the XR and robot systems. The robot system uses an intrinsic Z-Y′-X″ rotation order, whereas the XR-system (Unity) applies an extrinsic Z-X-Y order. As a result, Euler angles cannot be directly transferred between the two.

However, since Unity primarily uses quaternions to represent rotations, transformations between coordinate frames can be handled consistently in quaternion space. The conversion from Euler angles to quaternions is straightforward in Unity, as the transformation from axis-angle representation to quaternions is built into the Unity API. Each individual axis rotation (around X, Y, and Z) can be converted into a quaternion, and these quaternions are then multiplied in the specified rotation order to produce the final orientation. To extract Z-Y′-X″ Euler angles from a quaternion, the angles $\phi$ (rotation around the x-axis), $\chi$ (y-axis), and $\psi$ (z-axis) can be calculated using the discriminant $\Delta = q_r q_y - q_x q_z$, as shown in Equation 2.9 [40].

$$\begin{pmatrix} \phi \\ \chi \\ \psi \end{pmatrix} = \begin{pmatrix} \tan^{-1}\left(\dfrac{2(q_r q_z + q_x q_y)}{1 - 2(q_y^2 + q_z^2)}\right) \\ \sin^{-1}(2\Delta) \\ tan^{-1}\left(\dfrac{2(q_r q_x + q_y q_z)}{1 - 2(q_x^2 + q_y^2)}\right) \end{pmatrix} \tag{2.9}$$

### 2.3.5 *Robotic Ultrasound*

For robotic US, a *Clarius CR HD3* probe (Clarius Mobile Health, Canada) is attached to the robotic end effector using a 3D-printed mount. The US images

are streamed wirelessly and displayed on an *iPad Air 2* (Apple Inc., USA) using the Clarius Ultrasound App (see Figure 2.11).

To enable controlled experimentation and repeatable testing, customized US phantoms were developed. These phantoms were made from agar-agar, a commercially available gelling agent that is low-cost and allows for the simple construction of echogenic volumes [372] (concentration 30 g agar-agar/500 ml water). Target structures were embedded within the phantoms, created either from geometric rubber shapes or from agar-agar with a higher concentration (100 g agar-agar/500 ml water) to provide distinct contrast in the US image. In cases where rubber shapes were used, food coloring was added to the agar-agar to obscure the target structures from view. To precisely position these target structures within the phantom, brackets were 3D printed with dedicated seats for placing the structures. The brackets also included a marker shield protruding from the phantom, enabling its localization.

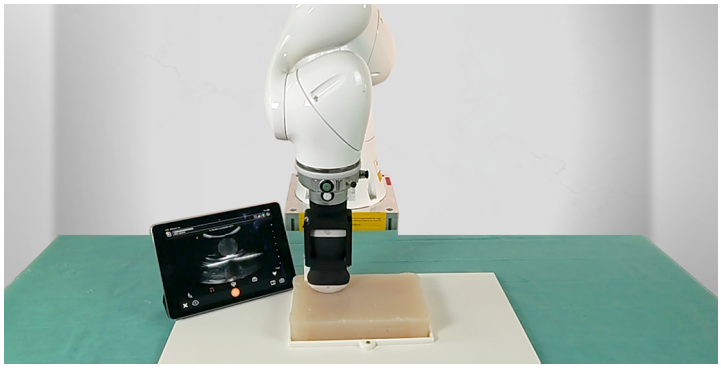


Figure 2.11: Setup for robotic US with phantom made of agar-agar and embedded agar-agar sphere visualized in US image.

### 2.3.6 *System Architecture*

Communication is required between the components described above. To summarize, the presented system consists of three main components:

- *Robot Control:* Runs on the robot controller, managing robot motion and sensor data, including data on robot pose and external forces.
- *XR Device:* Responsible for visual feedback, display of virtual components, registration, and both marker and hand tracking.
- *Tracking System:* Includes an external hand tracking sensor for increased flexibility, and an optical tracking camera for tracking task elements and supporting spatial registration.

External tracking and coordination are implemented in Unity and are running on a desktop PC. Communication between the three components is established using a UDP connection.

### 2.3.7 *Ethical Considerations*

All studies were carried out in accordance with the principles of the Declaration of Helsinki [353]. Prior to their involvement, participants were fully briefed on the study goals and procedures, and written informed consent was obtained. Participant privacy was preserved through the anonymization of all collected data, in accordance with local data protection regulations (DSGVO). The study procedures were designed to minimize any potential risk or discomfort to participants, ensure voluntary participation, and allow participants to withdraw from the study at any time.

# Part I

# XR-TO-ROBOT REGISTRATION

**Synopsis:** This part focuses on the alignment of XR within robotic workspaces. It describes two experiments investigating different aspects of XR-to-robot registration. The first experiment is an empirical user study examining the efficiency of three interactive registration methods. The second experiment is a technical evaluation of parameters influencing point-based registration accuracy.

This part contains material adapted from the following publications:

Tonia Mielke, Fabian Joeres, Danny Schott, and Christian Hansen. "Interactive Registration Methods for Augmented Reality in Robotics: A Comparative Evaluation." In: *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2023, pp. 501–506. DOI: 10.1109/ISMAR-Adjunct60411.2023.00109 [7]. © 2023 IEEE.

Tonia Mielke, Florian Heinrich, and Christian Hansen. "Enhancing AR-to-Robot Registration Accuracy: A Comparative Study of Marker Detection Algorithms and Registration Parameters." In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 19–23. DOI: 10.1109/ICRA55743.2025.11128039 [4]. © 2025 IEEE.

# 3 INTRODUCTION

## 3.1 MOTIVATION

For HRI, the transformation between different coordinate systems is of key interest. As detailed in Section 2.1.1.1, the robot is operated within its own coordinate frame and computes joint angles through inverse kinematics. However, movement commands may originate from other coordinate frames, such as the sensor or input device frames. Accurate mapping of these movement commands to robot motion requires knowledge of the transformations between coordinate frames, which can be derived through registration [61].

Beyond transformation of robot control commands, registration is also essential for integrating AR into robotic workspaces. AR can be employed to support programming and control by visualizing targets, waypoints, or paths [59, 114, 263], to enhance perception through visual cues [22], to improve safety by displaying workspace information [280, 311, 329], and to communicate robot intent in robotic systems with higher LORA [19, 119].

Although registration is a defining characteristic of AR, as established in Azuma's widely adopted definition [25], it is equally relevant in other domains grouped under the umbrella term XR. AV can be used to maintain the real robot in the loop while embedding it in virtual contexts, such as simulated environments for safe interaction testing [63] or real-scene overlays to improve user acceptance [309]. Similar to AR, employing AV requires registration to establish the spatial relationship between real and virtual elements. In contrast, VR is typically used for immersive teleoperation in fully virtual environments [351]. Thus, VR does not require aligning virtual with real-world objects. However, HRI in VR settings still depends on registration, as user input has to be aligned for enabling robot control.

In summary, rigid registration is essential to align both the interaction space and virtual content with the robotic workspace. In this thesis, mid-air hand gestures are used for HRI, a method also commonly employed in XR environments. As hand-tracking sensors are often integrated into devices such as HMDs [268], both interaction and integration of virtual components can be mediated through XR systems. Consequently, a unified registration approach can enable their spatial alignment with the robot. In this thesis, this alignment process is referred to as *XR-to-robot registration*. While prior work has explored hand gesture-based HRI and XR applications in robotics, the challenge of precise spatial alignment has received limited attention. To address this gap, this chapter investigates registration methods to answer the following RQ:

> **RQ1** | How can efficient and accurate XR-to-robot registration be achieved?

## 3.2 RELATED WORK

Although many commercial industrial robots provide tools to define external coordinate systems [159, 322], these processes do not provide interfaces for seamless integration with XR systems. Therefore, alternative approaches are required. This section provides an overview of related work on manual, point-based, and surface-based registration methods for robotic workspaces, as well as studies on parameters affecting point-based registration.

### 3.2.1 *Registration Methods*

As research on XR-to-robot registration is still limited, this section reviews both studies specific to XR-to-robot registration and broader approaches to robot-workspace registration. The registration methods can be categorized as manual, point-based, and surface-based registration (see Table 3.1).

In manual registration, the robotic workspace and the XR environment are interactively aligned. Therefore, a life-size virtual model of the robot is holographically displayed and aligned with the physical robot using interaction techniques [110, 263, 273]. Another approach is to manually position a tracked image marker in the environment at a known distance relative to the robot, typically performed by referencing the known location of the robot base and measuring the distance to the marker [41, 54, 178].

Point-based registration relies on defining corresponding points in both the XR and robot coordinate frames. Since the position of the robotic end effector in the robotic frame can be computed through forward kinematics, the challenge lies in determining the corresponding points in the XR frame. One approach is to define these points visually by either displaying them in XR [116] or defining them in relation to tracked image markers [141, 229], and then moving the robotic end effector to those positions. However, guiding the robot to visually defined points may be affected by perceptual inaccuracies, introducing an FLE that can negatively impact registration accuracy [107]. To mitigate this, the registration points can be physically defined by using fixtures at known positions in the environment. By utilizing a matching flange, the robotic end effector can be precisely positioned at the registration points [211, 225, 232, 325]. An alternative to predefined registration points is to track the robotic end effector within the XR coordinate frame, enabling registration at mid-air positions. The localization of the end effector can be achieved by attaching a tracking marker to the end effector of the robot [185, 228] or by moving a tracked device (e.g., a handheld tracker) to the location of the robotic end effector [246].

The third type of registration is surface-based registration, where point clouds of the robotic workspace are used for alignment. For XR-to-robot registration, this method can rely on depth point clouds acquired by the depth cameras integrated into HMDs [61, 241, 258].

Table 3.1: Overview of related work on XR-to-robot registration.

| Method | | Details | Evaluation |
|---|---|---|---|
| Manual | [273] | Manually aligning robot | None |
| | [110] | | Duration & accuracy |
| | [263] | | None |
| | [178] | Manually placing marker | Accuracy |
| | [54] | | Tracking accuracy |
| | [41] | | None |
| Point-based | [116] | Projected points | None |
| | [229] | Points relative to marker | None |
| | [141] | | None |
| | [246] | Tracked by controller | None |
| Surface-based | [61] | ICP with depth point cloud | None |
| | [241] | | Accuracy & runtime |
| | [258] | | Accuracy |

### 3.2.2 *Point-Based Registration Parameters*

The second experiment in this part of the thesis focuses specifically on point-based registration, as this method shows potential for efficient XR-to-robot registration. However, previous work has indicated that its accuracy depends on the characteristics of the registration points. Early investigations into these point characteristics originated in the domain of image-guided surgery, where point-based registration is commonly applied. West et al. [341] examined the impact of registration point placement. They observed that registration errors can often be reduced by using a greater number of well-distributed points, avoiding near-collinear configurations, and positioning the centroid of the registration points close to the target volume. These findings were confirmed and expanded by Hamming et al. [126], who emphasized the importance of a well-configured registration point layout over simply increasing the number of fiducials. Ershad et al. [96] further highlighted the impact of FLE on registration accuracy.

Although these insights are valuable, their application to XR-to-robot registration must be done carefully. Robotic systems introduce additional sources of error, including sensor inaccuracies and kinematic uncertainties [211]. While some findings are consistent across domains, such as the benefit of increasing the distance between points [211, 325] and using a higher number of points [211], Nguyen et al. [228] found that larger distances between points can actually decrease registration accuracy in robotic workspace registration during optical tracking. This suggests that the impact of point placement depends on the specific tracking technology used.

This issue is especially relevant for XR-to-robot registration scenarios, where marker-based tracking is often used with the cameras integrated in HMDs [74]. Prior work has shown that marker detection is influenced by factors such as marker size, viewing angle, and viewing distance [161]. Furthermore, different tracking algorithms perform differently under these conditions [74].

Therefore, in point-based registration between XR devices and robots, both the configuration of the registration points and the spatial relationship between the XR device and the robot might play a critical role in registration accuracy.

## 3.3 RESEARCH GAP

As summarized in Table 3.1, registration methods for XR-to-robot alignment are often merely presented, typically as part of works focusing on augmentation techniques, rather than systematically evaluated. In studies where registration methods are evaluated, the emphasis is primarily on accuracy [54, 110, 178, 241, 258], with only limited attention given to registration duration [110]. Additionally, none of the reviewed works consider the user's workload during the registration. However, investigations that go beyond accuracy, including both duration and user workload, are essential to gain more comprehensive insights into the efficiency of different methods for XR-to-robot registration.

Furthermore, regarding point-based registration, the importance of registration point characteristics has been highlighted in prior work on image registration and general robot-workspace registration [96, 126, 341], and the dependence of optimal parameters on the tracking technology has been explored [211, 228, 325]. However, these influencing factors have not yet been systematically investigated for XR-to-robot registration.

Consequently, two gaps remain. The first concerns the lack of research on the efficiency of registration methods, particularly with respect to factors beyond accuracy, such as duration and user workload. The second concerns the unexplored impact of specific registration parameters, such as point placement and configuration, on registration accuracy in XR-to-robot registration.

## 3.4 CONTRIBUTION

To address these gaps, this part of the thesis aims to enable more efficient XR-to-robot registration by providing the following contributions:

**Investigation on registration methods.** Chapter 4 systematically evaluates three registration methods, offering insights into the suitability of manual and point-based techniques for achieving efficient alignment in HRI.

**Investigation on registration parameters.** To improve the accuracy of point-based XR-to-robot registration further, Chapter 5 explores the influence of different registration point characteristics, tracking techniques, and XR device placement. Additionally, a refinement approach is presented that uses additional recorded points for surface-based registration improvement.

# 4 INVESTIGATION ON REGISTRATION METHODS

As previously described, related work has proposed manual, point-based, and surface-based registration methods for XR-to-robot alignment. However, these registration approaches are typically presented only as a means to enable XR integration into robotic workspaces, without a systematic evaluation of the methods themselves. In particular, aspects such as accuracy, duration, and user workload are often not investigated, even though they are essential for assessing the practical applicability of each method. To address this gap, this chapter presents a comparative evaluation of registration approaches, building on methods proposed in related work on XR-to-robot registration as well as more general approaches to robot-workspace registration.

Parts of this chapter were previously published in: Tonia Mielke, Fabian Joeres, Danny Schott, and Christian Hansen. "Interactive Registration Methods for Augmented Reality in Robotics: A Comparative Evaluation." In: *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2023, pp. 501–506. DOI: 10.1109/ISMAR-Adjunct60411.2023.00109 [7] © 2023 IEEE.

## 4.1 TECHNICAL METHODS

*Supplementary video*

This section introduces and describes the development and implementation of three registration methods for aligning XR content within the robotic workspace. The goal of this chapter is to identify registration methods that are not only efficient in use but also easy to implement, enabling developers to integrate them seamlessly into their frameworks. Therefore, this chapter focuses on registration approaches that leverage commonly built-in capabilities of HMDs, such as content visualization, spatial mapping, and hand tracking.

Surface-based registration methods require the acquisition of point clouds that represent the robotic workspace. However, while HMDs commonly include depth cameras, access to raw depth data is often restricted or non-trivial. Additionally, the registration is computationally intensive, requiring an external PC to receive the point cloud data and perform the registration [258]. In contrast, manual and point-based methods do not necessitate external hardware, and the required input, such as hand tracking data or marker and object tracking, can be more easily integrated into existing systems. Therefore, this chapter focuses on manual and point-based registration approaches.

In *Manual Registration*, the alignment is performed entirely manually using either tracked real [41, 54, 178] or virtual objects [110, 263, 273], placing them within the robotic workspace. Placing tracked real objects, such as markers, at known positions requires workspace setups where the robot base location is precisely known. Since this setup can be restrictive and dependent on the physical environment, this chapter adopts the more flexible approach of interactively aligning virtual models with their real-world counterparts.

In point-based approaches, corresponding points are determined in both the robotic and the XR coordinate space by steering the end effector to different registration points. Obtaining registration points in the robot reference frame is straightforward, as the end effector position can be derived through forward kinematics. However, different strategies exist for determining the corresponding points in the XR coordinate space. Two common approaches are: defining registration points at known positions to which the end effector is moved, or directly tracking the end effector in the XR space for flexible point acquisition. When registration points are predefined, they can be specified either visually [116, 141, 229] or physically [211, 225, 232, 325]. Since visually defined registration points may introduce FLE, this chapter focuses on the physical definition of registration points using a board (*board* registration). The second approach instead aims to visually track image markers affixed to the robot, allowing mid-air point acquisition (*mid-air* registration) [185, 229]. Figure 4.1 provides a visual representation of the components involved in the registration process and the relevant transformations.

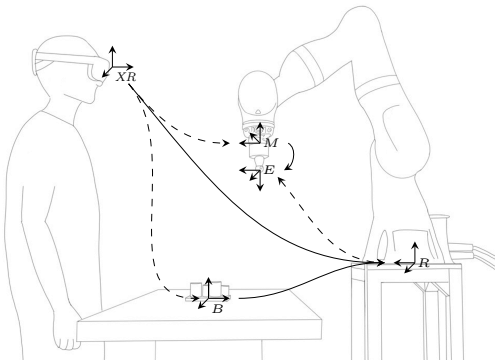
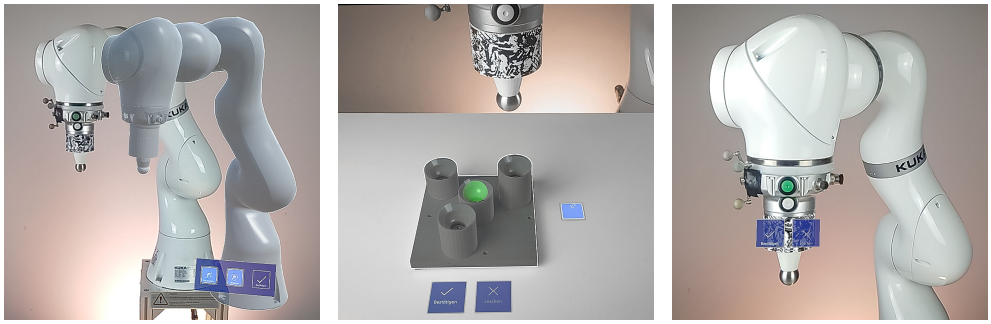



Figure 4.1: Visualization of coordinate systems used in registration: dashed lines depict transformations automatically derived from mechanical or optical tracking, while solid lines represent transformations obtained through the registration process. Reprinted from [7], © 2023 IEEE.

### 4.1.1 *Manual Registration*

In *manual* registration, the goal is to interactively align a virtual object as closely as possible with its real-world counterpart. As the position of the robot reference frame in relation to the robot itself is known, a virtual model of the robot is used for alignment. The registration begins with a life-sized virtual robot placed at a random orientation in fixed distance in front of the user. The holographic visualization is semi-transparent with opaque outlines (see Figure 4.2a). To manipulate the placement of the virtual robot, it can be grasped using the pinch gesture. The translational and rotational DOFs can then be adjusted either independently or simultaneously by performing manipulative hand gestures, which map the relative movement of the hand 1:1 to corresponding transformations of the virtual robot. Activation of the manipulation modes is controlled via virtual buttons.



(a) Manual Registration (b) Board Registration (c) Mid-Air Registration

Figure 4.2: Overview of UIs during registration. Reprinted from [7], © 2023 IEEE.

### 4.1.2 *Board Registration*

The board registration utilizes a 3D-printed board with four seats and a matching flange attached to the robotic end effector (see Figure 4.2b). The transformation between the XR and the robot coordinate frame $T_{XR}^{R}$ is derived by guiding the end effector to the seats on the board. The position of the end effector, and thus the registration points in the robot coordinate frame, $T_{R}^{E}$, is determined by robot kinematics. In the XR coordinate frame, the registration point positions are determined by tracking the board $T_{XR}^{B}$ through object tracking and deriving the individual registration point positions based on their relative locations, which are known from the board's geometry.

The initial step of the registration process involves localizing the board using object tracking and fixing its position using a virtual button. This approach aims at improving performance by eliminating the need for continuously tracking the board's static position. The second step of the registration process consists of subsequently localizing the four registration points in the robot coordinate system, by hand-guiding the robot to the seats. To this end, one seat at a time is highlighted in XR, indicating which point should be recorded. Once the end effector is positioned at the desired location, a virtual button has to be pressed to save the coordinates. After all points have been captured, the transformation between the resulting sets of corresponding points is calculated using the least-squares method (see Section 2.3.4).

### 4.1.3 *Mid-Air Registration*

To obtain the registration using the *mid-air* method, an image marker is attached to the robotic end effector. A cylindrical marker is used to ensure the marker's recognizability from different orientations (see Figure 4.2c). The transformation $T_{XR}^{R}$ between XR and the robot is obtained by moving the end effector $E$ to four arbitrary positions, which are then simultaneously recorded in both the robot coordinate system $R$ and the XR coordinate system. As only the position of the marker $M$ in the XR coordinate system $T_{XR}^{M}$ is known, but the position of $E$ is required, the transformation $T_{M}^{E}$ has to be calculated.

Hence, the initial stage of the registration process involves obtaining $T_M^E$ by a pivot calibration, where the end effector (and consequently the marker) is rotated around the TCP. To this end, an automated rotation routine is implemented, which starts upon marker recognition and enables continuous recording of the marker's position and orientation as the robot autonomously rotates. Once a sufficient number of points is recorded, the rotation stops, and $T_M^E$ is calculated using the Algebraic Two-Step method [363]. A RANSAC framework [105] is implemented to increase stability and robustness by reducing the potential impact of outliers on the calibration results.

Once the pivot calibration process is completed, the end effector can be moved to arbitrary positions to record calibration points. Similar to the board registration method, the robot can be hand-guided, and the points are recorded by pressing a virtual button. Manually guiding the robot to the registration points allows interactive point selection, ensuring sufficient tracking quality and avoiding obstacles in the workspace. After recording four points, the transformation $T_{XR}^R$ is calculated using the least-squares method.

### 4.1.4 *Implementation*

The prototype was implemented as described in Section 2.3. To enable interactive elements such as buttons, pop-ups, hand tracking, and object manipulation, the *Mixed Reality Toolkit* was used. The *Vuforia SDK* was employed for cylindrical image marker tracking (to track the end effector) and object tracking (to track the board and the multimodal ground truth marker). To steer the end effector, hand-guiding was implemented using impedance control with minimal stiffness (1 N/m and 1 N/rad) and high nullspace stiffness (100 N/rad), allowing the robotic arm to respond to external forces.

## 4.2 EVALUATION METHODS

To comparatively evaluate the described interactive registration approaches, a user study was conducted. This study followed a within-subject, one-factorial design, which is described in the following section.

### 4.2.1 *Tasks*

Participants performed XR-to-robot registration using each of the three techniques: *manual*, *board*, and *mid-air* registration. Thus, the tasks included:

- **Manual Registration:** Interactively aligning the virtual and real robots through hand gestures and confirming the alignment once satisfied.
- **Board Registration:** Locating the board within the HMD FOV, confirming accurate tracking, and recording four calibration points.
- **Mid-Air Registration:** Conducting a pivot calibration followed by the recording of four mid-air calibration points.

### 4.2.2 *Variables*

INDEPENDENT VARIABLES The independent variable was the registration technique, with three levels: *manual*, *board*, and *mid-air*. Each participant performed all three methods, with the order counterbalanced across participants.

DEPENDENT VARIABLES To assess registration performance and user experience, three dependent variables were evaluated: registration accuracy, duration, and perceived workload.

Registration accuracy was assessed using the TRE, which requires known corresponding points in both coordinate frames. For point-based registration, this would be straightforward, as specific target points can be tracked in both coordinate spaces and transformed using the registration results, allowing the Euclidean distance between them to be calculated. However, manual registration does not provide explicit point correspondences, making the computation of TRE less trivial. To address this, an alternative TRE assessment method was employed for all three registration approaches. A ground truth was established using an optical tracking system. By performing a point-based registration with a pivot-calibrated retro-reflective marker shield, the transformation $T^{R}_{optical}$, representing the relationship between the robot and the optical tracking system, was determined. To compare this ground truth with the registration results, the transformation $T^{optical}_{XR}$ between the XR system and the optical tracking system had to be computed. This was done using a multimodal marker serving as a shared reference between the two systems, enabling the computation of this transformation.

$$T^{\text{optical}}_{\text{XR}} = T^{\text{optical}}_{\text{reference}} \cdot T^{\text{reference}}_{\text{XR}} \tag{4.1}$$

To calculate the TRE, four target points located at the boundaries of the reachable and trackable workspace were selected. These points were transformed from the XR to the robot coordinate frame using both the ground truth transformation and the registration result. The TRE was then computed as the Euclidean distance between these projections:

$$\text{TRE} = ||T_{\text{groundTruth}}(p) - T_{\text{registration}}(p)|| \tag{4.2}$$

The duration of the registration process was assessed as TCT, measured from the initiation of registration to its confirmed completion. Perceived workload user feedback was evaluated using the raw NASA TLX [131].

### 4.2.3 *Hypotheses*

The conducted study was exploratory, making no a priori assumptions about the outcomes. Therefore, two-sided hypotheses were investigated. The null and alternative hypotheses considered were:

$H_0.x$ The mean $x$ is equal for all registration methods.
$H_1.x$ The mean $x$ differs for at least one registration method.
$x \in \begin{cases} 1\text{-TRE} \\ 2\text{-TCT} \\ 3\text{-TLX} \end{cases}$

### 4.2.4 *Sample Design*

As the study investigated general characteristics of XR-to-robot registration, participants with a technical background were recruited from the local university via online polls. Participants received a compensation of 15€.

### 4.2.5 *Procedure*

At the beginning of the experiment, participants were given a brief explanation of the research objectives, technical principles, and experimental procedures. Then, participants were asked to complete a demographic questionnaire and sign an informed consent form. As there were no prerequisites for participation in the study in terms of experience with XR and HRI, a brief training session was performed. The training consisted of two tasks: the first task involved interacting with virtual objects in XR, while the second task focused on hand-guiding the robot to positions displayed in XR. After completing the training, an experimental block was conducted for each registration method. At the beginning of each block, participants were given an overview of the method, followed by step-by-step instructions for the first trial. A second training trial was conducted, during which participants could ask any remaining questions. Finally, a final test trial was conducted in which the dependent variables were recorded. Upon completion of the registration, participants were asked to complete the NASA TLX questionnaire. After finishing all three registration blocks, a brief semi-structured interview was conducted. The experiment took an average of 45 minutes.

### 4.2.6 *Statistical Analysis*

For each dependent variable (TRE, TCT, raw TLX) and registration method (*board*, *mid-air*, *manual*), data were tested for normality and homogeneity using the Shapiro-Wilk and Levene's tests. As these assumptions were violated for all dependent variables, robust repeated measures ANOVAs for within-subjects designs based on trimmed means were calculated (see [348]). Effect sizes were reported using the $\delta_t$ estimate proposed by Algina et al. [16]. In cases of significant effects, pairwise post hoc paired-sample Yuen's tests based on trimmed means with Holm corrections for multiple testing were conducted. All statistical analyses were conducted using R (version 4.4.0).

## 4.3 RESULTS

This section summarizes and interprets the findings of the user study, including both quantitative and qualitative results.

### 4.3.1 *Data exclusion*

The Interquartile Range (IQR) Method was employed to detect extreme outliers. The results of one participant showed an inexplicably high TRE of 142 mm for the *mid-air* registration technique, which exceeded the range defined by the upper quartile ($Q_3$) plus three times the IQR. Consequently, this participant's data were excluded from the analysis. No other instances of extreme outliers were identified.

### 4.3.2 *Participants*

The 20 participants whose data were used for the analysis had a technical background (11 male, 7 female, and 2 diverse). They were aged between 23 and 31 years (Mean (M)=26.00, Standard Deviation (SD)=2.62) and were students (14), PhD students (4), or engineers (2) in the areas of Computer Science and Medical Engineering. Regarding the task-relevant experience, participants provided the following ratings on a scale from 1 (no experience) to 5 (very experienced): AR (M=2.75, SD=1.21), VR (M=3.20, SD=1.28), HRI (M=2.10, SD=1.12), and gaming experience (M=3.85, SD=1.27).

### 4.3.3 *Quantitative Results*

The descriptive results are summarized in Table 4.1, and the statistical results are outlined in Table 4.2. Descriptive statistics and significant post hoc comparisons are also visualized in Figure 4.3.

Table 4.1: Summary of descriptive results for all dependent variables (n=20). Entries are in the format: mean value [standard deviation].

| **Variable** | **TRE (mm)** | **TCT (s)** | **TLX** |
|---|---|---|---|
| *Manual* | 43.23 [17.22] | 106.71 [50.91] | 8.54 [3.48] |
| *Board* | 12.13 [6.42] | 59.09 [15.33] | 4.32 [2.33] |
| *Mid-Air* | 14.86 [4.28] | 85.26 [36.28] | 4.63 [2.29] |

Table 4.2: Summary of robust ANOVAs test results ($\alpha < .05$) reporting test statistic $F$ and effect size $\delta_t$.

| **Variable** | **F** | **p** | **Sig.** | $\delta_t$ | **Effect** | **Fig.** |
|---|---|---|---|---|---|---|
| **TRE** | 9.23 | 0.009 | * | 1.24 | Large | 4.3a |
| **TCT** | 80.60 | <0.001 | * | 3.08 | Large | 4.3b |
| **TLX** | 27.20 | <0.001 | * | 1.64 | Large | 4.3c |

### 4.3.4 *Qualitative Results*

For the analysis of qualitative feedback, statements collected during the semi-structured interview were paraphrased and clustered. Only statements that were made by at least two participants were included. A total of 171 individual statements were recorded, of which 133 occurred at least twice, leading to the formulation of 33 summarizing statements, presented in Table 4.3.

Table 4.3: Summary and frequency of statements (#) received during the semi-structured interview. Adapted from [7], © 2023 IEEE.

| **General** | **Board Registration** |
|---|---|
| + Robot control through hand-guiding is intuitive (2) and easy (2) | + Interaction easy (8) and intuitive (6) |
| - Interaction with the virtual buttons sometimes challenging (5) | + Feeling confident during interaction (8) |
| - Inaccuracies in object tracking (4) and image marker tracking (4) | + Predefined registration points (6) and haptic feedback (5) helpful |
| | + Registration expected to be precise (5) |
| | - Problems with occlusion by AR (3) |
| **Manual Registration** | **Mid-Air Registration** |
| + Seeing registration results helpful (4) | + Interaction easy (4) and intuitive (3) |
| + Interaction feels controlled (3), is intuitive (2), and effective (2) | + Feeling like can do nothing wrong (2) |
| - Interaction requires training (3) | - Problems with robot joint limits during hand-guiding (6) |
| - Registration takes time (2) | - Enabling the tracking of the marker requires maintaining an uncomfortable body posture (5) |
| - Issues with Hand-tracking (4) and difficulty grabbing the virtual robot when occluded by the physical one (4) | - Feeling insecure during registration (8) |
| - Difficulties in manipulation of virtual robot (3), especially for rotation (6) | - Uncertainty about criteria for accurate registration (4) |
| - Difficulties due to depth perception (2) | • Selecting registration points would be easier with broad placement suggestions from the system (3) |
| • Interaction would be easier if manipulation restricted to certain axes (3) | |

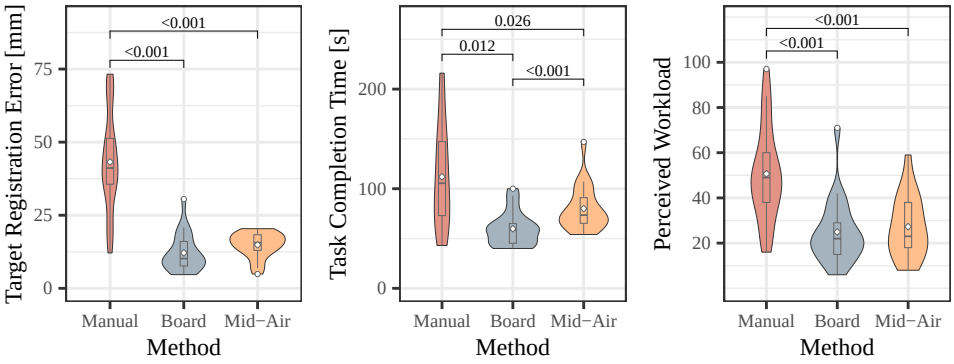
### 4.3.5 *Interpretation of Results*

Significant main effects of the registration method were found for TRE, TCT, and NASA TLX. Consequently, the null hypotheses $H_0.1$–$H_0.3$ were rejected in favor of the respective alternative hypotheses $H_1.1$–$H_1.3$. Post hoc pairwise comparisons revealed that *manual* registration was significantly outperformed by both point-based methods in terms of accuracy, duration, and perceived workload (see Figure 4.3). The lower accuracy may be due to difficulties with depth perception and interacting with the virtual model, as well as technical issues with hand tracking, as reported in the qualitative feedback. These challenges likely also contributed to the higher TCT and perceived workload, despite participants describing the interaction as intuitive and reporting a sense of control. One participant noted that "the physical robot is easier to move than the virtual model", highlighting the advantages of robot hand-guiding, which was seen as intuitive and easy compared to the more demanding XR interaction. Although *manual* registration performed worse

quantitatively, it offered valuable insights, with participants appreciating the direct visibility of results and the sense of control.

While participants reported high perceived accuracy with *board* registration, no significant differences in accuracy were found between *board* and *mid-air* registration. This suggests that, despite users feeling uncertain about point placement during *mid-air* registration, the flexibility in positioning did not negatively impact accuracy. However, this lack of confidence should be addressed, especially since participants reportedly felt more confident using the *board* method. One possible solution is to visualize suggested registration points or regions in XR. As point placement can affect accuracy [341], providing such guidance may improve both accuracy and user confidence. Another challenge with *mid-air* registration was the robot exceeding its joint limits, causing abrupt stops and requiring system restarts. Visualizing virtual registration points within the robot's safe workspace and adding visual boundaries could help users avoid these issues.

The study findings also show that *board* registration results in significantly lower TCT compared to the other two methods. The longer duration of *mid-air* registration may be due to the time needed for pivot calibration. This could be reduced by placing the marker at a predefined distance from the TCP, eliminating the need for separate calibration. Alternatively, a distinctively shaped end effector could be designed for direct object tracking.




(a) TRE (b) TCT (c) TLX

Figure 4.3: Violin Plots of experimental results of all dependent variables. Brackets indicate statistically significant post hoc pairwise comparison results. Means are indicated by ◇. Adapted from [7], © 2023 IEEE.

## 4.4 DISCUSSION

The findings of the described study show that the efficiency and effectiveness of XR-to-robot alignment significantly depend on the registration method employed. This section discusses the results, outlines the study's limitations, and formulates design implications based on the findings.

MANUAL REGISTRATION As previously described, the lower accuracy and higher duration and perceived workload for the *manual* registration method

might be attributed to difficulties in depth perception and manipulation of the virtual robot model. Problems regarding spatial understanding could possibly be alleviated by sophisticated approaches such as the virtual mirrors presented by Martin-Gomez et al. [209]. Since participants were not required to have previous experience in XR and expressed the need for training to learn the interaction, future work should analyze the learning curve and explore how more experience or longer training impacts the registration results.

POINT-BASED REGISTRATION Even for the point-based approaches, the TREs observed in this study remained relatively high (*board* registration: $12.13 \pm 6.42$ mm; *mid-air* registration: $14.86 \pm 17.22$ mm). While this level of accuracy may suffice for coarse registration tasks, such as aligning the interaction space for relative movements or providing visual feedback in the robotic workspace, more precise applications like defining or visualizing specific target locations in the robot's reference frame may require higher accuracy.

Tracking errors might have introduced an FLE, resulting in the performance of the point-based approaches being dependent on image and object tracking accuracy. For instance, Allen et al. [17] reported tracking errors of up to 10 mm for cylindrical markers using *Vuforia* on the *HoloLens 2*, suggesting that tracking inaccuracies likely affected the registration process. As a result, the observed TREs may reflect not only user performance but also technical limitations. Additionally, the accuracy may have been influenced by the selection and distribution of registration points [325, 341]. However, these parameters were not systematically investigated in this study.

GENERAL When interpreting the accuracy results, it is important to consider how the TRE was calculated. The ground truth may have included errors, as it was itself established through registration, and the object tracking used to define the shared reference between ground truth and registration results may have introduced additional inaccuracies. Furthermore, since the TRE was calculated using points at the outer boundaries of the workspace, any rotational registration errors would have had a stronger impact.

### 4.4.1 *Limitations*

One limitation of this study is that it focused solely on interactive manual and point-based registration methods, motivated by the goal of implementing techniques that can be easily integrated into existing HRI setups. However, registration accuracy could potentially be improved by incorporating surface-based registration methods. For example, adding a refinement step using point clouds of the robotic workspace may enhance accuracy [258]. Furthermore, the *mid-air* registration could be automated, pre-programming motion to registration points. This would eliminate the uncertainties in point placement and mitigate issues related to robot boundary constraints.

Another limitation arises from the use of only one specific hardware setup, which restricts the generalizability of the findings. Although general insights regarding registration techniques and interaction are believed to be

transferable, technical factors may have influenced the results. For instance, other XR devices might offer different marker tracking or hand tracking performance, which could affect registration outcomes. To support broader applicability, future work should investigate the performance of the methods across different devices and robotic platforms.

### 4.4.2 *Implications*

In conclusion, based on the study results, the following design implications are proposed to support efficient XR-to-robot registration in HRI setups:

**Employ point-based registration.** The findings suggest that point-based registration methods enable significantly more accurate registration while reducing perceived workload. Participants found it easier to control the physical robot for registration than to adjust the virtual model's DOFs manually.

**Provide guidance for registration points.** A key advantage of *board* registration over *mid-air* registration was that predefined registration points increased participants' confidence. When using *mid-air* registration, visual guidance for point placement should be provided to improve both user confidence and the quality of registration points.

**Provide visual feedback.** Participants appreciated the ability to see the registration result during *manual* registration. This highlights the value of visual feedback not only for assessing registration accuracy. Additional visual feedback should be incorporated to help visualize the workspace and robot limits, thereby supporting robot control during registration.

## 4.5 CONCLUSION

In this chapter, three registration methods were implemented and evaluated for enabling XR-to-robot registration. A manual registration method was developed, where XR content was aligned with the robot through hand interactions. Additionally, two point-based approaches were implemented: one defining registration points using a 3D-printed board, and another employing a marker attached to the robot for tracking. These techniques were empirically evaluated in a user study involving 21 participants, focusing on performance and user workload. The results indicate that manual registration was outperformed by the two point-based methods in terms of accuracy, duration, and perceived workload. Using a physical board for registration additionally enabled significantly faster registration than the other two methods.

By identifying point-based registration as an efficient and user-friendly method, this chapter establishes a foundation for practical XR-to-robot registration. However, since the achieved accuracy may still be insufficient for high-precision applications, it also motivates the investigation presented in the next chapter, which explores factors that may influence point-based registration accuracy.

# 5 INVESTIGATIONS ON REGISTRATION PARAMETERS

Chapter 4 showed that point-based registration methods have potential for enabling efficient XR-to-robot registration for HRI. However, the registration accuracies achieved might not be sufficient for high-precision applications.

In image-guided surgery, prior research has found that registration accuracy heavily depends on the quality of registration points, including factors such as their number, distribution, and spacing [96, 126, 341]. Similar findings have been reported for robot-workspace registration, where increasing both the number and distance between physically defined points improves accuracy [211, 325]. However, in contrast, robot-camera registration studies have indicated that greater distances between points can reduce accuracy, as they may exceed the sensor's optimal range [228]. This suggests that registration accuracy depends not only on the configuration of the points but also on the sensors involved, implying that findings from other domains may not fully transfer to XR-to-robot registration. Furthermore, marker detection performance has been shown to depend on factors such as XR device placement [161] and tracking technique [74].

While related domains offer important guidance, their findings might not be directly applicable to XR-to-robot registration, as specific tracking characteristics, spatial setup, and point properties may interact in this context. As such, a systematic investigation is needed to understand the key influencing factors specific to XR-to-robot alignment. Therefore, this chapter aims to address this gap by conducting a series of experiments investigating the effects of tracking techniques, registration point characteristics, and XR device placement on registration accuracy, with the goal of improving point-based registration accuracy for XR-to-robot alignment.

Parts of this chapter were previously published in: Tonia Mielke, Florian Heinrich, and Christian Hansen. "Enhancing AR-to-Robot Registration Accuracy: A Comparative Study of Marker Detection Algorithms and Registration Parameters." In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 19–23. DOI: 10.1109/ICRA55743.2025.11128039 [4]
© 2025 IEEE.

## 5.1 TECHNICAL METHODS

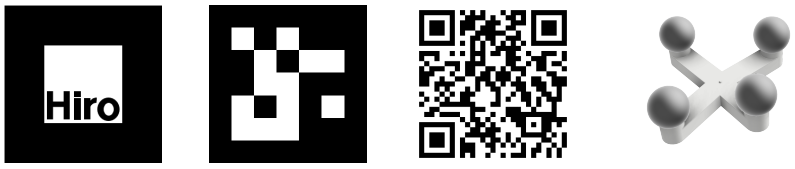
To investigate factors influencing XR-to-robot registration accuracy, a point-based registration method using an image marker attached to the robotic end effector was implemented. While Chapter 4 indicated that utilizing a 3D-printed board to define registration points offers advantages in registration duration, defining points mid-air allows greater flexibility in exploring different registration point configurations. Therefore, the *mid-air* registration approach described previously is used as the registration method in this chapter, with the technical details outlined in the following section.

### 5.1.1 *Tracking Techniques*

*Mid-air* registration relies on establishing corresponding points in the XR and robot coordinate frames by attaching a marker to the robotic end effector. To investigate the impact of different tracking techniques on XR-to-robot registration performance, four publicly available optical tracking approaches were implemented. These tracking algorithms include three square fiducial image markers and one approach that integrates a marker shield with retro-reflective spheres (see Figure 5.1). These tracking approaches were selected as they support the Universal Windows Platform (UWP), ensuring compatibility with the *HoloLens 2*, which is used as XR device in this thesis. The investigated tracking techniques were:

1. **ARToolKit** [259]: Integration of ARToolKit (v5.3.2) with UWP, available at: `https://github.com/qian256/HoloLensARToolKit`
2. **ArUco** [90]: Integration of ArUco Marker Detection with *HoloLens 2* from OpenCV available at: `https://github.com/doughtmw/display-calibration-hololens`
3. **Vuforia**: A widely used AR platform supporting image and object recognition, available at: `https://developer.vuforia.com/`
4. **Retro-Reflective Marker Tracking** [208]: Tracking of custom retro-reflective marker shields using the HoloLens 2 Research Mode, available at: `https://github.com/andreaskeller96/HoloLens2-IRTracking`

Windows Mixed Reality QR Code tracking was initially considered for inclusion. However, preliminary experiments showed that its reliance on stereo cameras with limited FOV and low resolution made it unsuitable for accurate registration.




(a) ARToolKit (b) ArUco (c) Vuforia (d) Retro-Reflective

Figure 5.1: Markers used for tracking for XR-to-robot registration. Reprinted from [4], © 2025 IEEE.

### 5.1.2 *Registration*

The registration is conducted as described in Section 2.1.4.3. First, the offset between the marker and the robot's TCP is determined through pivot calibration using the Algebraic Two-Step method [363], employing a RANSAC framework [105]. The registration is then performed by automatically moving the end effector to pre-programmed registration points, recording the positions in the robot and XR coordinate frames, and calculating the transformation using the least-squares method [23]. To ensure point correspondence

and reduce the impact of latency during automated registration point acquisition, the registration point is recorded only when the change in position is below an empirically tuned threshold of 1 mm/s.

### 5.1.3 *Refinement Approach*

While pausing the robot's movement until optical tracking convergence ensures accurate point correspondences, it increases the time required to measure each point, thus limiting the practical number of points that can be recorded. An alternative approach would be to continuously record points in both coordinate systems during movement and use those *path points* for point-based registration. However, tracking latency may reduce the reliability of point correspondences. To mitigate this, correspondences could be estimated by defining them based on the closest points between the two point sets. To improve accuracy further, these correspondences can be recalculated iteratively after applying the transformation, followed by recalculating the transformation itself. This method corresponds to the ICP algorithm, commonly used for surface-based registration [33]. As the ICP requires an initial transformation, *path points* will be used for registration refinement. First, point-based registration is conducted, followed by refinement using the ICP applied to points recorded along the paths between registration points.

### 5.1.4 *Implementation*

The implementation followed the approach described in Section 2.3. All tracking techniques used in this experiment were adapted for Unity version 2022.3.16. Since *Vuforia* does not support adjustment of camera intrinsics, the intrinsics used here were integrated into the other methods. The robot was controlled using high-stiffness Cartesian impedance control to ensure precise motion.

## 5.2 EVALUATION METHODS

To investigate the influence of different parameters on point-based XR-to-robot registration accuracy, a technical evaluation consisting of eight independent experiments was conducted. The first experiment evaluated a baseline condition with predefined registration characteristics, using a one-factorial design to assess the accuracy of the tracking techniques. This baseline configuration was retained as the reference in each of the following seven experiments, where a single characteristic was varied while all other parameters remained at baseline. For each characteristic, registration was performed using all tracking techniques, resulting in a two-factorial design.

### 5.2.1 *Variables*

INDEPENDENT VARIABLES In the first experiment, a one-factorial design was employed to investigate the *tracking technique*, including ARToolKit, ArUco, Retro-Reflective, and Vuforia. The remaining experiments used a two-factorial design, examining the *tracking technique* and one additional *registration characteristic*. The investigated registration characteristics included:

Experiment 2: **Viewing Angle** *0°*, 15°, or 30°.
Experiment 3: **Viewing Distance** 0.66 m, *5 m*, 1.33 m
Experiment 4: **Marker Size** 5 cm, *10 cm*, or 15 cm
Experiment 5: **Point Distance** 7.5 cm, *15 cm*, or 30 cm
Experiment 6: **Point Distribution** *equidistant*, coplanar, or random
Experiment 7: **Amount of Registration Points** *4* or 8 points
Experiment 8: **Refinements** *no refinement*, path points, or path points+ICP

These parameters were individually varied, with all other characteristics kept at the baseline configuration. The baseline parameters are indicated in italics above. Regarding point distributions: *Equidistant* setup consisted of a set of four evenly spaced points, with the set rotated across all axes to cover a spherical region. *Coplanar* setup rotated a coplanar point set within the three principal planes, constraining points to coplanar sections. In *random* setup, the points were randomly distributed across the spherical region (see Figure 5.2).

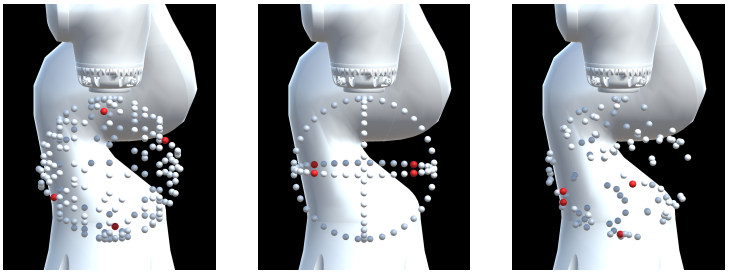


Figure 5.2: Point distributions investigated, with one exemplary point set highlighted in red. From left to right: equidistant, coplanar, random. Reprinted from [4], © 2025 IEEE.

DEPENDENT VARIABLES As detailed in Section 2.2, the most important measure for assessing registration accuracy is the TRE, which quantifies the error of target points not used during registration. To determine the TRE, the positions of four target points were measured in both the robot and XR coordinate systems. After registration, the resulting transformation was applied to the robot coordinate points, and the Euclidean distance to the corresponding XR coordinate points was calculated.

While the TRE provides insight into registration accuracy, it is also influenced by the tracking accuracy during measurement of the target points. To increase the reliability of the TRE, more target points could be used; however, this would slow down data acquisition and reduce practicality. Alternatively,

reliability can be improved by carefully selecting robust target points prior to the experiment. To identify points with high tracking accuracy, different candidate points were measured using the robot, the XR device, and an external tracking camera. By comparing the distances between these points across the three systems, those with minimal distortion, i.e., points whose distances are consistent in all tracking systems, were identified and used as target points.

### 5.2.2 *Hypothesis*

As the experiments were exploratory, no a priori assumptions were made about the outcomes. Therefore, two-sided hypotheses were investigated. The null and alternative hypotheses considered were:

$H_0^y$ The mean TRE is equal for all levels of $y$.
$H_1^y$ The mean TRE differs for at least one level of $y$.

$$y \in \begin{cases} \text{1-Tracking Technique} \\ \text{2-Viewing Angle} \\ \text{3-Viewing Distance} \\ \text{4-Marker Size} \\ \text{5-Point Distance} \\ \text{6-Point Distribution} \\ \text{7-Point Amount} \\ \text{8-Refinement} \end{cases}$$

### 5.2.3 *Procedure*

For data acquisition, the *HoloLens* was mounted on a tripod and positioned in front of the robot with the height fixed at 1.70 m throughout the experiment. The end effector was then moved to the four target points, followed by the registration points, with constant orientation. For each tracking method and investigated characteristic, 27 different point sets were considered. The use of 27 point sets was based on the coplanar point distribution, where four equidistant points were placed on one of the three principal planes. To uniformly cover possible orientations within each plane, this base configuration was rotated in 10° increments, resulting in 9 variations per plane. Across the three principal planes, this yielded a total of 27 distinct point sets.

### 5.2.4 *Statistical Analysis*

To evaluate the effects of the one-factorial first experiment, the data were analyzed using a repeated-measures ANOVA via the `aov_ez()` function from the `afex` package. Paired $t$-tests with Bonferroni corrections for multiple comparisons were conducted as post hoc tests.

For all other experiments, the data were analyzed using a two-way repeated-measures ANOVA to assess the interaction between tracking techniques and the respective registration parameter. Paired $t$-tests with Bonferroni corrections were again conducted as post hoc tests in cases of significant ANOVA results. For significant interaction effects, contrasts were computed within each tracking method cell using the `emmeans` package. For the viewing

distance analysis, a missing factor-level combination required fitting a linear mixed-effects model using the `lmer()` function from the `lme4` package. To evaluate the significance of fixed effects, a Type III Wald chi-square test was conducted using the `Anova()` function from the `car` package.

Greenhouse-Geisser corrections were applied in all repeated-measures analyses if a violation of sphericity was detected via Mauchly's test. All statistical analyses of measured TRE values were performed in R (version 4.4.0).

## 5.3 RESULTS

The following section presents the results for the experiments investigating the different registration characteristics. An overview of all descriptive results is provided in Table A.1, and a summary of the statistical analyses can be found in Table A.2. Pairwise comparison results are illustrated in the plots.

### 5.3.1 *Tracking Accuracy*

A significant main effect for tracking accuracy was found (test statistic $F = 259$, $p < 0.01$, effect size $\eta^2 = 0.882$). Significant pairwise comparison results for tracking accuracy are shown in Figure 5.3. TRE results across tracking methods for the baseline condition indicated that ARToolKit ($2.15 \pm 0.64$ mm) yielded significantly lower TREs compared to ArUco ($27.71 \pm 7.77$ mm), Vuforia ($3.23 \pm 1.20$ mm), and retro-reflective tracking ($3.87 \pm 1.19$ mm). Thus, $H_0^1$ can be rejected in favor of the respective alternative hypothesis $H_1^1$. ArUco tracking produced significantly worse TRE than all other methods. Given its substantially higher TRE (almost ten times greater), ArUco tracking was excluded from further analysis.

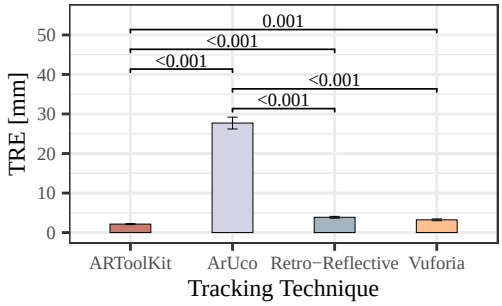



Figure 5.3: TRE for tracking techniques in baseline condition. Bars represent means. Error bars represent standard errors. Brackets indicate statistically significant post hoc pairwise comparisons. Adapted from [4], © 2025 IEEE.

The significant differences between the tracking techniques are also reflected by significant main effects of the tracking techniques on TRE in all subsequent two-way analyses. However, since the relationship between the methods remains consistent and the focus is on registration characteristics, these main effects will not be further discussed.

### 5.3.2 *Viewing Angle*

No significant viewing angle main effects ($F = 3.43$, $p = 0.068$) or interaction effects ($F = 3.30$, $p = 0.057$) on the TRE were observed. The null hypothesis $H_0^2$ could not be rejected. Descriptively, TRE was generally lower at 0° compared to 30° across all tracking algorithms. At 15°, ARToolKit's accuracy was intermediate, Vuforia showed a notably lower TRE compared to 0° and 30°, while retro-reflective tracking exhibited higher TRE (see Figure 5.4).

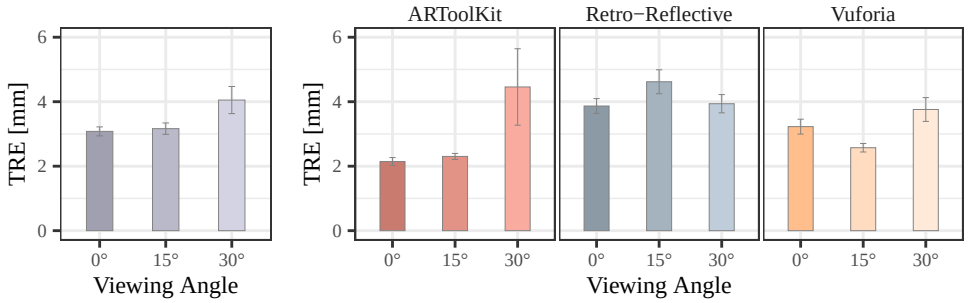



Figure 5.4: TRE dependence on viewing angles (left: non-significant *viewing angle* main effect; right: non-significant interaction effect). Bars represent means. Error bars represent standard errors. Adapted from [4], © 2025 IEEE.

### 5.3.3 *Viewing Distance*

As the retro-reflective tracking failed at the greatest distance (1.33 m), only the TREs for the remaining two distances were analyzed for this tracking method. There were no significant viewing distance main effects ($\chi^2 = 0.012, p = 0.994$) or interaction effects ($\chi^2 = 6.49, p = 0.09$) on the registration accuracy. Thus, null hypothesis $H_0^3$ could not be rejected. Descriptively, ARToolKit's registration accuracy was consistent across distances, whereas Vuforia's TRE increased with distance. Retro-reflective tracking deteriorated at closer ranges and failed at greater distances (see Figure 5.5).

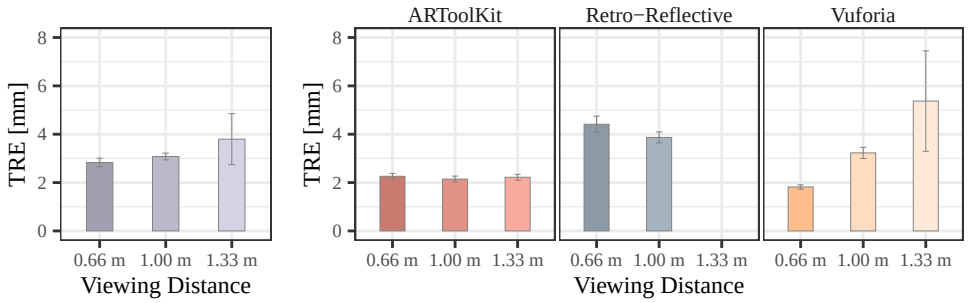



Figure 5.5: TRE dependence on viewing distances (left: non-significant *viewing distance* main effect; right: non-significant interaction effect). Bars represent means. Error bars represent standard errors. Adapted from [4], © 2025 IEEE.

### 5.3.4 *Marker Size*

A significant main effect of marker size on TRE was found ($F = 14.49$, $p < 0.001$, $\eta^2 = 0.109$), while there was no significant interaction effect ($F = 3.21$, $p = 0.081$). The main effect is depicted on the left of Figure 5.6. Post hoc pairwise comparisons showed that the smallest marker size (5 cm) resulted in significantly higher TRE values compared to both the medium (10 cm) and large (15 cm) markers. Therefore, the null hypothesis $H_0^4$ can be rejected in favor of the alternative hypothesis $H_1^4$.

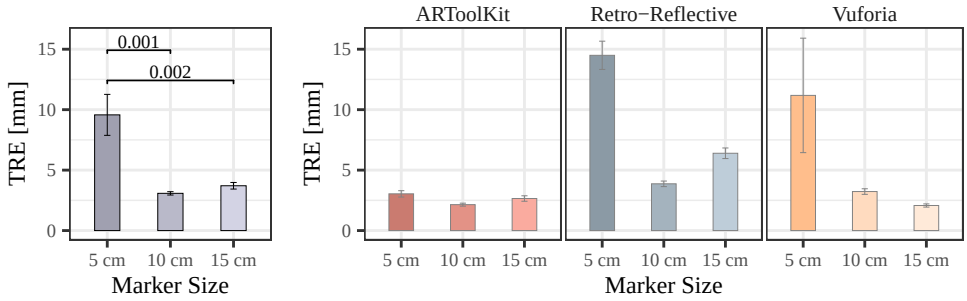



Figure 5.6: TRE dependence on marker sizes (left: *marker size* main effect; right: non-significant interaction effect). Bars represent means. Error bars represent standard errors. Brackets indicate statistically significant post hoc pairwise comparisons. Adapted from [4], © 2025 IEEE.

### 5.3.5 *Point Distance*

A significant main effect of point distance ($F = 62.43$, $p < 0.001$, $\eta^2 = 0.338$) and an interaction effect between point distance and tracking technique ($F = 62.38$, $p < 0.001$, $\eta^2 = 0.514$) on TRE were observed (see Figure 5.7). Therefore, $H_0^5$ can be rejected in favor of the corresponding alternative hypothesis $H_1^5$.

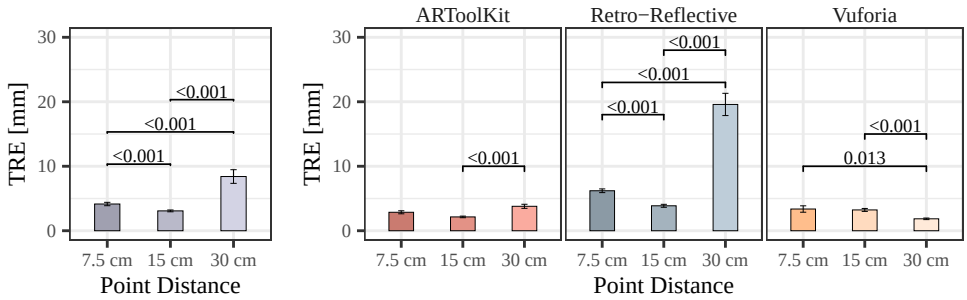



Figure 5.7: TRE dependence on point distances (left: *point distance* main effect; right: interaction effect). Bars represent means. Error bars represent standard errors. Brackets indicate statistically significant post hoc pairwise comparisons. Adapted from [4], © 2025 IEEE.

Pairwise comparisons for the main effect indicate that a point distance of 15 cm yielded significantly lower TRE than the other two distances, while a point distance of 30 cm resulted in significantly higher TRE than the others.

However, the significant interaction effect suggests a dependency on the tracking technique, challenging the isolated interpretation of the main effect.

For ARToolKit, the TRE was significantly higher at a point distance of 30 cm compared to 15 cm. Similarly, retro-reflective tracking performed worst at 30 cm, with 7.5 cm also resulting in higher TRE than 15 cm. In contrast, Vuforia showed the opposite trend, achieving significantly better registration accuracy at 30 cm than at 15 cm

### 5.3.6 *Point Distribution*

Significant point distribution main effects ($F = 16.63$, $p < 0.001$, $\eta^2 = 0.106$) and interaction effects ($F = 12.72$, $p < 0.001$, $\eta^2 = 0.196$) were found on registration accuracy (see Figure 5.8). Therefore, $H_0^6$ can be rejected in favor of the respective alternative hypothesis $H_1^6$.

Regarding the main effect, pairwise comparisons revealed significantly higher TRE for the random distribution compared to the other two. However, the presence of a significant interaction effect again highlights a synergy between point distribution and tracking method.

For ARToolKit, coplanar point distribution yielded significantly lower TREs compared to equidistant and random point placements. Retro-reflective tracking showed significantly higher TREs for random point placement than both equidistant and coplanar setups. No significant pairwise differences were observed for Vuforia.

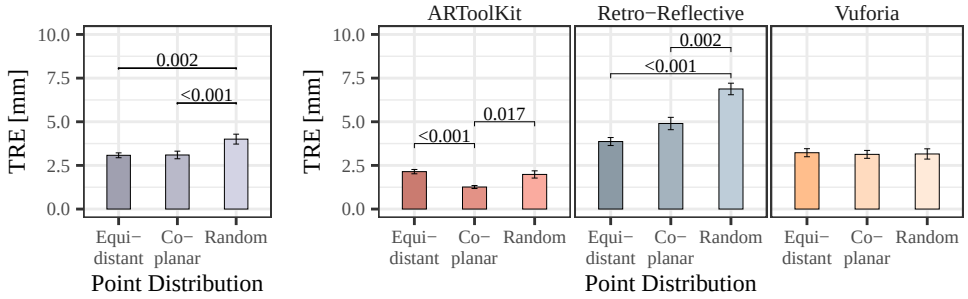



Figure 5.8: TRE dependence on point distributions (left: *point distribution* main effect; right: interaction effect). Bars represent means. Error bars represent standard errors. Brackets indicate statistically significant post hoc pairwise comparisons. Adapted from [4], © 2025 IEEE.

### 5.3.7 *Amount of Points*

While there was no significant main effect of the number of points ($F = 0.51$, $p = 0.43$), significant interaction effects ($F = 7.16$, $p = 0.007$, $\eta^2 = 0.087$) were found on TRE. Thus, $H_0^7$ can be rejected in favor of the respective alternative hypothesis $H_1^7$. For ARToolKit, the TRE was significantly lower with 8 registration points. While Vuforia showed a lower mean TRE with 8 points, it also produced more outliers, and the difference was not statistically significant.

Retro-reflective tracking, on the other hand, exhibited a significantly higher TRE with 8 points compared to 4.

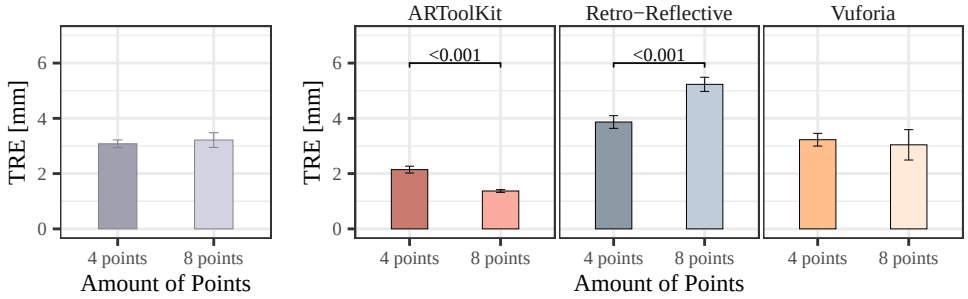



Figure 5.9: TRE dependence on point quantities (left: non-significant *amount of points* main effect; right: interaction effect). Bars represent means. Error bars represent standard errors. Brackets indicate statistically significant post hoc pairwise comparisons. Adapted from [4], © 2025 IEEE.

### 5.3.8 *Refinement Approaches*

To refine the point-based registration results, the use of additional points recorded during robot movement as paired points or as point clouds for ICP was analyzed as described in Section 5.1.3. Significant registration approach main effects ($F = 23.84$, $p < 0.001$, $\eta^2 = 0.046$) were found (see Figure 5.10). Thus, $H_0^8$ can be rejected in favor of the respective alternative hypothesis $H_1^8$. Using ICP significantly improved TRE compared to both using only four registration points and incorporating additional points as paired points. No significant interaction effects were found ($F = 1.25$, $p = 0.297$).

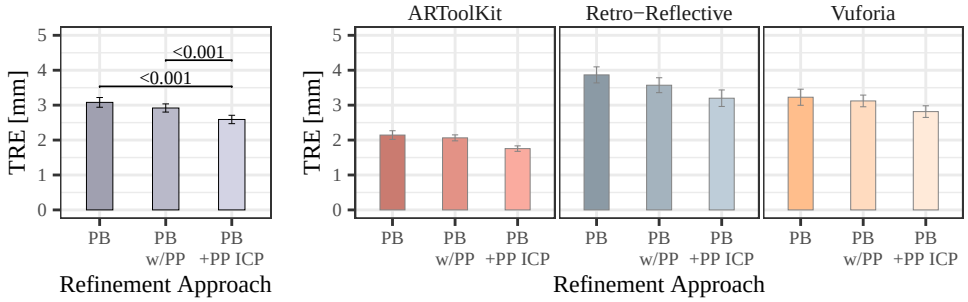



Figure 5.10: TRE dependence on refinement approaches (left: *refinement approach* main effect; right: non-significant interaction effect; PB: point-based registration with four points; PB /w PP: point-based with four points and path points; PB + PP ICP: point-based with four points refined by ICP using path points). Bars represent means. Error bars represent standard errors. Brackets indicate statistically significant post hoc pairwise comparisons. Adapted from [4], © 2025 IEEE.

## 5.4 DISCUSSION

In this section, findings are analyzed and compared with related work, summarized in Table 5.1. While different measures for registration accuracy are employed, these previous results can be used as general reference.

Table 5.1: Reported registration accuracy values from related work (*italics* indicate the accuracy metric used; accuracy reported as M ± SD; NR = not reported). Reprinted from [4], © 2025 IEEE.

| | Source | Accuracy | Registration Method |
|---|---|---|---|
| **XR-to-robot Registration** | this | 1.76 ±0.40 mm | Point-based using 4 points with refinement |
| | | *TRE using optical tracking and multimodal marker* | |
| | [Ch. 4] | 43.23 ± 14.22 mm | Manual alignment of virtual robot |
| | | 12.13 ± 6.42 mm | Point-based using 4 points (physical points) |
| | | 14.86 ± 17.22 mm | Point-based using 4 points (Vuforia) |
| | | *RMSE of three points on real/virtual robot* | |
| | [110] | 30.2 ± 23.9 mm | Manual alignment of virtual robot |
| | | 16.5 ± 11.0 mm | Manual alignment + Reflective AR display |
| | | *Manually placed virtual objects as ground truth* | |
| | [178] | 30.0 ± 9.3 mm | Manually placed Vuforia marker |
| | [94] | 10 ± NR mm | Point-based, tracked HoloLens |
| | | *RMSE between closest points of point clouds* | |
| | [258] | 27.53 ± 44.51 mm | Manual alignment of cube with robot base |
| | | 4.91 ± 1.39 mm | ICP with depth point cloud |
| | | *Translational error compared to ground truth* | |
| | [241] | 9 ± NR mm | ICP with depth point cloud |
| **Workspace Registration** | | *TRE using optical tracking* | |
| | [211] | 8.82 ± 4.85 mm | Point-based using 3 points (150 mm, Robot 1) |
| | | 4.64 ± 1.23 mm | Point-based using 3 points (150 mm, Robot 2) |
| | | 11.48 ± 6.24 mm | Point-based using 3 points (300 mm, Robot 1) |
| | | 2.71 ± 0.94 mm | Point-based using 3 points (300 mm, Robot 2) |
| | | *TRE using physically defined points* | |
| | [325] | 1.01 ± 0.20 mm | Point-based using 3 points (250 × 175 $mm^2$) |
| | | 0.70 ± 0.11 mm | Point-based using 3 points (750 × 525 $mm^2$) |
| | | 0.59 ± 0.09 mm | Point-based using 5 points (750 × 525 $mm^2$) |
| | | 0.44 ± 0.05 mm | SA, Point-based using 3 points (250 × 175 $mm^2$) |
| | | 0.41 ± 0.35 mm | SA, Point-based using 3 points (750 × 525 $mm^2$) |
| | | 0.38 ± 0.03 mm | SA, Point-based using 5 points (750 × 525 $mm^2$) |
| | | 0.42 ± 0.02 mm | Clustering |
| **Sensor** | | *TRE for 97 evenly sampled points* | |
| | [228] | 1.72 ± NR mm | Point-based using 3 points (camera left) |
| | | 2.13 ± NR mm | Point-based using 3 points (camera centered) |
| | | 1.76 ± NR mm | Point-based using 3 points (camera right) |
| | | *TRE using commanded and reached robot position* | |
| | [185] | 21.4 ± 12.3 mm | DART [288] using depth camera |
| | | 28.2 ± 24.1 mm | Point-based using 10 points (ArUco marker) |
| | | 27.4 ± 4.70 mm | Single RGB camera image |

TRACKING METHOD Comparing the results of this experiment with those reported for manual registration [110, 178, 258] reveals that point-based methods enable more accurate registration. This aligns with the findings

in Chapter 4. Although ArUco tracking performed significantly worse than the other methods in this study, the resulting TRE is comparable to manual approaches and consistent with the point-based accuracy reported for ArUco by Lee et al. [185]. Notably, point-based registration using the other tracking techniques achieved accuracies comparable to, or even exceeding, those reported for point-based registration with infrared cameras [228], as well as surface-based methods using depth [185, 258] and RGB cameras [185].

VIEWING ANGLE Previous research on infrared tracking cameras for sensor registration suggested that an angled view of the robotic workspace can improve registration accuracy [228]. This finding is attributed to the viewing angle altering the distance of the measured points, thereby influencing whether the points fall within the optimal tracking range. In contrast, studies on tracking accuracy with Vuforia and ARToolKit have shown that tracking accuracy depends not only on distance but also significantly on the viewing angle itself [74], indicating that image marker tracking may be affected by factors beyond just distance. However, unlike previous findings, this study did not reveal significant statistical effects of viewing angle on TRE. However, since descriptive differences were observed, future research should consider a broader range of data, incorporating both varying distances and angles simultaneously, to obtain more comprehensive results.

VIEWING DISTANCE AND MARKER SIZE The viewing distance and marker size are related, as changes in viewing distance alter the apparent size of the marker. The results show that larger markers enable more accurate registration, consistent with research linking reduced tracking error to shorter HoloLens-marker distances [74]. However, this effect does not translate to the analysis of viewing distance. This suggests that other factors might mitigate the observed effects. One explanation is the location of the marker within the FOV. When the marker is positioned closer to the XR device, it appears larger but may also lie near the edges of the FOV, where tracking accuracy tends to decrease if the intrinsic calibration does not fully compensate for lens distortion [319, pp. 60–61]. This may lead to increased FLE, which could negatively impact overall registration accuracy.

POINT DISTANCE Related work on image-guided surgery [126, 341] and robot-to-workspace registration using 3D-printed seats [325] suggests that increasing the distance between registration points can improve registration accuracy. In this experiment, this holds true only for changes in distance between 15 cm and 30 cm for Vuforia, and between 7.5 cm and 15 cm for retro-reflective tracking. In contrast, for both ARToolKit and retro-reflective tracking, decreasing the distance from 30 cm to 15 cm improves tracking accuracy. This may again be attributed to tracking inaccuracies at the edges of the HoloLens's FOV, or to reduced tracking performance of the robotic arm when operating in extended positions [324]. This aligns with previous findings that reported decreased registration accuracy for points located farther from the robot base [228].

POINT DISTRIBUTION The main effect and the interaction effect for ARToolKit and Vuforia indicated that random point distribution resulted in a significantly higher TRE compared to coplanar and equidistant placements. Random placement is the only method where the centroid of the registration points does not necessarily coincide with the centroid of the target points used to calculate TRE. This finding aligns with previous research suggesting that positioning the centroid of the registration points close to the centroid of the target points improves accuracy [126, 341].

AMOUNT OF POINTS The results for ARToolKit indicate that increasing the amount of registration points reduces TRE, consistent with related work [126, 325, 341]. However, previous studies have also highlighted that increasing the number of points can introduce outliers [96, 126]. For retro-reflective tracking, increasing the number of points increased TRE, suggesting that outliers may have negatively affected the accuracy.

REFINEMENT APPROACHES To refine registration results, *path points*, which are recorded during the movement between registration points, were integrated. Unlike static registration points, where the robot halted until tracking converged to a stable state to mitigate latency effects, the path points were collected continuously. While including path points in point-based registration did not significantly improve registration, using them for an ICP-based refinement after point-based registration with the registration points significantly enhanced accuracy. This suggests that iteratively updating the correspondence of path points improves registration outcomes by mitigating the effects of incorrect correspondences due to latency and outliers.

COMPARISON TO CHAPTER 4 As shown in Table 5.1, the TRE of the point-based registrations presented in Chapter 4 was significantly higher than the TRE reported in this chapter. This improvement might be attributed to the optimization of the registration parameters. In Chapter 4, participants were free to place registration points as they wished, which may have led to suboptimal parameter choices. However, it is worth noting that even the worst parameter configurations presented in this chapter resulted in lower TRE values than those in Chapter 4. One possible reason for this difference in registration accuracy could be tracking inaccuracies in Chapter 4, caused by the use of cylindrical markers and object tracking instead of planar image markers, or by relying on manual confirmation of tracking accuracy rather than automatic detection of tracking convergence. Another contributing factor may be the difference in how TRE was calculated in the two experiments. In the first experiment, a ground truth was established to project target points, whereas in the second experiment, target points were directly measured. Therefore, tracking errors affecting the placement of the ground truth, as well as the selection of target points at the edges of the workspace, might have amplified registration errors in the first experiment.

Nevertheless, applying the findings of this chapter to the interactive registration methods presented in Chapter 4 is expected to significantly improve

registration results. To this end, ARToolKit could be employed instead of Vuforia, and the refinement using path points could be incorporated. Additionally, the findings on registration point placement could be used to either enable automatic movement to registration points or provide visual guidance for manual placement.

### 5.4.1 *Limitations*

The main limitation of this experiment lies in using only one specific robot and XR device. While this approach allowed for a detailed and controlled investigation of XR-to-robot registration, it should be acknowledged that the results may be influenced by the characteristics and limitations of this setup. Future research should explore a broader range of devices to assess the generalizability of presented findings and the effects of different hardware.

This experiment focused on commonly used, publicly available tracking approaches. However, registration accuracy could potentially be enhanced by incorporating refinement techniques such as improved marker tracking through corner or edge refinement [161], machine learning [325], bias correction [111], or the inclusion of orientation vectors [221]. Although the proposed point-based method performs well compared to depth-based point cloud approaches [258] and deep learning methods based on RGB images [185], the latter techniques offer the advantage of requiring only a single image, which can accelerate the registration process. Therefore, future research should explore deep learning approaches [160, 185] to improve efficiency further.

Another limitation arises from the experimental design: separating the characteristics under investigation into individual experiments allowed for a more controlled analysis of each factor. However, this approach does not provide insight into potential interactions between characteristics, such as simultaneous changes in viewing distance and viewing angle. Future work should investigate these combined effects.

Finally, it should be noted that this experiment's error assessment did not account for pivot calibration errors, as the end effector was maintained in a consistent orientation during both the registration and target point acquisition. This isolated the registration accuracy, allowing the results to reflect the performance of the registration method itself rather than errors in pivot calibration. While this makes the presented results valid for scenarios where the transformation between the marker and TCP is known (e.g., through end effector geometry), future work should evaluate the accuracy of pivot calibration for different tracking techniques.

### 5.4.2 *Implications*

Based on the findings reported in this chapter, several general recommendations for improving XR-to-robot registration accuracy can be proposed:

**Use ARToolKit for marker tracking.** This study revealed significantly higher registration accuracy when using ARToolKit compared to Vuforia, ArUco, and retro-reflective marker tracking. ArUco, in particular, showed substantially worse accuracy than the other methods.

**Employ larger markers.** The results indicate that marker sizes of 5 cm led to lower accuracy compared to markers of 10 cm and 15 cm.

**Use path points for refinement.** Registration accuracy improved significantly when additional points were recorded during movement between registration targets and surface-based refinement techniques were applied.

In addition, the conducted experiments revealed that some factors are dependent on the tracking algorithm used. This leads to the following method-specific recommendations. As previously noted, ArUco was excluded from further analysis due to its substantially higher registration errors.

**ARToolKit.** To enhance registration accuracy, the viewing angle should be below 30°, the point distance should be approximately 15 cm, and the registration points should be arranged coplanar. Additionally, increasing the number of registration points significantly improves accuracy.

**Vuforia.** Vuforia is less sensitive to point distribution and quantity. However, the distance between the XR device and the registration points should be minimized, and point distances should be kept below 30 cm.

**Retro-reflective markers.** The viewing distance should be kept short, as this approach performs poorly at larger distances. Registration point spacing should be around 15 cm, and the centroid of the registration points should align with the centroid of the target points. Additionally, using four registration points instead of eight resulted in better performance.

## 5.5 CONCLUSION

In this chapter, an extensive investigation on the impact of different tracking techniques and registration parameters on the accuracy of XR-to-robot registration was presented. Overall, the best registration accuracy was achieved using ARToolKit tracking. While no significant effects of viewing angle and distance on registration accuracy were observed, larger markers and placing the centroid of the registration points near the centroid of the targets consistently led to more accurate results. The influence of other factors, such as point distance, amount, and distribution, was highly dependent on the tracking method used. The experimental results also demonstrated that a simple yet effective refinement approach, i.e., collecting points along the paths between registration points and using them for registration refinement via point cloud registration, significantly improved registration accuracy. While future work is needed to assess the transferability of the findings to other XR devices and robots, these results can serve as general guidelines for XR-to-robot registration.

# 6 SUMMARY

The first part of this thesis comprised two investigations into XR-to-robot registration, a crucial step for enabling both hand gesture interaction and seamless XR integration within robotic workspaces. In Chapter 4, three registration methods with potential for enabling efficient registration were identified based on related work. These manual and point-based methods were evaluated regarding accuracy, duration, and perceived workload in a comparative user study. The results indicated that point-based approaches could achieve more accurate registration while reducing user workload.

However, the accuracy achieved by point-based approaches in that study was not sufficient for high-precision scenarios. Therefore, Chapter 5 presented a follow-up technical investigation into parameters that may influence registration accuracy. By evaluating four commonly used tracking techniques and varying parameters such as viewing angle and distance, marker size, point distance, distribution, and number of points, several influential factors were identified and distilled into practical guidelines. In addition, a simple yet effective refinement approach was proposed, which utilized points recorded during movement between registration points for an additional refinement step based on point cloud registration. This refinement strategy, combined with careful parameter optimization, led to significantly improved accuracy compared to the initial experiment. Moreover, the achieved registration accuracies were higher than those reported in related work for manual, point-based, and surface-based XR-to-robot registration.

By first identifying point-based registration as a more accurate approach with lower user workload than manual registration, and subsequently improving its accuracy through a technical evaluation, the findings of the first part of this thesis contribute to more efficient XR-to-robot registration. As the implemented registration methods can be readily integrated into existing frameworks, the results can support practitioners in accurately leveraging XR for both visualization and intuitive input within robotic workspaces, representing an important step towards more efficient HRI.

# Part II

# HAND GESTURE-BASED HUMAN-ROBOT INTERACTION

**Synopsis:** This part addresses two challenges associated with hand gesture-based HRI through three experiments. The first two experiments focus on compensating for the lack of haptic feedback by employing force automation and sensory substitution. The third experiment investigates sensor placement and control modes to alleviate fatigue during hand gesture-based HRI.

This part contains material adapted from the following publications:

Tonia Mielke, Marilena Georgiades, Oliver S. Großer, Maciej Pech, Christian Hansen, and Florian Heinrich. "Enhancing Gesture-Based Human-Robot Interaction: Investigating the Role of Force Automation." In: *Proceedings of the 21st ACM IEEE International Conference on Human-Robot Interaction (HRI '26)*. Edinburgh, Scotland, UK: Association for Computing Machinery, 2026, pp. 1–10. DOI: 10.1145/3757279.3785568 [3].

Tonia Mielke, Florian Heinrich, and Christian Hansen. "SensARy Substitution: Augmented Reality Techniques to Enhance Force Perception in Touchless Robot Control." In: *Transactions on Visualization and Computer Graphics (TVCG)* 31.5 (Mar. 2025), pp. 3235–3244. DOI: 10.1109/TVCG.2025.3549856 [6].

Tonia Mielke, Florian Heinrich, and Christian Hansen. "Gesturing Towards Efficient Robot Control: Exploring Sensor Placement and Control Modes for Mid-Air Human-Robot Interaction." In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 19–23. DOI: 10.1109/ICRA55743.2025.11127519 [5]. © 2025 IEEE.

# 7 INTRODUCTION

## 7.1 INTRODUCTION

With advancements in artificial intelligence, machine learning, and sensing technologies, robots are increasingly capable of sensing, planning, and acting autonomously [194]. This raises the question: *Why do we still need HRI? Can't tasks simply be automated?* According to Sherwani et al. [301], the answer to this question is that we need HRI because "the strengths of humans are the weakness of robots and the other way around". While robots excel at repeatability and physical strength, they still often lack the flexibility and adaptability required in dynamic work environments [93]. HRI can bridge this gap by enabling humans to teach robots new tasks or take manual control when tasks are too complex or risky to be fully automated [266, 301, 361]. This includes handling failure situations [52], as well as providing control during risk-prone scenarios such as in medicine [361] or space robotics [125].

To enable such HRI, traditional robot control methods employ physical controllers or direct hand-guiding of the robot [35, 306]. However, more recently, NUIs have gained attention in HRI (see Section 2.1.2) [44, 183]. NUIs rely on interaction approaches familiar from everyday life, enabling intuitive and direct communication between humans and robots [256, p.472]. This includes modalities such as hand gestures, head movements, speech, and gaze [334]. In particular, hand gestures have shown potential for enabling efficient interaction in other application areas, such as interaction with medical images [137, 138], large displays [21], and XR applications [192, 364].

While the general use of hand gestures for robot control has been explored in related work, several challenges remain. Therefore, this chapter builds on prior research to systematically address gaps in hand gesture-based HRI. Specific challenges associated with this control method are identified and addressed to answer RQ2.

**RQ2** | How can efficient mid-air gesture-based HRI be designed?

## 7.2 RELATED WORK

Comparative studies have been conducted across different application domains to assess the efficiency of hand gesture interaction for HRI (see Table 7.1). For instance, Lee et al. [183] investigated hand gestures for robotic arm control in comparison to teach pendants and physical hand-guiding. Their findings indicated that hand gesture interaction allowed for faster task completion and was associated with lower perceived workload. Wu et al. [356] compared whole-body teleoperation, including hand gestures, with

joystick-based control in a remote interaction scenario. Their results found better performance for the whole body interface.

While these works highlight the potential advantages of robot control through hand gestures, other works highlight existing challenges. Zhou et al. [373] investigated different interaction methods for robotic surgery. In a first experiment, they compared hand gestures with upper body movement, keyboard input, a haptic device, and a handheld controller across multiple tasks, finding that the methods performed comparably. In a second experiment focused on a more complex task, they compared hand gestures to a haptic device, finding that the haptic device yielded superior performance. Similarly, Saren et al. [282] evaluated robot control using eye gaze, hand gestures, and voice commands, compared to a touchpad, in a pick-and-place task under both single-task and dual-task conditions. While touchpad and hand gesture interactions generally enabled faster task execution than the other methods, the perceived workload depended on the presence of a secondary task. Whitney et al. [343] compared hand-guiding to control via keyboard and mouse, as well as hand gestures. Their results indicate that while hand gesture control was faster and associated with lower workload than keyboard control, hand-guiding resulted in the lowest task duration and workload. In the context of surgical robotics, Borgioli et al. [44] and Chen et al. [61] both compared hand gesture-based interaction with standard controller-based control for the da Vinci surgical system. Borgioli et al. found that while hand gestures enabled intuitive and precise interaction, the traditional controller still outperformed them. In contrast, Chen et al. reported comparable performance between the two modalities.

Furthermore, the general feasibility of hand gesture control was investigated in two co-authored publications [8, 9]. In the first work [9], hand gesture control for robotic ultrasound was qualitatively evaluated. The findings suggested that while the interaction was intuitive and precise, challenges remain regarding the interaction space, fatigue, haptic perception, and involuntary vertical motion. In the second work [8], hand-guiding, haptic teleoperation, and hand gesture control were compared for an abstract robotic ultrasound task. It was found that hand-guiding enabled more accurate and faster manipulation, while haptic input offered slightly better usability and lower workload. Again, challenges regarding haptic perception and workload for hand gesture interaction were identified.

As summarized in Table 7.1, existing studies indicate that gesture-based interaction holds promise, reporting benefits such as naturalness [373], intuitiveness [183, 199, 356], flexibility [61], speed [183, 282], and the advantage of requiring less bulky and expensive hardware [44]. However, several studies also highlight remaining challenges that impact the performance of hand gesture-based control. These include issues such as user fatigue [44, 282, 373], missing visual or sensory feedback [183, 8], involuntary vertical hand movement leading to higher contact forces [9, 373], and limited tracking space [373]. These findings therefore motivate further research into the design and usability of hand gesture-based interfaces for HRI.

Table 7.1: Overview of studies on hand gesture-based robot control across different robotic application domains.

| Application | | Benefits | Challenges |
|---|---|---|---|
| Pick-and-Place | [183] | • Fast task completion<br>• Low workload<br>• Intuitive and easy | • Missing visual and sensory feedback |
| | [343] | • Faster & lower workload than keyboard control | • Higher duration & workload than hand-guiding |
| | [282] | • Fast | • Physically Demanding |
| Object Manipulation | [356] | • Precise & Intuitive<br>• Not mentally demanding<br>• Fast task completion | *none reported* |
| Surgery | [373] | • Comparable performance for simple tasks<br>• Natural | • Involuntary interaction increases contact force<br>• Leads to fatigue<br>• Limited interaction space |
| | [44] | • Intuitive & precise<br>• Less expensive and bulky | • Fatigue due to arm posture |
| | [61] | • Comparable performance to controllers<br>• High Flexibility | *none reported* |
| Ultrasound | [9] | • Intuitive<br>• Precise<br>• Easy to learn | • Involuntary vertical movement causing contact force<br>• Interaction space<br>• Contact force<br>• Fatigue |
| | [8] | *none reported* | • Missing sensory feedback<br>• High workload |

## 7.3 CONTRIBUTION

To address the issues associated with hand gesture interaction for robot control, this part of the thesis focuses on the two major challenges identified above: the lack of haptic feedback and user fatigue. In this context, it investigates RQ2 through the following contributions, based on three conducted experiments:

**Investigation of Partial Automation.** Section 8.2 explores an approach in which contact force is autonomously regulated while the remaining DOFs are manually controlled. This setup is used to study the potential of partial automation in easing HRI.

**Assessment of Hand Gesture for Robot Control.** Through a comparison between hand gesture-based robot control and the state-of-the-art method of hand-guiding, Section 8.2 comparatively evaluates the potential of gesture-based interaction across tasks requiring different levels of precision.

**Evaluation of Sensory Substitution Modalities.** Section 8.3 examines different sensory substitution modalities in XR, investigating the suitability of visual and vibrotactile feedback for conveying force information.

**Investigation on Visual Feedback Design.** To further investigate the characteristics of visual feedback, Section 8.3 explores different visual feedback designs and XR placements for sensory substitution of force information.

**Analysis of Sensor Placement.** Aiming to reduce fatigue during hand gesture-based HRI, Chapter 9 investigates different sensor placements for tracking hand gestures for robot control.

**Investigation of Control Modes.** Additionally, Chapter 9 explores different control modes for mapping hand gestures onto robot motion, focusing on position and rate control, and evaluates their efficiency for HRI.

# 8 FORCE ASSISTANCE

## 8.1 INTRODUCTION

As previously described, the loss of direct haptic perception poses a challenge in HRI [235, 9]. Integrating haptic feedback has proven to be a valuable tool for enhancing situational awareness by providing tactile cues [95]. While haptic input devices can deliver direct feedback by mimicking the robot's sensory responses, such solutions are not always feasible due to task-specific constraints such as spatial limitations or high associated costs [214, 235].

As a result, alternative approaches have emerged, one of which is the automation of contact force regulation. This involves detecting external forces and automatically adapting the probe's position accordingly. As this only requires the translation in one DOF, the force and position control can be decomposed, allowing users to control the remaining DOFs manually [267, 370]. This method has been shown to improve contact force consistency and enable safe interaction [100, 370].

Another approach is sensory substitution, in which tactile perception is mapped onto other sensory modalities. Force perception can be conveyed through auditory [157], vibrotactile [60], visual [308], or multimodal [145] stimuli. Unlike direct force feedback, sensory substitution allows small forces to be perceived at high stimulation levels without interfering with the operator's movements, thereby maintaining system stability [212]. Sensory substitution has been shown to reduce overall applied force [122], help maintain consistent contact force [314], and improve both stability and precision [374], even when used alongside haptic feedback.

Although both contact force automation and sensory substitution have demonstrated benefits for robot control during contact-intensive tasks, their potential has not been assessed in the context of hand gesture-based robot control. To address this gap, this chapter investigates the suitability of these two force assistance approaches for supporting users during contact-intensive tasks in HRI using hand gestures, aiming to answer RQ2.1.

> **RQ2.1** | Can force assistance improve efficiency in HRI?

### 8.1.1 *Related Work*

This section provides an overview of previous work on force assistance approaches. Specifically, it presents research on contact force automation in general HRI. Since sensory substitution of contact force is also relevant in areas beyond robot control, such as XR and surgical settings like laparoscopy, related work in these fields is also discussed.

#### 8.1.1.1 *Contact Force Automation*

To integrate contact force automation with interactive control, Raina et al. [267] and Zhao et al. [370] proposed hybrid position-force control approaches that automatically regulate contact force while allowing interactive control of the remaining DOFs via a haptic input device. These approaches were reported to have high user acceptance [267] and to offer safer, more convenient, and more reliable control compared to fully teleoperated systems [370]. Seitz et al. [297] investigated the remote control of robotic ultrasound using a joystick while maintaining stable contact pressure. Although they hypothesized that force automation improves interaction intuitiveness, they did not conduct a human-centered evaluation. Fang et al. [100] introduced different control regimes using admittance control to amplify user-applied forces, assist in smoothly reaching desired forces, and restrict motion when excessive forces are applied. Their evaluation compared four conditions: freehand ultrasound scanning, freehand scanning with visual contact force feedback, robot-assisted scanning without maximum force constraints, and robot-assisted scanning with maximum force constraints. The results showed that the robotic system effectively reduces human-applied force, while the maximum force constraint simplifies interaction, enhances contact force stability, and improves ultrasound image quality.

#### 8.1.1.2 *Sensory Substitution for Force Perception*

The design of sensory substitution for force perception involves several key factors, which are discussed in the following subsections. The first is the modality used to convey force information. For visual feedback specifically, two additional aspects are critical: the design of the feedback representation and its placement.

SENSORY SUBSTITUTION MODALITIES Several comparative studies on multimodal and unimodal approaches to assess the suitability of different modalities for sensory substitution conveying force information have been conducted. Jonetzko et al. [157] compared three methods of force visualization with auditory feedback for grasping tasks. While all feedback modalities reduced grasping force, no performance differences were found across conditions. Moreover, although users preferred multimodal, redundant feedback, it appeared to be less efficient. Kitagawa et al. [171] compared visual force feedback with discrete auditory feedback for a robotic suturing task. Their results indicated that visual outperformed auditory feedback in terms of force consistency. Haruna et al. [132] investigated visual, auditory, and vibrotactile feedback to signal excessive force during a grasping task. Their results revealed that visual feedback was the most effective in reducing grasping force without increasing cognitive load. Chan et al. [60] explored force augmentation for robot trajectory planning using visual and vibrotactile feedback. Their findings suggested that unimodal vibrotactile sensory substitution resulted in smaller force errors than visual or multimodal feedback. In contrast, Howard and Szewczyk [145] observed that visual feedback had

overall better performance when compared to different vibrotactile feedback types for force representation in laparoscopy. Additionally, they found that multimodal feedback enhanced task precision compared to visual and tactile feedback alone.

VISUALIZATION CONCEPTS Different designs of visual sensory substitution to facilitate force perception have been explored not only in HRI but also in domains such as minimally invasive surgery and remote collaboration. Visualization concepts include highlighted areas, bars, arrows, numerical displays, heat maps, and force gauges. These visualizations convey force information through color, size, numbers, and directional indicators. An overview of the different approaches found in related work is provided in Table 8.1.

Table 8.1: Overview of visual sensory substitution concepts for conveying force information in related work.

| **Type** | **Characteristic** | | **Application** |
|---|---|---|---|
| Highlighted Areas | Color | [14, 24, 271]<br>[39, 167]<br>[157] | Robotic Surgery<br>Remote Collaboration<br>Robot Teleoperation |
| | Size | [118] | Surgery |
| Bars | Size | [145]<br>[167]<br>[290, 367] | Laparoscopy<br>Remote Collaboration<br>Robot Teleoperation |
| | Size and color | [24, 171, 314]<br>[39]<br>[308]<br>[122, 224] | Robotic Surgery<br>Remote Collaboration<br>Robotic Ultrasound<br>Robot Teleoperation |
| | Size, color, and direction | [374] | Robot Teleoperation |
| Arrows | Size | [60, 157] | Robot Control |
| | Size and color | [82] | Robot Control |
| | Size and direction | [308]<br>[143] | Robotic Ultrasound<br>Laparoscopy |
| Numerical Displays | Number | [82, 157] | Robot Control |
| Heat Maps | Color | [24]<br>[308] | Robotic Surgery<br>Robotic Ultrasound |
| Force Gauges | Pointer | [117] | Robotic Ultrasound |

To compare different visual feedback designs, Black et al. [39] compared two force visualizations in a remote ultrasound collaboration task: continuous color changes of a virtual probe and an error bar that adjusted its size and color based on force offset. Their study found the error bar to be superior, with all participants preferring it due to its clear, intuitive representation. Similarly, Kim et al. [167] compared two approaches for sharing hand force in remote collaboration and found that a size-changing bar was more effective than a superimposed colored mesh on the remote collaborator's hands. Aviles-Rivero et al. [24] compared a colored circle, heatmap, bar, and traffic light

for robotic surgery applications, and derived several design implications: avoid overlapping the visualization with the region of interest or cluttering the display with excessive text; place small visualizations close to the region of interest; use intuitive color coding and simple geometric shapes; and provide multimodal cues, such as combining position and color, to enhance information conveyance. Recently, Song et al. [308] presented an experiment comparing four XR visual feedback designs for robotic ultrasound: a vector arrow, a squeezable arrow, a linear gauge, and a force heatmap. They found that the linear gauge, which represented contact force through both height and color change, was the most effective and user-friendly.

VISUALIZATION PLACEMENT The placement of XR content to enhance situational awareness has been shown to impact reaction time and comprehension [352]. According to Billinghurst et al. [36], information can be displayed in a world-stabilized, head-stabilized, or body-stabilized manner. In the context of HRI, this can be extended to include robot-stabilized information display [311].

*World-Stabilized Placement.* Given that most existing literature on force visualizations focuses on robotic surgery [238], the region of interest is typically visible through the endoscopic image displayed on a screen. Thus, a common approach is world-stabilized placement, presenting the force feedback, either adjacent to the image [24, 143, 171, 290, 314] or directly overlaid onto the image at the position of the instruments [14, 24, 118, 122, 145, 271]. Similarly, Yusof et al. [367] and Moortgat-Pick et al. [224] investigated teleoperation for grasping tasks, displaying the remote area on a monitor and integrating force feedback alongside the video stream. Gilbertson and Anthony [117] studied the impact of force visualization in freehand ultrasound examinations, displaying a force gauge beside the ultrasound image on an external monitor.

*Robot-Stabilized Placement.* In contrast to these screen-stabilized approaches, Jonetzko et al. [157] and Chan and Quintero et al. [60] investigated robot-stabilized placement of visual sensory substitution. They utilized the *Microsoft HoloLens* to display force visualizations directly at the position of the robotic end effector. De Franco et al. [82] compared screen- and robot-stabilized visual force feedback for a polishing task. While the former displayed real-time force plots on a desktop monitor, the latter utilized a holographic arrow rendered via the *HoloLens*. Results indicated that the robot-stabilized interface allowed for more consistent force maintenance, while participants encountered difficulties focusing on the task with the desktop feedback.

*Head- and Body-Stabilized Placement.* While both screen- and robot-stabilized visualization placements anchor the information at world positions, either in the direct or remote environment, Plabst et al. [250] expanded their study beyond world-stabilized positions, investigating the impact of head-, body-, and world-stabilized placements for notification visualization. They found wrist-mounted (body-stabilized) notifications promising for high-interaction scenarios, while world-stabilized notifications were better suited for stationary

tasks. Fixed notifications in the user's FOV were preferred for tasks involving frequent attention shifts. Rzayev et al. [277] found that while head-stabilized information reduced response times, it increased distraction compared to world- and body-stabilized displays. Similarly, Li et al. [193] showed that object-stabilization improved user experience by minimizing visual interference. Chua et al. [65] identified the center-right FOV position as optimal for balancing performance and usability in head-stabilized interfaces. Lee and Woo [181, 182] also found that displaying notifications in the middle-right FOV reduced response time and task load. Nunez Sardinha et al. [231] investigated the placement of User Interfaces (UIs) for robot control. They found that robot-stabilized and environment-stabilized placements resulted in lower workload and higher performance compared to interfaces placed within a head-stabilized placement.

### 8.1.2 *Research Gap*

Previous work has thus demonstrated the general feasibility of both partial automation and sensory substitution to assist users in HRI for contact-intensive tasks. However, there remain several open questions:

Regarding force automation, prior studies have shown its potential for robot control using haptic input devices [267, 370], joysticks [297], and hand-guiding [100]. However, the application of force automation in hand gesture interaction has not yet been explored. This is particularly relevant because hand gesture interaction results in the loss of haptic perception, which has been shown to introduce challenges in robot control [9].

In the context of sensory substitution, previous studies have provided comparative evaluations of different techniques, indicating that visual and vibrotactile modalities are suitable for force display [60, 132, 145]. Furthermore, other works suggest that the efficiency of force visualization may depend on both the design [24] and the spatial placement of the visualization [82, 250]. However, these studies do not consider how variations in design and positioning influence the need for multimodal sensory substitution. Since these factors may interact and affect the effectiveness of different modalities, a gap remains in the multifactorial evaluation of sensory substitution approaches.

## 8.2 INVESTIGATION ON PARTIAL AUTOMATION

The first approach to force assistance investigated in this thesis is partial automation. In this approach, the contact force is regulated by using sensors to detect external forces and adjusting the end effector's position accordingly. Since only the autonomous control of a single DOF is required to achieve the desired contact force, the remaining DOFs can be manually controlled.

As previously described, this approach has not yet been investigated for hand gesture interaction, despite the reported challenges in force perception inherent to this control method [9]. Therefore, this section aims to examine the potential of partial automation for HRI using hand gestures. To contextualize the results, hand-guiding is included as a baseline, enabling a direct comparison with a state-of-the-art interaction technique. To analyze both interaction methods and the impact of partial automation on them, two types of HRI tasks with different precision demands are investigated.

Parts of this section are published in: Tonia Mielke, Marilena Georgiades, Oliver S. Großer, Maciej Pech, Christian Hansen, and Florian Heinrich. "Enhancing Gesture-Based Human-Robot Interaction: Investigating the Role of Force Automation." In: *Proceedings of the 21st ACM IEEE International Conference on Human-Robot Interaction (HRI '26)*. Edinburgh, Scotland, UK: Association for Computing Machinery, 2026, pp. 1–10. DOI: 10.1145/3757279.3785568 [3].

### 8.2.1 *Technical Methods*

*Supplementary video*

The benefits of force automation have been demonstrated for different interaction methods in robotic ultrasound [100, 267, 370]. Given that hand gesture control has also been identified as a promising approach in this domain [9], robotic ultrasound is selected as the exemplary use case (see Section 1.2.4). This section describes the technical methods used to investigate whether the reported positive effects of partial automation in HRI also apply to hand gesture interaction.

#### 8.2.1.1 *Interaction Techniques*

To contextualize the performance of hand gesture-based robot control, hand-guiding is employed as a baseline. Hand-guiding is selected as it is commonly integrated into commercial robot control systems [327], and it is the predominant interaction method in robotic ultrasound [64, 196, 297, 368]. To enable the separation of DOFs during robot control, a foot pedal with two switches is used, allowing the user to toggle between translational and rotational DOFs.

HAND GESTURE INTERACTION The Hand Gesture Interaction (HI) technique is implemented as described in Section 2.3.2. An earlier version of this technique was presented in the co-authored publication by Schreiter et al. [9]. That version used position control to map hand movement to end effector translation and rate control to map forearm rotation to end effector

rotation. Since this method did not allow simultaneous manipulation of all DOFs, this thesis incorporates an improved version of the hand gesture-based robot control, utilizing position control for both hand position and rotation to directly map hand movement onto robot motion. As additional visual feedback, a virtual dummy probe is employed. This holographic visualization of a probe is placed in the operator's hand during interaction (see Figure 8.1a). Depending on the active mode, the virtual probe either translates or rotates in response to the user's hand movements. To achieve movement to commanded positions without reacting to external forces, impedance control with high stiffness (1000 N/m, 200 N/rad) is used. Hand motion is tracked using an HMD, which is registered with the robot as described in Section 2.3.4.

HAND-GUIDING In Hand-Guiding (HG), the compliant robotic arm is moved under gravity compensation. While prior research suggests the potential of using more sensitive sensors [66], hand-guiding using the robot's built-in force/torque sensors remains the standard interaction technique, as it is already integrated into commercial robot control systems [327]. Therefore, the robot's torque sensors are used to implement impedance control with minimal stiffness (1 N/m, 1 N/rad) and high nullspace stiffness (100 N/rad), ensuring that the robotic end effector follows external forces.

(a) Hand gesture interaction (b) Hand-guiding

Figure 8.1: Interaction techniques investigated. Reprinted from [3].

#### 8.2.1.2 *Force Control*

The objective of the Force Automation (FA) is to autonomously maintain a predefined contact force $F_d$. This is achieved through hybrid position/force control, which decouples the positioning and force laws. Force control is realized using impedance control to establish a mass-damper-spring relationship that balances the Cartesian end effector position and external force $F_{ext}$, measured through the robot's joint torque sensors. Consequently, the vertical position is autonomously adjusted to reach $F_d$, while the remaining degrees of freedom are controlled using the aforementioned interaction methods (see Figure 8.2). The force values are defined in alignment with Dhyani et al. [85] for abdominal examinations, setting $F_d = 10\,\text{N}$ and $F_{max} = 30\,\text{N}$.

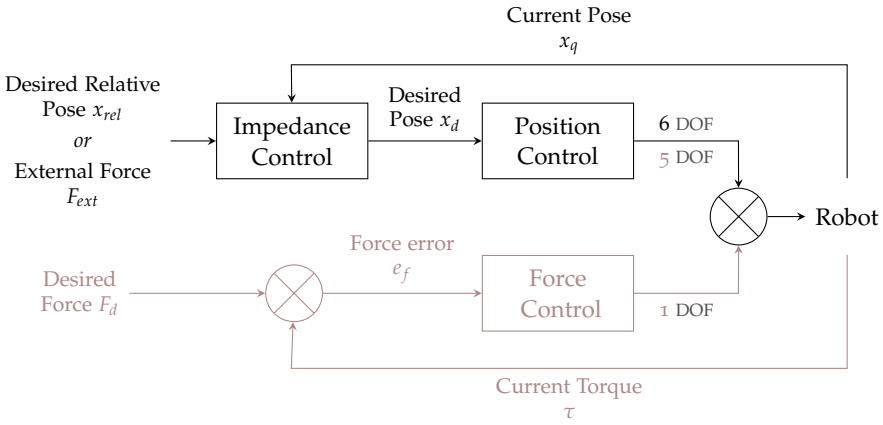



Figure 8.2: Control architecture of HRI for robot control ■ without and ■ with force automation. Adapted from [3].

#### 8.2.1.3 *Implementation*

The prototype was implemented as described in Section 2.3, with an ultrasound probe attached to the robot using a 3D-printed mount. Hand tracking was facilitated by the sensors of the *HoloLens 2*.

### 8.2.2 *Evaluation Methods*

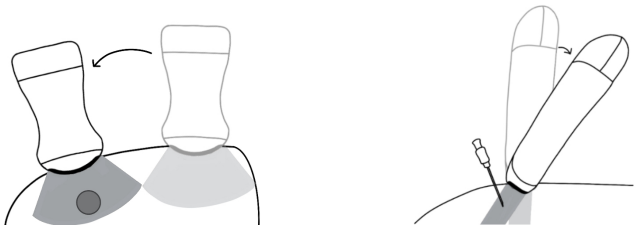
A user study was conducted to evaluate the impact of force automation on hand gesture HRI and compare it to the baseline of hand-guiding. This study employed a within-subjects design, described in the following section.

#### 8.2.2.1 *Tasks*

Related work has shown that the effectiveness of interaction techniques in HRI is task-dependent [282, 373]. To gain deeper insights into hand gestures for HRI and the effect of partial automation, this study investigates two different tasks that vary in the precision they require. In the context of robotic ultrasound, broad exploratory movements are used to initiate autonomous procedures, while smaller, precise motions are required to visualize target structures or correct autonomous behavior. One clinical scenario that combines both types of movements is needle-based interventions, including procedures such as central venous access, biopsies, and more complex interventions like brachytherapy and thermal ablation [257]. In these cases, broad exploratory movements help identify an appropriate acoustic window, while fine, precise control is necessary to accurately visualize the needle tip. Therefore, needle-based interventions are selected as an exemplary use case to investigate the suitability of different interaction techniques, as well as the role of force assistance in enhancing their performance. The tasks were identified and defined in collaboration with radiologists.

ACOUSTIC WINDOW TASK The initial task in the intervention involves identifying an *acoustic window* that facilitates imaging of the target region. This step is crucial for either initiating autonomous procedures [64, 196, 297, 368] or during teleoperation. In this study, the Acoustic Window Task (AW) involves maneuvering the transducer from a standardized starting position across the surface of a phantom, with the goal of visualizing a target structure embedded within it (see Figure 8.3a).

NEEDLE TIP TASK After initial probe positioning, the imaging focuses on visualizing the needle tip. While this could involve autonomous tracking of the needle, operators may need to correct offsets or adjust the visualization post-needle advancement. Therefore, the second task of this study is the Needle Tip Task (NT). This involves visualizing the needle tip of pre-positioned needles, with the transducer automatically positioned to require only minor adjustments for needle tip visualization (see Figure 8.3b). Although the needles are static in the experimental setup, during an actual intervention, this task would coincide with needle advancement. Therefore, participants are instructed to hold the needle in their dominant hand and control the robot using only their non-dominant hand.



(a) Acoustic Window Task (b) Needle Tip Task

Figure 8.3: Tasks investigated during user study. Reprinted from [3].

#### 8.2.2.2 *Variables*

INDEPENDENT VARIABLES The independent variables of the two-factorial study design were the interaction *technique* (HI, HG) and the force *automation* (presence or absence of FA).

DEPENDENT VARIABLES As dependent variables TCT and perceived workload were measured. TCT was calculated from the moment the robot reached its starting position until participants indicated satisfaction with their results. Task completion was then verified by the study conductor through visual inspection, ensuring adequate visibility of the target structure. Adequate visibility was defined as the visibility of approximately 75% of the sphere's outlines for AW or clear visibility of the needle tip for NT by confirming it was the last point along the needle. This method, performed by a trained researcher, was chosen for its practicality, efficiency, and ability to minimize task setup complexity. Perceived workload ratings were obtained using the NASA TLX questionnaire [131].

#### 8.2.2.3 *Hypotheses*

No prior assumptions regarding the outcomes were made. Thus, two-sided hypotheses were examined. Based on the variables described above, the following null and alternative hypotheses were investigated.

$H_0^y.x$ The mean $x$ is equal for all levels of $y$.
$H_1^y.x$ The mean $x$ differs for at least one level of $y$.

$$x \in \begin{cases} \text{1-TCT} \\ \text{2-TLX} \end{cases} \qquad y \in \begin{cases} \text{1-Technique} \\ \text{2-Automation} \end{cases}$$

The resulting null and alternative hypotheses are each investigated across the two different experimental tasks (NT, AW).

#### 8.2.2.4 *Sample Design*

As the task involved interacting with the robot to conduct ultrasound imaging, basic knowledge of ultrasound imaging was required. Therefore, medical students from the local university were recruited to participate in the user study through online polls. They were compensated with 30 € for their participation.

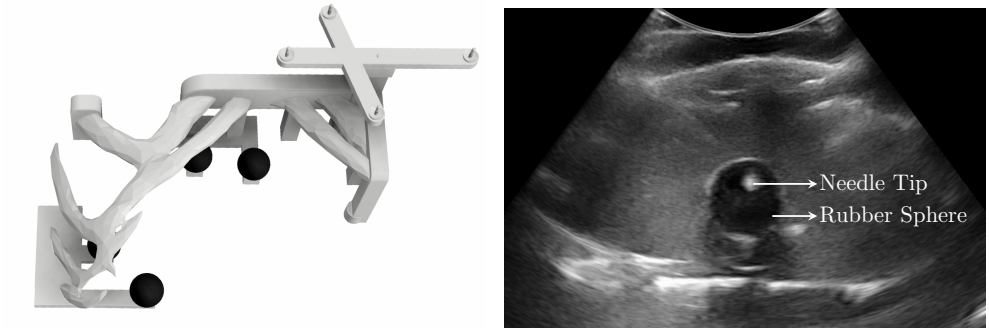
#### 8.2.2.5 *Apparatus*

The study utilized a custom-made phantom, developed to emulate realistic ultrasound properties and simulate acoustic shadowing caused by bones. As needle guidance was chosen as use case, the phantom, consisting of agar-agar, included a 3D-printed partial rib cage employing the CT-derived segmented model by De Jong et al. [158] as well as rubber spheres representing target regions (see Figure 8.4a). To account for the challenge of localizing target regions with ultrasound, four rubber spheres were strategically placed at varying accessibility levels (see Figure 8.4b).

At the beginning of the tasks, the robot was automatically positioned at a start position. To ensure consistent starting positions of the ultrasound probe on the phantom, a marker shield was affixed to the partial rib cage, facilitating the localization of the phantom. A second phantom, mirroring the layout of the first, contained one pre-positioned needle within each rubber sphere. An 18-gauge echogenic needle, commonly used for liver and kidney biopsies, was used. To support orientation during the study and provide pre-interventional insight into the phantom's internal structure, CT images were acquired beforehand. These were made available to participants via a standard DICOM viewer on an external screen.

#### 8.2.2.6 *Procedure*

After collecting informed consent and demographic data, participants underwent a training session to freely explore the phantom using manual ultrasound, familiarizing themselves with the structures. Subsequently, for




(a) CAD Model of 3D printed ribcage with rubber spheres and marker shield

(b) Ultrasound image of phantom showing rubber sphere and needle tip

Figure 8.4: Experimental apparatus for the user study. Reprinted from [3].

all four combinations of control methods (HG, HI, both with and without FA), the two tasks (AW and NT) were conducted. The order of tasks and control methods was counterbalanced across participants. Before each task, participants were able to train under the current experimental condition until they felt confident enough to proceed. They then performed the task for each of the four rubber spheres or needles, respectively. After each task, participants completed the NASA TLX questionnaire. Following all trials, a semi-structured interview was conducted to obtain more in-depth feedback. The study duration was approximately 60 minutes per participant.

#### 8.2.2.7 *Statistical Analysis*

Since the task involved localizing four target structures, TCT values were averaged across trials with identical experimental conditions for each participant. For statistical analysis, the data were first checked for homogeneity with Levene's tests and for normality using Shapiro-Wilk tests. For each dependent measure, at least one experimental condition violated these assumptions. Thus, robust two-way ANOVAs for within-subjects designs based on trimmed means were calculated to evaluate main and interaction effects (also see [348]). The $\delta_t$ estimate, as proposed by Algina et al. [16], was calculated as effect size for the main effects. To investigate identified interaction effects, robust Yuen's trimmed means tests were conducted, with p-value adjustments using Hochberg's method, for post hoc pairwise comparisons [348]. All statistical analyses were conducted using R (version 4.1.1).

### 8.2.3 *Results*

This section presents and interprets both the quantitative and qualitative results of the conducted user study.

#### 8.2.3.1 *Participants*

Twenty-eight (28) medical students (19 female, 9 male) participated in the study. Their ages ranged from 22 to 37 years (M=27.25, SD=3.74), and they were between their 3rd and 12th semesters of university studies (M=8.70, SD=1.71). Participants self-reported their familiarity with ultrasound (M=3.18, SD=0.72) and HRI (M=1.79, SD=0.83) on a 5-point Likert item, where 1 indicated no experience and 5 a high level of familiarity. They also reported their technical affinity (M=3.46, SD=0.88) on the same scale (1=low, 5=high).

#### 8.2.3.2 *Quantitative Results*

The descriptive results are summarized in Table 8.3, and the statistical effects are reported in Table 8.2. Although no interaction effects were detected for the TCT in the NT task, the descriptive data revealed notable trends, prompting further pairwise comparisons. The results are visualized in Figure 8.5, Figure 8.6, Figure 8.7, and Figure 8.8, with the post hoc pairwise comparison results shown in the plots.

Table 8.2: Summary of robust ANOVA results for TCT and TLX ($\alpha < .05$) in AW and NT tasks across *Techniques* (hand-guiding and hand gesture interaction) and *Automation* (with or without force automation). Test statistic $Q$ and effect size $\delta_t$ are reported. Adapted from [3].

| Effect type | Factor | Q | p | Sig. | $\delta_t$ | Effect | Fig. |
|---|---|---|---|---|---|---|---|
| **TCT - AW** | | | | | | | |
| Main | *Technique* | 16.682 | <0.001 | * | -0.682 | Medium | 8.5a |
| | *Automation* | 7.773 | 0.005 | * | -0.436 | Small | 8.5b |
| Interaction | *Technique* × *Automation* | 19.804 | <0.001 | * | - | - | 8.5c |
| **TLX - AW** | | | | | | | |
| Main | *Technique* | 30.176 | <0.001 | * | -0.844 | Large | 8.6a |
| | *Automation* | 41.059 | <0.001 | * | -0.661 | Medium | 8.6b |
| Interaction | *Technique* × *Automation* | 13.241 | <0.001 | * | - | - | 8.6c |
| **TCT - NT** | | | | | | | |
| Main | *Technique* | 2.722 | 0.099 | | 0.240 | - | 8.7a |
| | *Automation* | 8.017 | 0.005 | * | -0.607 | Medium | 8.7b |
| Interaction | *Technique* × *Automation* | 3.298 | 0.070 | | - | - | 8.7c |
| **TLX - NT** | | | | | | | |
| Main | *Technique* | 3.155 | 0.076 | | -0.156 | - | 8.8a |
| | *Automation* | 16.71 | <0.001 | * | -0.421 | Small | 8.8b |
| Interaction | *Technique* × *Automation* | 7.48 | 0.006 | * | - | - | 8.8c |

Table 8.3: Summary of descriptive results for all dependent variables ($n = 28$) for AWs and NTs tasks. All entries are in the format: mean value [standard deviation]. In the 'accumulated' row, values for HI and HG are combined to evaluate the main effects of force automation. Adapted from [3].

| Variable | TCT - AW [s] | TLX - AW | TCT - NT [s] | TLX - NT |
|---|---|---|---|---|
| **Accumulated** | 26.90 [9.36] | 30.30 [18.37] | 20.33 [7.21] | 30.77 [17.61] |
| Without Automation | 29.06 [10.48] | 36.06 [19.63] | 22.01 [7.46] | 33.91 [17.83] |
| With Automation | 24.74 [7.59] | 24.54 [15.10] | 18.65 [6.60] | 27.62 [16.95] |
| **Gesture Interaction** | 29.62 [9.65] | 36.41 [19.01] | 19.27 [6.36] | 32.71 [19.51] |
| Without Automation | 34.30 [8.36] | 46.55 [15.22] | 22.29 [6.89] | 39.67 [20.42] |
| With Automation | 24.94 [8.63] | 26.28 [17.08] | 16.25 [4.02] | 25.74 [16.07] |
| **Hand-Guiding** | 24.18 [8.28] | 24.18 [15.59] | 21.38 [7.88] | 28.82 [15.40] |
| Without Automation | 23.82 [9.83] | 25.57 [18.03] | 21.73 [8.11] | 28.15 [12.74] |
| With Automation | 24.53 [6.53] | 22.80 [12.90] | 21.04 [7.79] | 29.49 [17.88] |

### 8.2.3.3 *Qualitative Results*

During the semi-structured interviews, 295 individual statements were collected. To analyze the data, duplicate statements from the same participants were removed. Statements made by more than two participants were clustered and summarized, resulting in 203 statements being condensed into 40 key points (see Table 8.4).

Table 8.4: Summary and frequency of statements (#) received during the interview. Adapted from [3].

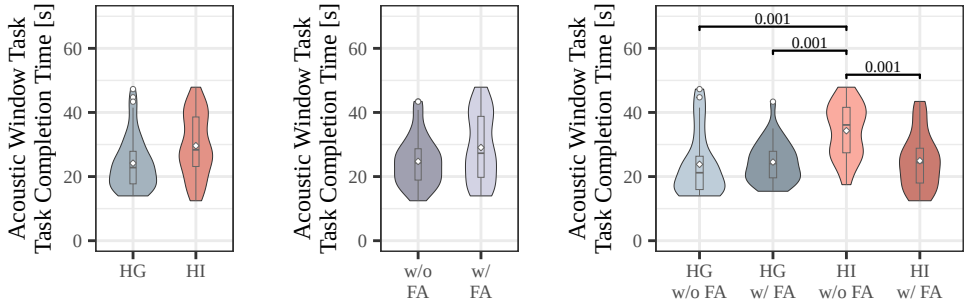
| | | Hand-Guiding | Hand Gesture Interaction |
|---|---|---|---|
| **Advantages** | + | Haptic feedback helpful (5) | Sensitive (10), rotation works well (6) |
| | + | Sensation of holding something (5) | Good responsiveness (5) |
| | + | Feels controlled (4), easy (3), familiar (3), precise (3) | Virtual probe helpful (4) |
| | + | | Quick to learn (4), precise (4), not exhausting (3), intuitive (2) |
| **Disadvantages** | - | Physically exhausting (8) | Assessing contact force difficult (10) |
| | - | Tilt (8) and rotation (6) challenging | Easy to make unwanted movements (7) |
| | - | Occasionally jerky (4) | High cognitive load (6), unfamiliar (5) |
| | - | Interaction with non-dominant hand challenging (4) | Need to look at robot (4) |
| | - | Uncertainty where to grab robot (4) | Moving hand in plane difficult (3) |
| | • | Adding handle (4) | Hand in FOV distracting (3) |
| | • | More sensitive sensors (3) | Hard (3), Lacking resistance (2) |
| | - | | Coarse movements difficult (2) |
| **FA** | ↓ | Force automation makes no difference (15) | |
| | ↓ | Force automation helpful (6) | Force automation helpful (20) |
| | ↓ | Force automation disturbing (4) | |

| General |
|---|
| - Coordinating pressing the pedals for control mode activation difficult (4) |
| - Pressing pedal causes involuntary arm movement (3) |
| ↓ The option to adjust $F_d$ for FA would be necessary in real-world scenarios (3) |
| ↓ Force automation improves imaging quality (2) |

#### 8.2.3.4 *Interpretation of Results*

This section interprets the quantitative and qualitative results obtained from the acoustic window and the needle tip tasks.

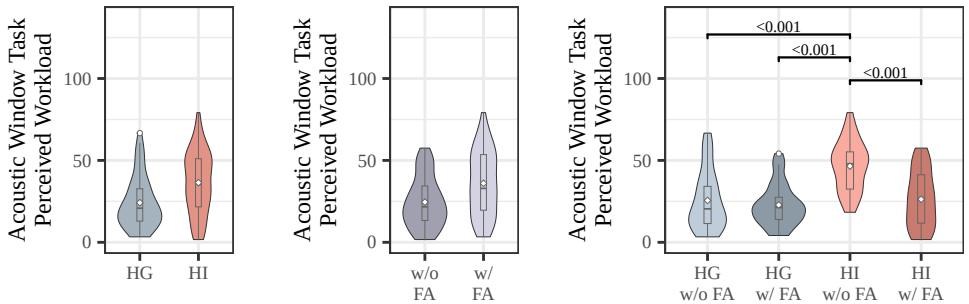
ACOUSTIC WINDOW TASK Significant main and interaction effects of the interaction technique and force automation on TCT were found (see Figure 8.5). Accordingly, $H_0^1.1$ and $H_0^1.2$ were both rejected in favor of their respective alternative hypotheses, $H_1^1.1$ and $H_1^1.2$. Further analysis revealed that these main effects were primarily driven by significant interaction effects, with post hoc pairwise comparisons revealing that the TCT was significantly higher for hand gesture interaction without force automation compared to the other methods (see Figure 8.5).




(a) *Technique* main effect (b) *Automation* main effect (c) *Technique*×*Automation* interaction

Figure 8.5: TCT results of the acoustic window task. Means are indicated by ◇. Statistically significant pairwise comparison results are indicated by brackets. Adapted from [3].

A similar pattern emerged for the NASA TLX results, where significant main and interaction effects were also observed. Post hoc comparisons indicated that hand gesture interaction without force automation resulted in a significantly higher perceived workload than any of the three other methods (see Figure 8.6). Therefore, $H_0^2.1$ and $H_0^2.2$ were both rejected in favor of their respective alternative hypotheses, $H_1^2.1$ and $H_1^2.2$.




(a) *Technique* main effect (b) *Automation* main effect (c) *Technique*×*Automation* interaction

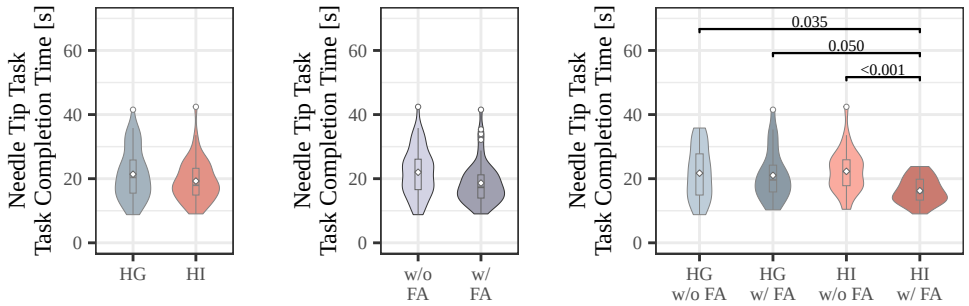
Figure 8.6: TLX results of the acoustic window task. Means are indicated by ◇. Statistically significant pairwise comparison results are indicated by brackets. Adapted from [3].

These findings, showing that hand gesture interaction without force automation performed significantly worse than all other methods, align with the qualitative feedback. Participants identified the primary challenge as the lack of haptic feedback, which made it difficult to accurately estimate and apply the optimal force required for good image quality without exceeding the force threshold. However, force automation appeared to mitigate this issue, significantly reducing both perceived workload and task duration.

In contrast, no significant difference was found between the force automation conditions during hand-guiding. This, too, aligns with qualitative feedback: most participants did not notice a difference, some found the automation helpful, while others considered it disruptive. This may be due to the haptic perception already present in hand-guiding, which reportedly made the interaction feel familiar and provided a greater sense of control.

As no significant differences in TCT and TLX were found between hand gesture interaction and hand-guiding with force automation, further post hoc analysis using Bayes factor analysis was conducted to examine the support for the null hypothesis. Bayes factors were calculated using Bayesian t-tests [142]. Moderate evidence supporting the null hypothesis was found (TCT: $BF_{10} = 0.20$, TLX: $BF_{10} = 0.27$). These results suggest that, despite participants expressing a preference for direct haptic feedback during hand-guiding, the reportedly intuitive, fast-reacting, and understandable interaction offered by hand gesture interaction with force automation produced comparable outcomes in terms of perceived workload and TCT.

NEEDLE TIP TASK In the needle tip visualization task, significant main effects of force automation on the TCT were observed (see Figure 8.7). Therefore, $H_0^1.1$ and $H_0^1.2$ were both rejected in favor of their respective alternative hypotheses, $H_1^1.1$ and $H_1^1.2$. While no interaction effects were identified, pairwise comparisons revealed that hand gesture interaction with force automation exhibited a significantly lower TCT compared to the other methods.



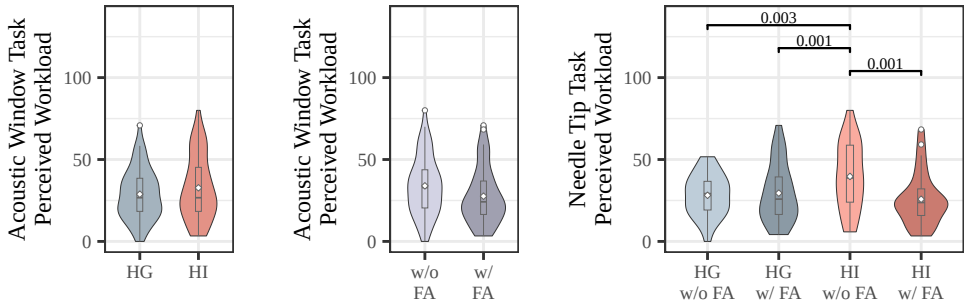

(a) *Technique* main effect (b) *Automation* main effect (c) *Technique*×*Automation* interaction

Figure 8.7: TCT results of the needle tip task. Means are indicated by ◇. Statistically significant pairwise comparison results are indicated by brackets. Adapted from [3].

This is also reflected in the qualitative feedback, where hand gesture interaction was described as sensitive and precise, while hand-guiding was perceived as physically exhausting and jerky. Moreover, participants de-

scribed difficulties in rotating the transducer during hand-guiding due to high resistance when grasping the robot by the probe. Attempting to alleviate this issue by grasping farther up the robotic arm resulted in a longer lever, which eased interaction but also led to less naturalness. This challenge posed by probe rotation, particularly affecting TCT for needle tip visualization, can be attributed to the frequent use of fanning motions for tracking the needle tip, in contrast to the predominantly translational movements employed during the acoustic window task.

The perceived workload results for the needle tip task showed significant main and interaction effects for both interaction technique and force automation (see Figure 8.8). Therefore, $H_0^2.1$ and $H_0^2.2$ were both rejected in favor of their respective alternative hypotheses, $H_1^2.1$ and $H_1^2.2$. Similar to the results from the acoustic window task, pairwise comparisons revealed that perceived workload was significantly higher for hand gesture interaction without force automation compared to all other methods. This may again be attributed to the lack of haptic feedback, which likely required a higher mental workload to control the robot effectively.




(a) *Technique* main effect (b) *Automation* main effect (c) *Technique*×*Automation* interaction

Figure 8.8: TLX results of the needle tip task. Means are indicated by ◇. Statistically significant pairwise comparison results are indicated by brackets. Adapted from [3].

### 8.2.4 *Discussion*

This work investigated the potential of force automation to mitigate the missing haptic feedback during HRI using hand gestures. This section discusses findings, describes limitations, and presents resulting design implications.

FORCE AUTOMATION Regarding force automation, the results indicate that for hand gesture interaction, automation significantly improved workload and interaction duration. This demonstrates that force automation helps users compensate for the lack of haptic feedback and enables more efficient robot control. This aligns with previous findings showing that contact force automation eases interaction [267, 370]. However, for hand-guiding, force automation did not yield significant improvements, suggesting that the effectiveness of force automation depends on the interaction method, as force perception during hand-guiding may reduce the need for assistance.

It should be noted that the force automation was constrained by a specific force level. As patient-specific adjustments to the constant force or the temporary application of a higher force may be necessary to ensure imaging quality, future work should include possible manual adjustments or more sophisticated control schemes.

INTERACTION TECHNIQUES The results additionally indicate that the suitability of interaction methods is task-dependent. For broad acoustic window tasks, hand gesture interaction with force automation performed comparably to hand-guiding, while it outperformed hand-guiding for precise tasks. This context-dependency aligns with prior research [282, 373], highlighting the need for task-specific evaluation of interaction methods. Regarding the implemented hand-guiding, participants reported jerky robot behavior and perceptible resistance. This may be attributed to gripping the robot near the end effector, resulting in a shorter lever, as well as the use of force/torque sensors, which are prone to noise at low force levels. Although more sensitive sensors, such as those used by Cirillo et al. [66], could address these issues, employing the integrated sensors reflects the manufacturer-provided interaction method and is thus considered a valid baseline.

In contrast to earlier work, which indicates worse performance for touchless hand gesture interaction than for touch-based methods like hand-guiding and haptic devices [293, 373], this study demonstrates that hand gesture interaction can match or even surpass these methods when the lack of haptic perception is compensated. Therefore, the findings highlight the potential of hand gesture interaction for HRI, motivating further investigation of this interaction modality throughout the remainder of this thesis.

#### 8.2.4.1 *Limitations*

While the results provide valuable insights into the task-specific suitability of different control methods for HRI, it is important to note some limitations. Firstly, the efficiency assessment was focused on TCT as an objective measure. Given the nature of the tasks, where the primary goal was to achieve sufficient visualization for successful navigation during intervention rather than precise alignment, it was chosen to measure only the duration of task completion rather than accuracy metrics, assuming that participants performed comparably across techniques, which was also validated by the experimenter. While it was aimed to capture nuances in efficiency demands by assessing two tasks with varying complexity, incorporating additional tasks as well as assessing accuracy measurements could enhance the understanding of interaction technique performance and speed-accuracy trade-off.

In this study, medical students with initial ultrasound experience were selected as subjects. While the training provided ensured that they had no difficulty with imaging, experienced medical professionals may be better at identifying factors such as contact force from the ultrasound image. However, it is believed that the results are primarily dependent on the interaction technique rather than medical expertise, as feedback indicated that robot

handling was the key factor in the differences between control approaches. While it is believed the results are therefore generalizable, future work should include medical professionals for more in-depth feedback.

The final limitation arises from the study's specificity. Robotic ultrasound was selected as an exemplary use case, with tasks chosen to reflect scenarios relevant to needle-based interventions. As a result, the findings are task-dependent and influenced by the particular characteristics of the selected procedure. However, considering the task characteristics, specifically the varying levels of required precision, it is believed that the general findings may be transferable to other use cases. In particular, the result that hand gesture interaction shows potential for precise tasks, and that force automation can compensate for missing haptic feedback, may be applicable in contexts beyond robotic ultrasound. Future research could explore these interaction methods further and examine their applicability in broader domains.

#### 8.2.4.2 *Implications*

The study's findings lead to the following implications:

**Automate contact force during hand gesture interaction.** Contact force automation significantly reduced TCT and perceived workload during hand gesture interaction, helping to overcome challenges related to the absence of haptic feedback.

**Employ hand-guiding or hand gesture interaction with force automation for broad tasks.** The results for the acoustic window task show that hand gesture interaction performs comparably to hand-guiding when contact force is automated. This suggests that both approaches are suitable for broad control, which is essential for exploratory movements or for initializing autonomous procedures that require coarse pre-positioning.

**Use hand gesture interaction for precise tasks.** In the needle tip task, hand gesture interaction with force automation outperformed hand-guiding in terms of task completion time. This demonstrates the potential of hand gesture interaction for high-precision tasks, such as fine-tuning autonomous motions or performing precise tasks like grinding or polishing.

### 8.2.5 *Conclusion*

In summary, this study investigated different control methods and the effects of autonomous adjustment of contact force for robotic ultrasound. The results emphasize the importance of force automation in hand gesture interaction, which results in reduced task duration and perceived workload while compensating for the lack of haptic feedback. In addition, assisted hand gesture control showed comparable performance to hand-guidance in coarse movements and outperformed it in precision tasks such as needle tip visualization. These findings highlight the potential of using hand gestures for robot control and the need for additional support to address the absence of haptic feedback.

## 8.3 INVESTIGATION ON SENSORY SUBSTITUTION

Besides contact force automation, a widely used approach for addressing the challenges caused by the absence of haptic feedback is sensory substitution. In this approach, haptic sensations are mapped onto alternative sensory modalities, such as auditory [157], vibrotactile [60], visual [224], or multimodal feedback [145]. Prior work has demonstrated distinct advantages for visual [132, 171], vibrotactile [60], and multimodal [145] feedback. Visual feedback, in particular, can be implemented using XR, which allows for flexible placement of feedback elements, with this spatial location influencing perceptibility [82]. While different sensory substitution strategies have been explored in previous studies (see Section 8.1.1.2), interactions between visualization placement, design, and multimodality have not been thoroughly investigated. Additionally, these approaches were not studied in the context of hand gesture-based HRI. Therefore, this section focuses on the targeted design of sensory substitution for force representation in HRI.

Parts of this chapter were previously published in: Tonia Mielke, Florian Heinrich, and Christian Hansen. "SensARy Substitution: Augmented Reality Techniques to Enhance Force Perception in Touchless Robot Control." In: *Transactions on Visualization and Computer Graphics (TVCG)* 31.5 (Mar. 2025), pp. 3235–3244. DOI: 10.1109/TVCG.2025.3549856 [6].

### 8.3.1 *Technical Methods*

To investigate different sensory substitution approaches, robotic ultrasound was again selected as a representative use case. In this context, it is essential not only to avoid excessive force to ensure patient safety but also to maintain a consistent force to achieve stable tissue deformation. Accordingly, the sensory substitution methods implemented in this study do not convey absolute force values but instead represent the applied force along the axial direction of the probe, expressed relative to a desired force $F_d$.

*Supplementary video*

Based on the previous work presented in Section 8.1.1.2, two common visual concepts were implemented, one 2D and one 3D, each encoding relative force deviations through changes in shape and color. In addition, a vibrotactile feedback concept was developed to communicate force information using variations in patterns and frequencies (see Figure 8.9). To explore different visualization placements, an HMD was used. The use of an HMD enabled flexible prototyping of feedback positions while maintaining consistent visual properties, thereby providing a controlled environment for evaluating the influence of visualization placement on user performance.

#### 8.3.1.1 *Visual Feedback Concepts*

Related work has highlighted the potential of redundant cues such as color and shape changes for force visualization [24, 39]. Therefore, this section investigates two common feedback designs for force display: a bar and an arrow, both conveying force information through changes in color and shape.

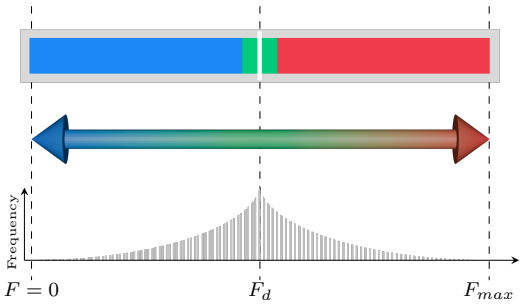



Figure 8.9: Overview of sensory substitution concepts. From top to bottom: bar, arrow, and vibrotactile feedback design. Reprinted from [6].

BAR The first visualization features stacked bars (see top visualization in Figure 8.9 and top row of Figure 8.10). The height of the 2D bar continuously indicates the currently applied force $F$, while discrete color coding indicates its relation to $F_d$. The color scheme is defined as follows:

$$color_{Bar} = \begin{cases} \text{blue} & F < F_d - 1\,\text{N} \\ \text{red} & F > F_d + 1\,\text{N} \\ \text{green} & \text{else} \end{cases}$$

Additionally, a white line marks the goal force $F_d$. The threshold of 1 N above or below $F_d$ was empirically tuned, aiming to provide users with precise feedback to distinguish small deviations in applied force. The visualization's orientation is continuously adjusted to be facing towards the observer.

ARROW Instead of mapping the visualization's size to the absolute force magnitude, a common alternative is to represent the relative force deviation from $F_d$. In this case, the visualization shrinks as the applied force approaches the target and disappears when the desired force is reached. This method is often implemented using an arrow, which can also convey force direction.

To this end, a 3D arrow was implemented that varies in height, color, and orientation (see bottom row in Figure 8.10). The arrow's orientation reflects the direction of the applied force. For forces smaller than $F_d$, the arrow points in the direction of the applied force, with the color interpolated between blue and green. For forces larger than $F_d$, the arrow points in the opposite direction, with the color interpolated between green and red. Thus, the pointing direction always indicates the movement direction required to reach $F_d$. Both the length and the color transitions are scaled linearly (see middle visualization in Figure 8.9).

#### 8.3.1.2 *Visual Feedback Placement*

To investigate the impact of visualization placement, three distinct positions were examined. As previously described, XR content in HRI can be classified as head-stabilized, body-stabilized, world-stabilized, or robot-stabilized [36, 311]. For visualizing force information, world-stabilized placement is commonly used, anchoring the visualization at task-relevant locations in the environment, such as a screen [24, 117, 367]. Similarly, in HRI, the robotic end effector

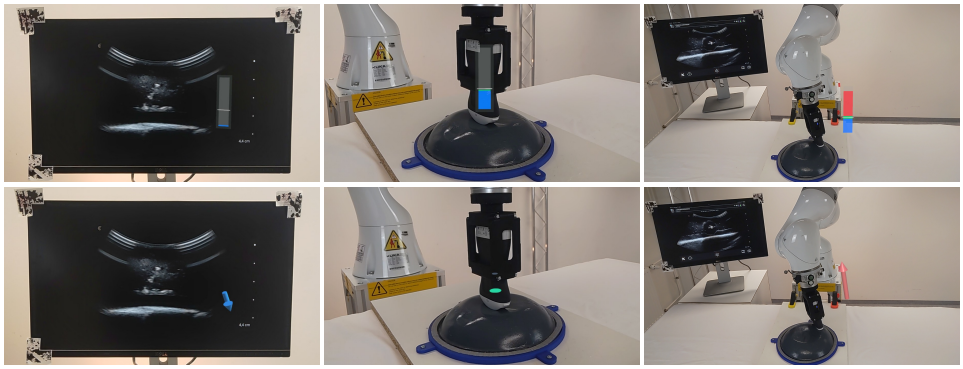


Figure 8.10: Overview of force visualization types and placements. The top row shows bar visualizations, and the bottom row shows arrow visualizations, both at different force levels (left to right: too low, appropriate, and too high). The left column shows placement on the screen displaying the sonographic image, the middle column shows placement on the ultrasound probe in contact with the phantom, and the right column shows placement anchored to the user's FOV. Reprinted from [6].

is often used to anchor information at task-relevant positions [60, 82, 157]. Therefore, world-stabilized and robot-stabilized placements were selected for investigation. Regarding body- and head-stabilized placements, prior work has shown that head-stabilized content supports attention shifts [250] and reduces response times [277]. Consequently, the third placement examined was a fixed position within the user's FOV. An overview of these placements is provided in Figure 8.11.

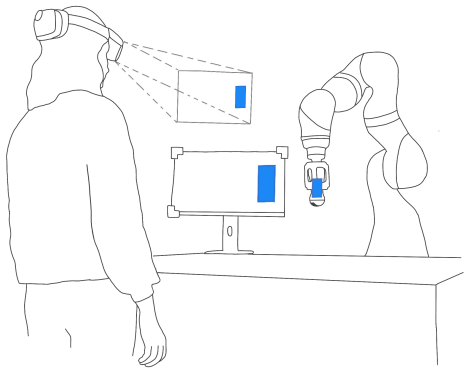


Figure 8.11: Overview of visualization placements. Reprinted from [6].

Since the investigated task is robotic ultrasound, in which the ultrasound image is displayed on a screen, this screen serves as the anchor for the world-stabilized placement (*screen* position). The screen's location is determined using image markers attached to its corners, which are detected once and then fixed in position. Based on the findings of Lee and Woo [181, 182] as well as Chua et al. [65], the virtual content is positioned in the center-right, adjacent to the ultrasound image.

For robot-stabilized placement, the feedback is located at the robotic end effector. Since the ultrasound probe is mounted to the end effector, this placement is referred to as the *probe* position and is chosen based on its frequent use in the literature. In this setup, virtual content is superimposed on the ultrasound probe, with its position continuously tracked via robot forward kinematics and transformed through an initial registration at the start of the interaction. As the intervention focuses on the general position of the probe and its contact point on the body, placing the visualization at the upper part of the probe is assumed not to obstruct the view.

The third position aims to explore head-stabilized information by anchoring the visualization within the user's FOV (*FOV* position). Analogous to the screen position, the visualization is placed in the center right. Specifically, the virtual content is placed 1 m from the user, vertically centered, and positioned within the right quarter of the FOV.

#### 8.3.1.3 *Vibrotactile Feedback Concept*

Based on the findings reported by Howard and Szewczyk [145], which suggest that pulsed vibrotactile feedback performs better than continuous feedback with varying amplitudes, this study adopts a pulsed approach. The vibrotactile feedback utilizes vibration pulses with varying pause lengths to convey deviation from $F_d$. Following the work of Chan et al. [60], different vibration patterns are used to indicate forces that are higher or lower than $F_d$. Specifically, forces smaller than $F_d$ are represented as a double pulse with a pause of 10 ms between the pulses, while forces higher than $F_d$ are represented as a single pulse. To enhance the detectability of differences in forces closer to $F_d$ and facilitate faster detection, a scaling approach similar to that used by Marquard et al. [207], as described by Equation 8.1, is employed. $T_p$ is scaled to be between the minimal pause length $T_{\text{min}} = 10\,\text{ms}$ at $F_d$ and the maximal pause length $T_{\text{max}} = 100\,\text{ms}$ at the maximal allowable force $F_{\text{max}}$ using the relative force $F_{\text{rel}}$.

$$\begin{aligned} F_{\text{rel}} &= \min\left\{1,\ \frac{F - F_d}{F_{\text{max}} - F_d}\right\} \\ T_p &= \sqrt{2F_{\text{rel}} - (F_{\text{rel}})^2} \cdot (T_{\text{max}} - T_{\text{min}}) + T_{\text{min}} \end{aligned} \tag{8.1}$$

This approach was selected because it adapts a widely known logic used in car parking systems: as the desired force is approached, the pause length $T_p$ decreases, creating a faster vibration pattern that becomes most rapid at $F_d$. Using the highest vibration pulse frequency near $F_d$, rather than increasing the frequency proportional to the force (as suggested by Howard and Szewczyk [145]), was chosen to provide the highest information density close to $F_d$, making deviations easily detectable. While this contrasts with the linear scaling used for the visual force representations, the information density is assumed to be comparable across modalities, since the visualizations present multiple channels of information through shape and color changes, making small changes around $F_d$ easily perceptible.

The vibrotactile feedback is delivered through a vibrotactile wristband worn on the dominant hand, with the patterns transmitted via an auditory signal. Therefore, Equation 8.1 is utilized to scale $T_p$ between simple inaudible sounds with a duration of 100 ms at a frequency of 100 Hz created using the Windows Beep function.

#### 8.3.1.4 *Force Data*

The force data were derived from the joint torque sensors of the robot manipulator. Analogous to Section 8.2, the force thresholds followed the values reported by Dhyani et al. [86] for abdominal examinations in patients with a high BMI, with $F_d = 10\,\mathrm{N}$ and $F_{\max} = 30\,\mathrm{N}$. Due to the relatively low forces involved and the minimal load introduced by the ultrasound probe, the Cartesian force estimation was susceptible to digital noise. To mitigate this, a Kalman filter [340] was applied.

#### 8.3.1.5 *Robot Control*

The robot was controlled as described in Section 2.3.2, using hand gestures to manipulate the translation and rotation of the robotic end effector. As additional visual feedback, a virtual dummy probe was employed, displayed as a hologram in the user's hand during interaction. The interaction mode (separate or simultaneous translation and rotation) was toggled using one foot pedal, while another pedal was used for clutching.

#### 8.3.1.6 *Implementation*

The setup consisted of the components described in Section 2.3. Hand tracking for robot control was facilitated by the sensors of the *HoloLens 2*. Vibrotactile feedback was provided by a *Lofelt Basslet* (Lofelt, Germany) bracelet, which is a haptic wristband featuring a high-bandwidth vibration actuator driven by auditory signals transmitted via Bluetooth.

The system consisted of two *Unity* applications. The first ran on the HoloLens and handled robot registration, determined the ultrasound screen position using the *Vuforia SDK*, visualized force feedback, and mapped hand gestures to robot movements. The second application ran on an external PC, managed communication between the HoloLens and the robot, filtered force data, and, for the vibrotactile condition, converted force values into inaudible auditory signals sent to the bracelet. For temporal alignment, both feedback modalities were updated at the same frame rate.

### 8.3.2 *Evaluation Methods*

To comparatively evaluate the sensory substitution techniques, a three-factorial within-subjects design user study was conducted. This section describes the study design, task, and experimental procedure.

#### 8.3.2.1 *Task*

To evaluate the different sensory substitution concepts, a task requiring motion while maintaining a specific contact force was necessary. Therefore, an exploratory task was chosen in which the ultrasound probe had to be navigated to identify geometric shapes within a phantom. To this end, a homemade agar-agar phantom was constructed, containing three distinct geometric shapes (a sphere, a pyramid, and a cube) made of rubber.

To require both translational and rotational movement, the phantom was designed in the shape of a semi-sphere. Its lateral curvature necessitated probe rotation to maintain image quality. Figure 8.10 shows the appearance of the phantom alongside corresponding sonographic images.

Participants were instructed to successively visualize the three target structures while maintaining a constant force as close as possible to $F_d$. To achieve this, the robot was first moved from its mid-air starting position to the surface. Participants then maneuvered the probe along the phantom's surface to locate an acoustic window where the outline of each structure was clearly visible before moving on to the next target.

To prevent minor initial forces during surface contact from influencing the results, participants were instructed to press the probe against the surface until $F_d$ was reached at the start of each task. Upon reaching $F_d$, a blue LED on the robot blinked, signaling the beginning of force recording.

#### 8.3.2.2 *Variables*

INDEPENDENT VARIABLES The independent variable was the force display *technique*, which was composed of the factors of visualization *type*, visualization *position*, and *vibrotactile* feedback. The different factors include:

- *Type*: Bar and Arrow force visualization concepts
- *Position*: Placement at screen, probe, and FOV
- *Vibrotactile*: Presence or absence of vibrotactile feedback

Additionally, the study investigated two baseline conditions for investigating the sensory substitution *modality*. In one condition, the visualization *type* and *position*, as well as the *vibrotactile* factors, were set to "none," representing no sensory substitution (*technique*: none). In the other condition, only *vibrotactile* feedback was active (*technique*: tactile). The combination of these factor levels resulted in 14 prototypes (2 *types* $\times$ 3 *positions* $\times$ 2 *vibrotactile* + 2 baseline).

DEPENDENT VARIABLES As dependent variables, both objective and subjective variables were measured. The first dependent variable was the TCT, which was measured from the moment participants first reached $F_d$ until the completion of the task. Task completion involved visualizing the three geometries, with the study conductor confirming adequate visibility through visual inspection. The other two objective variables concerned the applied force, which was continuously recorded and paused when no interaction was

active to prevent constant force exertion during static robot positions from influencing results. To maintain a consistent sampling rate across trials and participants, the force data were interpolated to an equidistant sampling at a sampling rate of 1 kHz. The two dependent variables were the Average Absolute Deviation (AAD), calculated as the average absolute deviation from $F_d$ at each time point, and Average Maximal Applied Force (AMax), computed by averaging the force of local maxima. Local maxima were filtered to include only those with a prominence of 1 N to mitigate the influence of jitter on results. Additionally, participants were asked to rate two questions on a 6-point Likert item (with verbal anchors ranging from "strongly disagree" to "strongly agree"). The questions regarded the *perceived difficulty* ("The currently applied force was difficult to estimate.") and the *perceived continuity* ("I was continuously aware of the magnitude of the applied force.").

#### 8.3.2.3 *Hypotheses*

No a priori assumptions regarding the outcomes were made, as this was an exploratory study. Thus, two-sided hypotheses were examined. Based on the three-factorial design and the baseline conditions described above, the following null and alternative hypotheses were considered:

$H_0^y.x$ The mean $x$ is equal for all levels of $y$.

$H_1^y.x$ The mean $x$ differs for at least one level of $y$.

$$x \in \begin{cases} \text{1-TCT} \\ \text{2-AAD} \\ \text{3-AMax} \\ \text{4-Difficulty} \\ \text{5-Continuity} \end{cases}$$

$$y \in \begin{cases} \text{1-Type} \\ \text{2-Position} \\ \text{3-Modality} \end{cases}$$

#### 8.3.2.4 *Sample Design*

As robotic ultrasound served as an exemplary use case, a basic understanding of sonographic images was necessary. However, due to the abstract nature of the task, no specific clinical expertise was required. Therefore, medical students from the local university were invited to participate via online polls and received 30 € as compensation.

#### 8.3.2.5 *Procedure*

The study began with participants completing a consent form and a demographic questionnaire upon arrival. They then received a brief introduction to the study objectives and the system. This was followed by a training session during which participants practiced freehand ultrasound imaging on the phantom, familiarized themselves with the positioning of the target structures, and practiced interacting with the robot. Once participants reported feeling comfortable with the imaging procedure and the HRI, they performed one trial for each sensory substitution technique, resulting in a total of 14 trials. To

reduce potential order effects, the sequence of *techniques* was counterbalanced across participants using a $14 \times 14$ Latin square. For each trial, participants had the opportunity to practice under the current conditions until they felt confident to proceed. They then performed the experimental trial, in which they guided the robot to each geometric structure once. After completing each trial, participants provided subjective ratings on perceived difficulty and continuity. Finally, after completing each trial, a semi-structured interview was conducted. The experiment lasted approximately 60 minutes on average.

#### 8.3.2.6 *Statistical Analysis*

To evaluate effects of the three investigated factors on the observed measures and their respective interaction, the data of the objective variables TCT, AAD, and AMax were first checked for normality using Shapiro-Wilk tests and for homogeneity using the Lavene's tests. Three-way repeated measures ANOVA was conducted in case this assumption was met. Otherwise, robust three-way ANOVAs for within-subjects designs based on trimmed means were calculated (also see [348]). For ordinal data of the Likert-item variables perceived difficulty and perceived continuity, Wilcoxon signed-rank tests for paired samples were conducted to evaluate main effects of factors with two factor levels, i.e., *type* and *vibrotactile*, and Friedman tests were calculated for main effects with more than two factor levels, i.e., *position*. The existence of potential interaction effects was evaluated using robust three-way ANOVAs. For standard ANOVAs, the test statistic $F$ and the effect size $\eta^2$ were reported. In the case of robust ANOVAs, the test statistic $Q$ was reported. Effect sizes for main effects in robust analyses were estimated using $\delta_t$, as proposed by Algina et al. [16], while interaction effects were reported using the more conventional $\eta^2$.

These investigations did not consider data from the two baseline conditions, *none* and *tactile*. To assess them, data from the remaining conditions were aggregated into two points per participant: *visual*, representing the mean for all visual-only conditions, and *multimodal*, representing the mean for all visual and vibrotactile conditions. The resulting data formed a one-factor repeated measures design with four levels. This *modality* factor was then evaluated as described earlier. Normality was checked, followed by repeated measures ANOVAs or robust ANOVAs if violated. Subjective items were analyzed with Friedman tests.

Furthermore, post hoc pairwise comparisons were conducted for all significant main effects involving more than two factor levels. Pairwise paired t-tests with Bonferroni corrections were conducted for normally distributed objective variables. Robust Yuen's trimmed means tests with p-value adjustments using Hochberg's method were used for non-normal objective data. Pairwise Durbin-Conover tests with Bonferroni corrections were calculated for the subjective measures. All statistical analyses were conducted using R (version 4.4.0).

### 8.3.3 *Results*

This section presents the results of the conducted user study and interprets the findings. The descriptive results are summarized in Table A.3.

#### 8.3.3.1 *Participants*

In total, twenty-one (21) medical students participated in the study (15 female, 6 male) with ages ranging from 22 to 32 years (M =26.14, SD= 3.23). They were in their 3rd to 5th year of university (M=4.41, SD=0.70). None of the participants reported any form of color vision deficiency. Participants reported their familiarity with ultrasound (M=3.48, SD=0.68) and HRI (M=2.00, SD=0.95) on a five-point Likert item, where 5 indicates high familiarity. Their self-reported technical affinity ranged from two to five (M=3.29, SD=0.72) on the same scale.

#### 8.3.3.2 *Quantitative Results*

The following outlines the results of the conducted statistical tests. Emphasis is placed on significant findings. More details on the results of the statistical test can be found in Table A.4. Additional plots illustrating descriptive results can be found in Figure A.1.

Regarding the tests evaluating the three-factorial design, TCT and AAD measures were not normally distributed. Hence, robust ANOVAs were conducted. The tests revealed a significant *vibrotactile* main effect on TCT (test statistic $Q = 8.869$, p-value $p = 0.003$, effect size $\delta_t = -0.28$) and a significant *type* main effect on AAD ($Q = 8.068$, $p = 0.005$, $\delta_t = -0.35$). All other main effects showed no statistical significance. Also, neither significant two-way nor significant three-way interaction effects were found.

The AMax variable was normally distributed. The repeated measures ANOVA revealed a significant three-way interaction effect between all three factors (test statistic $F = 3.902$, p-value $p = 0.028$, effect size $\eta^2 = 0.03$). No significant main and two-way interaction effects were found on this variable.

The subjective measures (perceived difficulty and continuity) were not affected by interaction effects. Friedman tests showed significant *position* main effects on subjective difficulty (test statistic $\chi^2 = 6.237$, p-value $p = 0.044$, effect size $W = 0.15$) and subjective continuity ($\chi^2 = 7.629$, $p = 0.022$, $W = 0.18$).

Regarding the analyses of the *modality* factor, the TCT data were not normally distributed. Hence, a robust repeated measures ANOVA was conducted. This test revealed a significant effect ($F = 4.804$, $p = 0.016$, $\delta_t = -0.09$). The respective AAD and AMax data were normally distributed. A repeated measures ANOVA showed a significant effect of the *modality* factor on AAD ($F = 7.362$, $p < 0.001$, $\eta^2 = 0.20$). Friedman tests were conducted on subjective difficulty and subjective continuity. These tests also revealed significant effects of the *modality* factor (perceived difficulty: $\chi^2 = 30.429$, $p < 0.001$, $W = 0.48$; perceived continuity: $\chi^2 = 40.215$, $p < 0.001$, $W = 0.64$).

#### 8.3.3.3 *Qualitative Results*

For the qualitative feedback analysis, the statements made by the participants during the semi-structured interview were paraphrased and clustered. Clustering focused on statements consistent between at least two participants, with duplicates within participants removed. In total, 251 individual statements were recorded based on participants' verbal feedback. Of these, 206 were expressed by at least two participants. These statements were condensed into 43 summarizing statements, presented in Table 8.5.

Table 8.5: Summary and frequency of statements (#) received during the semi-structured interview. Adapted from [6].

| **Type** | |
|---|---|
| *Bar* | *Arrow* |
| + easy to read (2) | + intuitive (11), continuous (6), precise (2) |
| + correct force precisely recognizable (2) | + visualization of direction helpful (9) |
| + intuitive proportions (2) | + peripheral changes perceptible (8) |
| - too sensitive due to small green area (4) | + continuous color scale striking (2) |
| - sensitivity causes overestimation (4) | - force direction confusing (3) |

| **Position** | | |
|---|---|---|
| *Screen* | *Probe* | *FOV* |
| + helpful when observing US image (17) | + helpful when looking at probe to estimate robot movement (6) | + helpful for viewing both screen and probe (3) |
| - outside of view when looking at probe (2) | - out of view when looking at screen (11) | - requires looking at additional position (6), is not in focus (6), confusing (3), disturbing (2) |

| **Vibrotactile** |
|---|
| + intuitive pattern (7), helpful(6), correct force easily recognizable (2), pleasant (2) |
| + draws attention to changes (6), helpful when visualization out of view (4), supports visual feedback (4), extra sensory channel alongside visual input (4) |
| - pattern ambiguous (9) and needs concentration (6), requires training (6) |
| - more focus on visualization, if present (6), ignored with habituation (3) |
| - leads to information overload (3), not helpful (2), disturbing (2), and stressful (2), information delivery too slow (2), difficult as sole feedback modality (2) |
| • preferable as alarm for $F_{max}$ (6), pattern should increase frequency with force (2) |

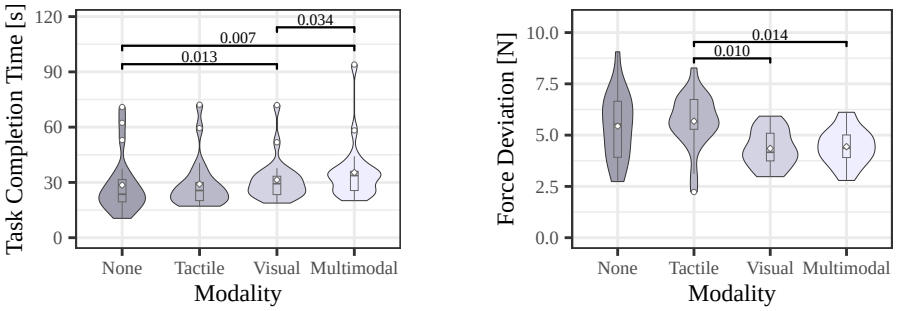
#### 8.3.3.4 *Interpretation of Results*

This section aims to interpret the quantitative and qualitative results obtained for each of the investigated factors.

MODALITY The significant results from pairwise comparisons of objective measures across the feedback modalities are depicted in Figure 8.12. The TCT results show significant differences across feedback modalities, with visual feedback leading to higher TCT than the baseline, indicating that visual feedback slows down interaction (see Figure 8.12a). Therefore, the null

hypothesis $H_0^3.1$ was rejected in favor of the corresponding alternative hypothesis $H_1^3.1$. This could be attributed to the additional cognitive load induced by processing visual information, particularly in conjunction with analyzing sonographic images for navigation. Furthermore, TCT was significantly higher with multimodal feedback than without or with visual feedback alone, as indicated by both a significant effect in the analysis of the *modality* factor and a significant main effect of the *vibrotactile* factor in the three-factorial design. Multimodal feedback resulting in even slower performance may be attributed to information overload and the heightened concentration required to process the combined feedback, as reported by participants during the semi-structured interviews.

In terms of AAD, the results indicate a significantly higher force deviation for tactile feedback alone compared to multimodal and visual feedback (see Figure 8.12b). Thus, the null hypothesis $H_0^3.2$ was rejected in favor of the corresponding alternative hypothesis $H_1^3.2$. This discrepancy may originate from the slower perception of information during tactile feedback, particularly evident when manipulating the probe, where force changes take longer to be perceived relative to the movement speed. Moreover, participants reported difficulties in recognizing the pattern of tactile feedback, potentially leading to confusion between patterns representing forces that are too high or too low. This confusion might have resulted in participants inadvertently adjusting the force in the wrong direction. Notably, there was no significant difference in AAD between the multimodal and visual feedback conditions, suggesting a tendency among participants to disregard vibrotactile feedback when visual feedback was present. This is consistent with feedback during the interview. As no significant effects of the *modality* on AMax were found, $H_0^3.3$ could not be rejected.

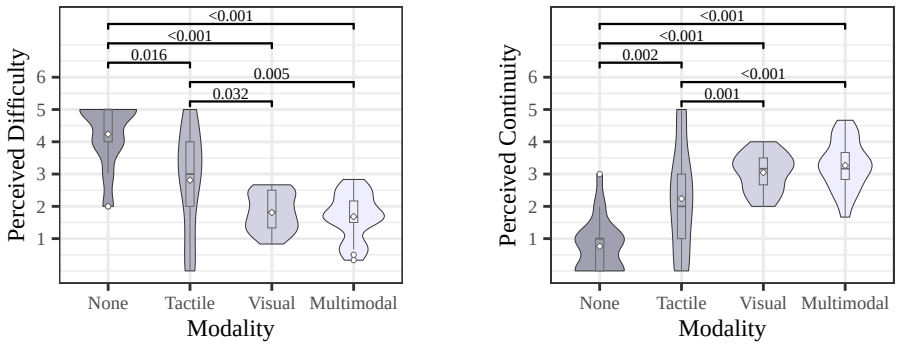



(a) Task completion time (b) Force deviation

Figure 8.12: Objective measures across feedback modalities. ◇ indicate means. Statistically significant pairwise comparison results are indicated by brackets. Adapted from [6].

Significant effects of *modality* on both subjective measures were found (see Figure 8.13). Thus, the null hypotheses $H_0^3.4$ and $H_0^3.5$ were rejected in favor of the corresponding alternative hypotheses $H_1^3.4$ and $H_1^3.5$. The subjective measures reveal improvements in perceived difficulty and continuity for all feedback conditions compared to the baseline no-feedback condition. Moreover, results indicate that both visual and multimodal feedback significantly reduce perceived difficulty and enhance perceived continuity compared to

vibrotactile feedback alone. These findings align with the results for AAD and suggest that ambiguities regarding the vibration pattern may contribute to lower user satisfaction with vibrotactile feedback. In contrast, visual feedback might provide a more direct and clearer information representation, leading to enhanced performance and higher user satisfaction. Notably, while multimodal and visual feedback significantly increase completion time, they reduce perceived difficulty. This apparent discrepancy may be explained by the fact that providing feedback reduces perceived difficulty by increasing participants' ability to assess the force applied. However, this may also slow down the interaction as participants need to respond to the feedback to adjust their force application.




(a) Perceived Difficulty (b) Perceived Continuity

Figure 8.13: Subjective measures across feedback modalities. ◇ indicate means. Statistically significant pairwise comparison results are indicated by brackets. Adapted from [6].

VISUALIZATION TYPE Significant main effects regarding visualization type are depicted in Figure 8.14. Therefore, the null hypothesis $H_0^1.2$ was rejected in favor of the corresponding alternative hypothesis $H_1^1.2$.

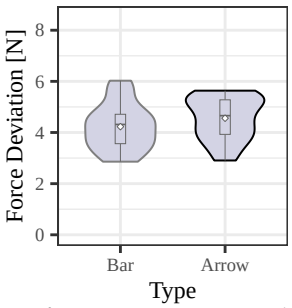
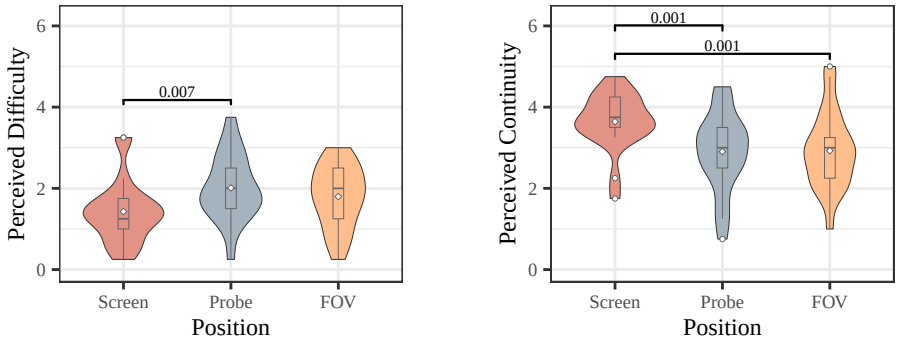



Figure 8.14: Main effects of *Type* factor. Means are indicated by ◇. Adapted from [6].

The pairwise comparison indicates that the bar visualization resulted in significantly lower AAD than the arrow. This could be due to the discrete coloring and subdivision, which effectively highlights $F_d$. The lower force deviation for the bar is consistent with user feedback describing the bar as precise and sensitive. Despite participants' reported preference for the arrow's gradual changes in size and color, this feature may have contributed to a less accurate perception of small force deviations. Especially as the arrow

becomes increasingly smaller as the currently applied force approaches $F_d$, making subtle changes more difficult to identify. As no other significant main effects for visualization type were found, the null hypotheses $H_0^1.1$, $H_0^1.3$, $H_0^1.4$, and $H_0^2.5$ could not be rejected.

VISUALIZATION POSITION For the visualization position, significant main effects were found for perceived difficulty and perceived continuity (see Figure 8.15). Thus, the null hypotheses $H_0^2.4$ and $H_0^2.5$ were rejected in favor of the corresponding alternative hypotheses $H_1^2.4$ and $H_1^2.5$. The pairwise comparison revealed that the screen position had a significantly lower perceived difficulty in maintaining $F_d$ and a significantly more continuous feedback perception compared to the probe position. This could be explained by the statements given during the semi-structured interview, where although some participants reported looking at the probe to observe the robot's movements, the majority stated mainly looking at the sonographic image on the screen and therefore preferring the feedback on the screen. Regarding FOV positioning, pairwise comparisons showed significantly less continuous feedback perception compared to screen placement. Despite the visualization being always within the FOV due to its body-stabilized position, participants found it out of focus, requiring a conscious gaze shift. This might be attributed to perceptual issues, as the visual feedback was positioned at a different depth than the task-relevant content, making it harder to perceive. Participants also reported confusion with the dynamic positioning, which likely further reduced the perceived continuity of force perception. No other significant main effects were found; thus, $H_0^2.1$-$H_0^2.3$ could not be rejected.




(a) Perceived Difficulty (b) Perceived Continuity

Figure 8.15: Main effects of *Position* factor. Means are indicated by ◇. Statistically significant pairwise comparison results are indicated by brackets. Adapted from [6].

INTERACTIONS BETWEEN FACTORS In addition, a three-way interaction was observed for AMax (see Figure 8.16). The effect descriptively suggests that without vibrotactile feedback, particularly the bar visualization on the screen, performs worse than the other combinations of *type* and *position*. This finding seems to be in contrast to the subjective feedback, where a preference for screen position was described. However, a possible explanation could be that although the screen display enables continuous perception, it primarily

engages peripheral vision. Since the bar visualization employs a discrete color scheme, changes leading to maxima may not be immediately perceived if the focus is not directly on the visualization, potentially resulting in higher maxima. Other visualization placements may require a more deliberate shift in focus, leading to a more accurate perception of changes.

When vibrotactile feedback is available, this trend shifts, with the bar visualization at the screen position performing better than at the probe position. This aligns with the results for perceived difficulty, where recognizing the applied force at the probe position is rated as more challenging. Vibrotactile feedback, therefore, appears to enhance the perception of the bar visualization on the screen by potentially alerting participants to changes and guiding their attention to deliberately shift towards the visualization. The influence of vibrotactile feedback and visualization position on AMax for arrow visualization appears to be less pronounced. This may be attributed to changes in size and color being easily perceived even when only peripherally visible, making AMax relatively independent of other factors for arrow visualization. This is consistent with participants' feedback, which described the arrow's changes as perceptible peripherally and noted the striking color scale.

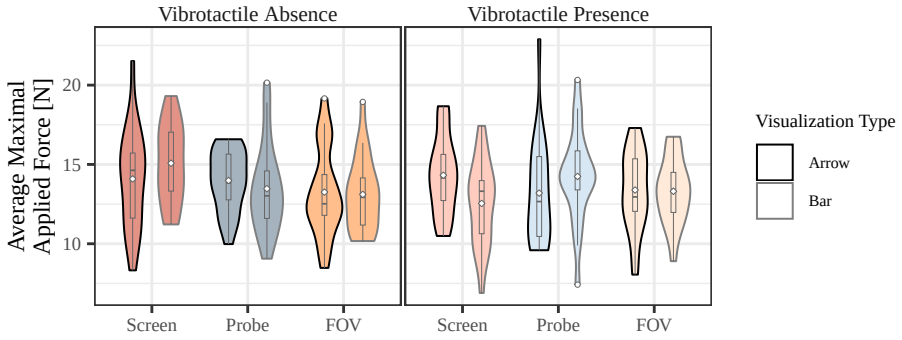



Figure 8.16: Three-way interaction effect for AMax. Means are indicated by ◇. Black outlines represent the *arrow* and gray outlines the *bar* visualization. Adapted from [6].

### 8.3.4 *Discussion*

This section discusses the study findings, contextualizes them within related work, addresses the study's limitations, and summarizes the results into design implications.

MODALITY The study results indicate that while participants found vibrotactile feedback intuitive and helpful, force error, perceived difficulty, and continuity were primarily dependent on the presence of visual feedback. This indicates a tendency for participants to prioritize visual over vibrotactile sensory substitution, which aligns with the qualitative feedback indicating neglect of vibrotactile cues. These results are consistent with previous research by Howard et al. [145], reporting superior performance with visual compared to vibrotactile feedback. Additionally, in line with Jonetzko et

al. [157], a reduced efficiency with multimodal feedback was observed, indicating information overload. A major limitation of this study is employing only one type of vibrotactile feedback. The scaling approach was selected based on literature highlighting its suitability for vibrotactile feedback and its assumed comparability to the information density of visual feedback. However, discrepancies between visual and vibrotactile scaling may have influenced force perception. Additionally, previous work suggests that while non-linear scaling can improve accuracy, it may increase perceived difficulty and completion time [135]. Therefore, further investigation is needed into the scaling and the alignment of different scaling methods and their impact on task performance. Unlike the employed relative force magnitude vibrotactile pattern, Chan et al. [60] employed constant patterns to indicate force direction alone, achieving superior performance compared to visual feedback. Given that participants expressed a preference for constant vibration for excessive force feedback, or increasing frequency with increasing force, exploring these alternatives could yield different results.

VISUALIZATION TYPE Similar force visualizations were compared in the recent work of Song et al. [308][1]. Although their arrow implementation differed, using discrete rather than continuous coloring and representing absolute force through arrow length, their linear gauge concept shared similarities with the bar concept investigated in this section. They found that the linear gauge improved task ease, workload, and completion time compared to the arrow designs, which they attributed to clear boundary markings and low visual complexity. While the study reported in this section found significant effects only on force deviation rather than on subjective factors, the general findings align with those of Song et al., highlighting the effectiveness of visual feedback that conveys force through height and color changes.

Differences in results between this section's study and that by Song et al. may stem from variations in specific design elements. These design choices also influenced overall outcomes, such as the bar visualization supporting more accurate force application, while participants reported better peripheral perceptibility with the arrow visualization. However, comparing two distinct approaches, including different characteristics such as 2D versus 3D and discrete versus continuous color and shape changes, makes it difficult to isolate the effects of individual variables. Future research could therefore focus on these specific aspects to gain more targeted insights.

The visualizations additionally differed in the information displayed, with the arrow indicating the force direction. Although the participants described the directional display as helpful, it did not significantly impact quantitative measurements, likely due to task simplicity. Future studies with more complexity may yield different outcomes.

While the study results indicate that the bar visualization allowed for smaller deviations from $F_d$ to be perceived, some participants noted that the threshold for the green area was too small, making the bar feel overly sensi-

1 The work of Song et al. [308] was published online on November 28, 2024, while the content of this section was submitted to IEEE VR on September 11, 2024.

tive. Although the threshold was designed to enhance precise perception of changes around the desired force, the rapid changes in indicator color could lead to difficulties in force interpretation. Therefore, future work could evaluate the impact of increasing the size of the visualization threshold on both the accuracy of force application and user feedback, to determine whether user satisfaction can be improved without compromising accuracy. Future work could additionally explore scaling approaches, such as logarithmic and linear scaling as done in Heinrich et al. [135], to assess their impact on task performance.

VISUALIZATION POSITION Results indicate that the suitability of visualization placement is dependent on the operator's focus during the task. While some participants reported looking at the probe during interaction to perceive the robot's reaction, the majority indicated looking at the probe for initial orientation but then focusing primarily on the screen. This aligns with previous findings regarding feedback placement, which found that force visualization at task-relevant locations can enhance force perception [82]. Beyond placement alone, the results also suggest that guiding the user's attention towards the visualization may further improve effectiveness. This could be achieved through vibrotactile cues or, alternatively, by using visual attention guidance in XR [295].

GENERAL While the sensory substitution technique design is based on literature findings, qualitative feedback suggests possible design flaws that may have influenced study results. However, since the main feedback primarily concerns the nature of the techniques rather than specific design decisions, it is believed that these flaws do not diminish the overall significance of the findings. Some of these findings, such as the risk of tactile clutter reducing comprehension and sensory overload with vibrotactile feedback, have been previously investigated in studies focusing on multimodal feedback design [323]. Nevertheless, this study provides valuable insights by offering empirical evidence, replication, and validation for a new context.

The presented study employed an optical see-through HMD for feedback visualization, which has a high implementation cost. This was chosen to ensure consistent visual properties across different visualization placements, allowing comparability. However, based on the study's results showing the benefits of placing the visual feedback adjacent to the ultrasound image, this feedback could be implemented more cost-effectively by integrating it into the display. While the HMD in this study also supported hand tracking and interaction, alternative feedback visualization approaches may be more viable in scenarios where different input methods are used for robot control.

To compare unimodal and multimodal feedback, two baseline conditions were included. However, more data were available for the other conditions than for the baseline, leading to an unbalanced design. To ensure a balanced sample for the statistical tests, the data were aggregated. Since no interaction effects were found for TCT, AAD, perceived difficulty, and continuity, it is believed that this aggregation did not result in information loss, and the

results remain valid. Notably, only 21 participants were included in the study. While the repeated measures per factor level increase the power of the analysis due to the factorial design, the interaction effects have lower power, and additional participants are required to examine these effects robustly.

#### 8.3.4.1 *Limitations*

The main limitation of this study lies in its specificity to robotic ultrasound, which may restrict the generalizability of findings to other HRI contexts. However, several insights may extend to scenarios where direct force perception is limited, such as laparoscopy or industrial robotic tasks like polishing and grasping. While some adaptation may be required, the core findings on feedback modality, visualization design, and placement are likely transferable.

To apply the visual feedback to different contexts, alternative scaling strategies might be needed to represent subtle force fluctuations near target values accurately. For instance, applications involving higher forces (e.g., industrial settings) or requiring greater precision (e.g., surgical tasks) could benefit from nonlinear scaling near the target force. In tasks where achieving a specific force is not the goal, such as robotic grasping, where absolute force levels are more important, the feedback could be adjusted by setting $F_d = 0$, allowing it to reflect absolute forces directly.

Regarding visualization placement, the importance of a screen in co-located interaction is specific to robotic ultrasound. While probe- and screen-based placements are application-dependent, generalizable insights can be drawn from world- and head-stabilized sensory substitution techniques. Overall, although the strategies developed here are believed to have broader relevance, further research is needed to validate their effectiveness in other domains.

#### 8.3.4.2 *Implications*

In summary, based on the study results, the following design implications are proposed for integrating sensory substitution to enhance force perception during HRI, which can be generalized beyond robotic ultrasound:

**Employ visual sensory substitution to enhance force perception.** Study findings suggest that any form of sensory substitution significantly reduces force perception difficulty compared to no feedback. Additionally, the presence of visual feedback was shown to be particularly effective in lowering force errors and difficulty while enabling continuous perception of applied force. Regarding multimodality, care should be taken in feedback design, as the combination of multiple modalities can reduce task efficiency by increasing task duration. However, incorporating discrete vibrotactile feedback to complement visual feedback by providing alerts for changes may be beneficial in optimizing performance.

**Utilize bar visualizations to accurately depict force.** The study results indicate that representing force magnitude with a bar visualization, where the size is proportional to force magnitude and the deviation from a desired force is indicated by discrete coloring, offers a clear representation and helps

minimize force errors. However, if the visualization needs to be peripherally perceivable, mapping force error onto size and incorporating gradual color changes could enhance perceptibility.

**Employ world-stabilized placement of virtual content near task-relevant locations.** The findings suggest that placing virtual content on the screen displaying ultrasound data reduces the perceived difficulty and enhances continuous perceptibility. Qualitative feedback indicates a user preference for world-stabilized visualization placement. In contrast, head-stabilized placement moves with the user's head, causing the visualization to shift within the environment and fall outside the direct line of sight, requiring a conscious gaze shift and impairing perception.

### 8.3.5 *Conclusion*

In this study, different XR concepts for sensory substitution to enhance force perception during HRI were investigated. These concepts differed in modality, visualization design, and placement. The suitability of unimodal and multimodal visual and vibrotactile approaches was investigated through a user study assessing task completion time, average maximal applied force, average force error, and subjective feedback for the exemplary use case of robotic ultrasound. The results indicate that any form of feedback improves force perception, with visual feedback being particularly effective in reducing force errors and being preferred by users. Specifically, a 2D bar visualization, changing height and color, was found to be better at reducing force errors compared to a 3D arrow visualization displaying relative force through gradual color, direction, and size changes. Moreover, users subjectively preferred the display of feedback on the screen displaying the sonographic image over positioning it in their FOV or at the ultrasound probe. These results guided the formulation of guidelines for sensory substitution in HRI and are an important step towards improved force perception in HRI.

# 9 SENSOR PLACEMENT AND CONTROL MODES

## 9.1 INTRODUCTION

As described in Section 7.2, previous work has identified fatigue during hand gesture interaction for HRI as a key challenge [44, 373]. This effect is commonly referred to as the *gorilla arm effect*, which describes the fatigue experienced in users' arms during prolonged use [45]. This issue is also relevant in hand gesture interaction within XR. In this context, researchers have found that fatigue is influenced by the *interaction space*, i.e., the range of motion and the specific area in which the hand must move during interaction [140].

When using optical sensors for hand tracking, spatially fixed sensors define specific interaction spaces in which hands can be reliably tracked and gestures performed. While hand interaction can also be facilitated using handheld motion-tracking controllers [183, 236, 316], which allow for more flexible input spaces, vision-based approaches enable touchless, "come as you are" interaction. By removing the need for users to be equipped with additional hardware, these approaches support more spontaneous and accessible interaction [256, p. 524]. Different sensor placements have been explored for HRI, including head-mounted cameras [61], sensors placed in the robot's environment [29, 62, 174], sensors mounted on the robot itself [15], wrist-mounted devices [328], and handheld controllers [183, 236, 316]. However, the impact of these sensor placements on fatigue and interaction efficiency has not yet been systematically investigated.

Beyond sensor placement, the interaction space also depends on the required range of motion, which is influenced by the control mode. Different control mappings, such as position and rate control, and gesture designs, can result in varying ranges of motion, potentially affecting user fatigue. While comparisons of control modes have been conducted in the context of haptic input devices [169, 336] and head gestures [275], no direct comparison has yet been made for hand gesture interaction.

To address this gap, this chapter investigates the influence of sensor placement and control mode on interaction efficiency and user fatigue in gesture-based HRI, thereby addressing RQ2.2.

> **RQ2.2** | How can fatigue be reduced in hand gesture-based HRI?

Parts of this chapter were published in: Tonia Mielke, Florian Heinrich, and Christian Hansen. "Gesturing Towards Efficient Robot Control: Exploring Sensor Placement and Control Modes for Mid-Air Human-Robot Interaction." In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 19–23. DOI: 10.1109/ICRA55743.2025.11127519 [5], © 2025 IEEE.

### 9.1.1 *Related Work*

In the literature, different interaction spaces and control methods have been proposed. However, prior work specifically addressing these aspects in the context of gesture-based robot control remains limited. Therefore, this section reviews relevant studies on interaction spaces for hand gesture interaction and control modes in application areas beyond manipulative robot control using hand gestures.

#### 9.1.1.1 *Interaction Spaces*

Different strategies have been proposed to optimize gesture poses within the FOV of HMDs. Hincapié-Ramos et al. [140] developed a metric to quantify arm fatigue in hand gesture interaction, evaluating factors such as plane location, arm extension, plane size, and selection method. Building on this, Belo et al. [97] introduced a visualization toolkit to help creators design ergonomic UIs by highlighting the ergonomic costs of different interaction zones. Schön et al. [289] extended this research to complex rotation tasks, finding that placing XR content closer to the user's body increased comfort. Feuchtner and Müller [103] explored integrating an additional sensor to the HMD, enabling users to interact at waist level for prolonged interaction while their virtual hand remained visible in an overhead position.

Beyond optimizing interactions within the FOV, other approaches have explored alternative sensor placements. Siddhpuria et al. [305] investigated gesture input at the user's side, using a hands-down posture enabled by a smartwatch. Their study reported high user satisfaction, low fatigue, and strong memorability when interacting with large displays. Similarly, Liu et al. [197] explored bimanual interaction at a hands-down position using sensors on both thighs, demonstrating the technical feasibility of this approach. Brasier et al. [49] expanded these findings, comparing different hands-down sensor placements, such as wrist, waist, or leg, for 2D cursor control. They found that indirect input, remapping hand movements from off-screen planes, achieved similar performance to direct input while reducing fatigue.

#### 9.1.1.2 *Control Methods*

In the context of gesture-based robot control, related work can be broadly categorized into two types of interaction: discrete and continuous gestures. Discrete gestures are symbolic or command-based, triggering distinct actions such as selecting an object or initiating a robot behavior. These gestures can include, for example, pointing gestures used to define discrete end effector goal positions [56, 148, 264], to indicate movement directions for the robot [31, 102], or a set of gestures representing different commands [265].

In contrast, continuous gestures involve real-time control of the robot's motion based on ongoing hand movements. These include techniques such as position control and rate control (see Section 2.1.1.3). In position control, hand gestures are directly mapped to the motion of the robot, either at the joint level [316, 357] or the end effector level [61, 170, 174, 183]. In contrast,

rate control maps changes in hand position and orientation to corresponding end effector velocities rather than absolute positions [73, 236, 310, 328].

Outside of hand gesture interaction, previous work has compared position control and rate control for HRI across different input modalities. Kim et al. [169] investigated these modes in a pick-and-place scenario using joysticks, reporting that position control outperformed rate control in terms of performance. However, they noted the potential benefits of rate control for broader movement tasks. Wang et al. [336] demonstrated that position control led to faster and more accurate needle insertions in robotic needle steering, while also reducing cognitive load. Similarly, Rudigkeit et al. [275] compared position and rate control using head gestures. They found that position control enabled faster and more intuitive interaction, whereas rate control improved accuracy, which could enhance operational safety. Mick et al. [218] compared rate and position control using an isometric input device in a 3D target-reaching task. Their results showed that rate control outperformed position control in terms of task performance.

Beyond robotics, position and rate control have also been compared for 2D input tasks. For example, Crossan et al. [77] evaluated control modes for head gestures in mobile applications and found position control to be faster and more accurate than rate control. Similarly, Dougherty et al. [89] found position control to be more efficient than rate control for foot gestures in a 2D positioning task. Teather et al. [315] compared position and rate control for 2D pointing using tilt interaction and likewise found that position control offered faster and more accurate performance.

### 9.1.2 *Research Gap*

Previous studies have demonstrated the benefits of indirect interaction spaces in reducing user fatigue during hand gesture interactions. However, these studies focused on 2D interfaces [49, 140]. Therefore, the applicability of these techniques for 3D robot manipulation tasks remains unexplored. In addition, hand tracking can be performed close to the robot by placing sensors directly on the robot itself. Although this approach has proven effective for recognizing deictic gestures in HRI [15], its potential for enabling direct robot control through hand gestures has not been evaluated.

Moreover, related work explored the suitability of position and rate control for robot operation using external input devices [169, 218, 336] and head gestures [275]. However, these control paradigms have not yet been systematically investigated in the context of hand gestures, particularly with respect to their impact on fatigue.

Together, these gaps highlight the need to investigate how different interaction spaces and position and rate control can be applied to hand gesture-based control for three-dimensional robot manipulation, and especially how these factors influence fatigue during HRI.

## 9.2 TECHNICAL METHODS

*Supplementary video*

To investigate the influence of different interaction spaces on the efficiency of hand gesture-based HRI, this section examines two key factors: *sensor placement* and *control mode*. The technical methods used to implement these aspects are described in the following.

### 9.2.1 *Sensor Placements*

The most common sensor placements in the literature involve placing the sensor in the environment (e.g., on a *desk*) [29, 62, 174] or on the *head* (e.g., using an HMD) [61]. These two placements are used as benchmarks in this study. In addition, two alternative body-mounted placements are explored. The first is on the user's *leg*, which has been shown to be effective for display interactions [197, 305]. This placement allows users to interact with their hands at their sides while maintaining a relaxed, arms-down posture. The second is at the *waist*, another promising approach that facilitates interaction while keeping the arms low [49]. Finally, placing the sensor directly on the *robot* is considered, a method that has proven effective for robot control using deictic gestures [15]. An overview of the different placements is shown in Figure 9.1 and Figure 9.2.

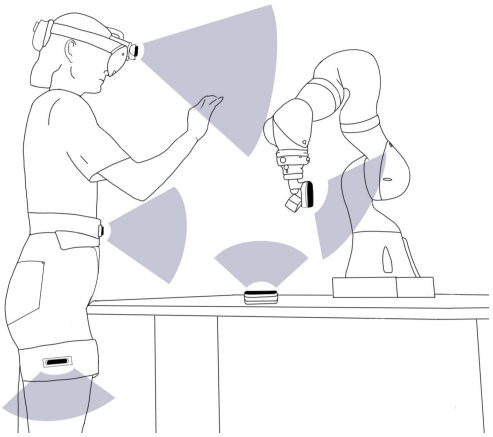


Figure 9.1: Overview of sensor placements. Reprinted from [5], © 2025 IEEE.

### 9.2.2 *Registration*

To enable flexible interaction spaces, an external hand-tracking sensor is employed, which can be positioned at the previously described placements. To correctly map tracked hand motion to robot movement, the sensor and the robot's coordinate frames must be aligned. Part I demonstrated that point-based registration can accurately establish this alignment. However, it requires the robotic end effector to be localizable in both coordinate spaces. As the hand-tracking sensor does not support marker tracking to determine the end effector position, an HMD is employed to mediate the registration.

The transformation $T^R_{HMD}$ between the HMD and the robot $R$ is calculated using a pivot-calibrated image marker attached to the robotic end effector and a point-based registration approach (see Section 2.3.4). Then, the transformation $T^{HMD}_S$ between the sensor $S$ and the HMD is derived using the hand-tracking capabilities of both devices. By positioning the hand within the FOV of both devices, its position can be simultaneously recorded in each coordinate frame, allowing four corresponding points to be collected for point-based registration. Based on $T^R_{HMD}$ and $T^{HMD}_S$, the transformation $T^R_S$ between the sensor and the robot can then be computed.

$$T^R_S = T^R_{HMD} \cdot T^{HMD}_S \tag{9.1}$$

The relationship between the sensor and the robot coordinate frames is dynamic for certain sensor placements. For example, body-mounted sensors are affected by user movement, while robot-mounted sensors are influenced by end effector motion. Therefore, continuously updating the sensor's relative position and orientation is essential. Since the sensor itself has no inside-out tracking capabilities, the fixed relative position between the sensor and its mounting point, either the user or the robotic end effector, is used as a reference. For body-mounted sensors, the HMD's inside-out tracking provides real-time updates of the user's position relative to the robot. In robot-mounted configurations, the sensor's motion is tracked using the robot's kinematic model. These real-time position updates are used to continuously adjust the registration between the sensor and the robot coordinate frame, ensuring accurate alignment throughout the interaction.

### 9.2.3 *Robot Control*

The robot control is implemented as described in Section 2.3.2, using a pinch gesture to activate the interaction and enabling switching between translation, rotation, or full 6-DOF control via a foot pedal. Two different mappings between the manipulative hand gesture and robot motion are investigated: position control and rate control.

#### 9.2.3.1 *Position Control*

In position control, changes in hand position and orientation are directly mapped, scaled by a factor $k$, to the robotic end effector movements. The target end effector position $P^E_n$ and rotation $R^E_n$ are therefore calculated using the robot's initial position $P^E_0$ and rotation $R^E_0$, along with the change in hand position between mode activation $P^H_0$ and the current frame $P^H_n$, and the relative rotation between mode activation $R^H_0$ and the current frame $R^H_n$, denoted as a quaternion (see Equation 9.2 and Equation 9.3).

$$P^E_n = P^E_0 + k \cdot (P^H_n - P^H_0) \tag{9.2}$$

$$R^E_n = k \cdot (R^H_n \cdot (R^H_{n-1})^{-1}) \cdot R^E_0 \tag{9.3}$$

#### 9.2.3.2 *Rate Control*

In rate control, relative hand positions and orientations are mapped to end effector velocities using a transfer function. To prevent small hand movements from causing involuntary interactions, a threshold $\Delta X_{min}$ is employed. Beyond this threshold, the relative position/orientation $\Delta X$ to the start position/orientation is linearly scaled to velocities until a maximum value $\Delta X_{max}$ is reached (see Equation 9.4).

$$s(\Delta X) = s_{max} \cdot \begin{cases} 0 & , \quad \Delta X \leq \Delta X_{min} \\ 1 & , \quad \Delta X \geq \Delta X_{max} \\ \dfrac{\Delta X_n - \Delta X_{min}}{\Delta X_{max} - \Delta X_{min}} & , \quad \text{else} \end{cases} \tag{9.4}$$

Using this equation, the current hand displacement $\Delta P^H = P_n^H - P_0^H$ can be used to calculate the end effector's goal position $P_n^E$ (see Equation 9.5). Similarly, the rotation $\Delta R^H = R_n^H - R_0^H$ is used to scale the angular displacement, with the rotation $R(angle, axis)$ being applied around the axis $A$ defined by $\Delta R^H$ (see Equation 9.6).

$$P_n^E = P_{n-1}^E + s_n(\Delta P^H) \cdot \frac{\Delta P^H}{||\Delta P^H||} \tag{9.5}$$

$$R_n^E = R(s_n(\Delta R^H), A) \cdot R_{n-1}^E \tag{9.6}$$

### 9.2.4 *Implementation*

The implementation was realized as described in Section 2.3. Hand tracking was enabled using a *Leap Motion Controller 2*. Sensor placements were managed using 3D-printed mounts and Velcro fasteners for the waist and leg (see Figure 9.2). For additional input, a foot pedal with two switches was used: one to toggle the robot control mode and the other to confirm task completion. The application for hand tracking, registration, and gesture-to-robot control was developed in *Unity* and ran on a desktop PC. A *HoloLens 2* was used for interface display, task visualization, and registration, with a Unity-based application that utilized *Vuforia* for marker tracking. Robot motion was executed using a high-stiffness impedance control scheme for precise Cartesian movements.

(a) Desk (b) Head (c) Leg (d) Waist (e) Robot

Figure 9.2: Setup for different sensor placements. Reprinted from [5], © 2025 IEEE.

## 9.3 EVALUATION METHODS

To assess the suitability of different sensor placements and control methods for the hand gesture-based control of a robotic arm, a within-subjects user study with repeated measures was conducted. Each participant tested every combination of sensor placement and control method across four repetitions of the evaluation tasks, resulting in 32 trials per person (4 placements $\times$ 2 control methods $\times$ 4 tasks). Additionally, a preliminary study was conducted to determine relevant parameters for the robot control.

### 9.3.1 *Preliminary Study*

To optimize the robot control modes, a preliminary pilot study was conducted with 5 participants (2 female, 1 diverse, 2 male), all of whom had a technical background. The goal was to determine optimal parameters for both position control and rate control. For position control, parameters included the mapping ratio of hand translation and rotation to end effector movement. For rate control, the range of hand motion was examined.

For position control tuning, participants were presented with three different mappings (1:1, 1:0.75, and 1:0.5) for translation and rotation. They were then asked to rank these mappings based on their experience. For rate control, participants were asked to move their hands from an individually defined resting position in each Cartesian direction and to rotate them around each individual axis, returning to the resting position after each movement. The transfer function values were then defined by determining the minimum of the range values and the maximum deviation from the resting position for each participant, and averaging these values across participants.

In terms of position control scaling, participants preferred 1:1 scaling for translation and 1:0.5 scaling for rotation. They found that unscaled translation provided a more efficient and natural interaction, while scaled-down rotation made the robot's behavior feel more predictable and easier to control. The latter preference was likely since small end effector rotations cause larger movements in the robot's joints, making unscaled rotation perceived as too fast. These optimized scalings, along with motion ranges for different sensor placements, were applied in the final study. When evaluating sensor placements, the waist position had significantly limited rotational freedom, likely due to the sensor's difficulty detecting the hand unless the palm faced it. Consequently, rotation becomes uncomfortable, restricting the range of motion and impacting the feasibility of this placement. Therefore, the waist placement was excluded from further investigation. It is worth noting that using different sensors capable of recognizing hands in different orientations could potentially mitigate this issue, making waist placement a more viable option in future studies.

### 9.3.2 *Tasks*

To evaluate the suitability of the different control methods, a precision task was used, as precision has been shown to cause a high level of fatigue in HRI [195]. To additionally require both translation and rotation, an alignment task was chosen, where participants aligned the robotic end effector with a target position. Therefore, a 3D-printed flange with a specific geometric shape was employed, and the target position and rotation were displayed as a hologram of the end effector in AR (see Figure 9.3).

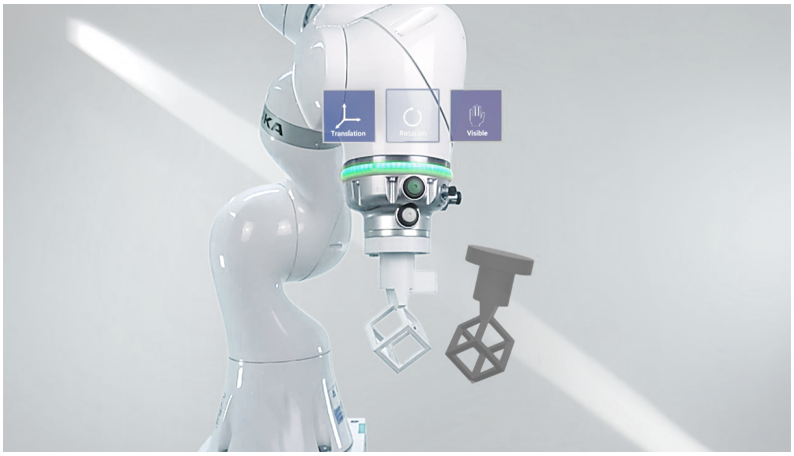



Figure 9.3: UI during study task displaying target hologram and feedback about active mode and hand visibility. Reprinted from [5], © 2025 IEEE.

### 9.3.3 *Variables*

INDEPENDENT VARIABLES The independent variables of the two-factorial study design were the *sensor placement* (desk, head, leg, robot) and the *control mode* (rate control, position control).

DEPENDENT VARIABLES To assess performance and user perception, both objective and subjective measures were collected. For performance measures, the TCT was recorded, starting from the first detected gesture to participants confirming task completion via a foot pedal press. In the final position, the translation error was measured as the Euclidean distance, and the rotation error was assessed as the shortest angle in degrees between the two rotations. Workload was assessed using the NASA TLX [131], and fatigue was assessed using the Borg CR10 [43] questionnaire.

### 9.3.4 *Hypotheses*

The study was exploratory in nature, with no prior assumptions made regarding the outcomes. As a result, two-sided hypotheses were examined. Based on the variables described above, the following null and alternative hypotheses were considered:

$H_0^y.x$ The mean $x$ is equal for all levels of $y$.
$H_1^y.x$ The mean $x$ differs for at least one level of $y$.

$$x \in \begin{cases} \text{1-TCT} \\ \text{2-Translation Error} \\ \text{3-Rotation Error} \\ \text{4-TLX} \\ \text{5-Borg CR10} \end{cases}$$

$$y \in \begin{cases} \text{1-Control Methods} \\ \text{2-Sensor Placement} \end{cases}$$

### 9.3.5 *Sample Design*

As this user study investigated a general precision task, no prior task-specific knowledge was required. Therefore, participants with a technical background were recruited from the local university through online polls. Each participant received a compensation of 15 €.

### 9.3.6 *Procedure*

Upon arrival, participants were asked to complete a written informed consent form and provide demographic information. They were then given a brief introduction to the general purpose of the study and the task to be performed. The study began with the first trial for the first sensor position and control method. The order of sensor positions and control methods was counterbalanced, with each participant performing both control methods for the same placement sequentially in a fixed order. Each trial consisted of a training and an evaluation block separated by a short break. During each of the blocks, participants were presented with four target positions and were instructed to align the end effector with the virtual targets. During the training block, the method was explained, and questions could be asked. Following each trial, participants completed the NASA TLX and Borg CR10 questionnaires. A short break followed each trial until participants felt comfortable to continue, to minimize the carry-over effects of effort. After completing all eight trials (4 sensor positions $\times$ 2 control methods), a semi-structured interview was conducted. The study took an average of 60 minutes.

### 9.3.7 *Statistical Analysis*

First, Shapiro-Wilk and Levene's tests were performed for each measure to assess normality and homogeneity assumptions, respectively. Across variables, these assumptions were not met. Therefore, robust two-way ANOVAs for within-subjects designs based on trimmed means were calculated (see [348]). The $\delta_t$ estimate proposed by Algina et al. [16] was employed for effect sizes. Post hoc paired sample Yuen's tests based on trimmed means with Holm corrections for multiple testing were conducted as pairwise comparisons between the sensor placement factor levels. All statistical analyses were conducted using R (version 4.4.0).

## 9.4 RESULTS

This section outlines both the quantitative and qualitative results obtained from the user study.

### 9.4.1 *Participants*

A total of 29 participants (19 male, 7 female, 3 non-binary) took part in the study. Participants' ages ranged from 20 to 39 years (M=27.38, SD=4.02), and their heights ranged from 162 cm to 197 cm (median: 174 cm). The sample consisted of undergraduate students (17), graduate students (6), and engineers (5). To standardize the study setup, only right-handed participants were recruited. Participants were asked to self-assess their technical affinity, motor skills, and task-relevant experience on a 5-point Likert item, where 5 indicates a high level of affinity, skill, or experience. Participants reported their technical affinity (M=4.31, SD=0.60), motor skills (M=3.93, SD=0.65), experience in HRI (M=2.62, SD=0.94), experience in XR (M=3.10, SD=1.42), and experience in hand gesture interaction (M=2.38, SD=1.32).

### 9.4.2 *Quantitative Results*

The statistical analyses revealed significant main effects of sensor placement and control mode across multiple variables. The only exception was the control mode factor on translation error, which did not yield significant differences. Additionally, no statistically significant interactions were observed. Therefore, only descriptive results for each individual factor condition are reported in Table 9.1. Additional illustrations of all descriptive results are shown in Figure A.2, and a summary of all statistical outcomes is presented in Table 9.2. Pairwise comparison results are illustrated in the plots (see Figure 9.4 and Figure 9.5).

Table 9.1: Summary of descriptive results for all dependent variables (n=29). Entries are in the format: mean value [standard deviation].

| **Variable** | **TCT (s)** | **Translation Error** | **Rotation Error** | **TLX** | **Borg CR10** |
|---|---|---|---|---|---|
| **Placement** | | | | | |
| *Desk* | 51.72 [18.68] | 20.52 [8.02] | 9.50 [3.96] | 28.03 [17.11] | 1.88 [1.82] |
| *Head* | 52.96 [20.83] | 25.37 [15.20] | 9.94 [3.87] | 29.35 [17.20] | 2.09 [1.93] |
| *Leg* | 69.12 [24.41] | 24.18 [11.57] | 11.85 [6.77] | 37.36 [20.84] | 2.07 [1.95] |
| *Robot* | 53.23 [17.69] | 18.73 [8.76] | 8.96 [3.25] | 32.00 [20.17] | 2.77 [2.16] |
| **Control Mode** | | | | | |
| *Position Control* | 49.07 [17.57] | 21.60 [11.83] | 9.78 [4.86] | 27.11 [17.46] | 1.91 [1.68] |
| *Rate Control* | 64.44 [22.67] | 22.81 [11.16] | 10.34 [4.67] | 36.26 [19.69] | 2.49 [2.22] |

Table 9.2: Summary of the robust ANOVAs' results ($\alpha < .05$). Test statistic $Q$ and effect size $\delta_t$ are reported. Adapted from [5], © 2025 IEEE.

| Effect type | Factor | Q | p | Sig. | $\delta_t$ | Effect | Fig. |
|---|---|---|---|---|---|---|---|
| **TCT** | | | | | | | |
| Main | *Sensor Placement* | 6.32 | <0.001 | * | -0.054 | Small | 9.4a |
| | *Control Mode* | 65.92 | <0.001 | * | -0.786 | Medium | 9.5a |
| Interaction | *Placement × Mode* | 0.47 | 0.699 | | - | - | A.2 |
| **Translation Error** | | | | | | | |
| Main | *Sensor Placement* | 5.30 | 0.001 | * | -0.392 | Small | 9.4b |
| | *Control Mode* | 0.88 | 0.347 | | - | - | |
| Interaction | *Placement × Mode* | 0.59 | 0.618 | | - | - | A.2 |
| **Rotation Error** | | | | | | | |
| Main | *Sensor Placement* | 4.12 | 0.006 | * | -0.210 | Small | 9.4c |
| | *Control Mode* | 6.83 | 0.009 | * | -0.180 | Small | 9.5b |
| Interaction | *Placement × Mode* | 1.06 | 0.363 | | - | - | A.2 |
| **NASA TLX** | | | | | | | |
| Main | *Sensor Placement* | 3.20 | 0.023 | * | -0.106 | Small | 9.4d |
| | *Control Mode* | 29.91 | <0.001 | * | -0.416 | Small | 9.5c |
| Interaction | *Placement × Mode* | 1.49 | 0.216 | | - | - | A.2 |
| **Borg CR10** | | | | | | | |
| Main | *Sensor Placement* | 3.71 | 0.011 | * | -0.112 | Small | 9.4e |
| | *Control Mode* | 9.47 | 0.002 | * | -0.233 | Small | 9.5d |
| Interaction | *Placement × Mode* | 1.98 | 0.116 | | - | - | A.2 |

### 9.4.3 *Qualitative Results*

For the qualitative feedback analysis, participant statements from the semi-structured interviews were paraphrased and grouped. The clustering process emphasized statements shared by at least two participants. Out of 347 individual statements collected, 244 were mentioned by at least two participants. These were then condensed into 54 summarizing statements (see Table 9.2).

### 9.4.4 *Interpretation of Results*

The analysis revealed no interaction effects between the two independent variables. This indicates that control mode and sensor placement do not influence each other. Thus, the effects of control mode and sensor placement can be interpreted separately.

SENSOR PLACEMENT Statistically significant main effects across all dependent variables between sensor placements were found (see Figure 9.4). Therefore, the null hypotheses $H_0^2.1$-$H_0^2.5$ were rejected in favor of their respective alternative hypotheses $H_1^2.1$-$H_1^2.5$.

Sensor placement at the robot resulted in significantly lower translation errors compared to leg and head placements (see Figure 9.4b), likely due

Table 9.3: Summary and frequency of statements (#) received during the semi-structured interview. Adapted from [5], © 2025 IEEE.

| **Sensor Placement** | |
|---|---|
| *Desk* | *Head* |
| + Comfortable (5), intuitive (4), natural (2) | + FOV obvious (5) and feels unrestricted (2) |
| + Large FOV (3) | + Intuitive (4), familiar (2), natural (2) |
| - Static FOV lacks flexibility (5) | - Tiring (9) |
| - Outstretched arm tiring (2) | - Hand in line of sight distracting (7) |
| - Issues with hand tracking (2) | - Sensor weight is uncomfortable (2) |
| *Leg* | *Robot* |
| + Effortless (9), intuitive (2) | + Intuitive (13), easy (2), natural (2) |
| + Hand not in line of sight (2) | + Flexible FOV (5) |
| - Rotation (8) and hand visibility issues (7) | - (In rate control:) hand leaves sensor FOV due to robot movement (7) |
| - FOV small (5) and unclear (2) | - Uncomfortable hand position (7) |
| - Requires training (5), imprecise (2), unnatural (3), unclear mapping (2) | - Unclear FOV (3) |
| - Issues with gesture recognition (3), maintaining distance to sensor (2) | - Hand tracking issues (3) |
| - Registration off due to leg movement (2) | - Rotation difficult (2) |
| **Control Mode** | |
| *Position Control* | *Rate Control* |
| + Intuitive (14), direct (6), easy (6) | + Effective for small spaces (9) and large movements (5) |
| + Precise (6), fast (5) | + Precise (9), more comfortable (2) |
| + Familiar (2), natural (2) | - Difficult to return to start position (5) |
| - Frequent need to re-grab (2) | - Overshooting desired position (4) |
| | - Not intuitive (2), tiring (2) |

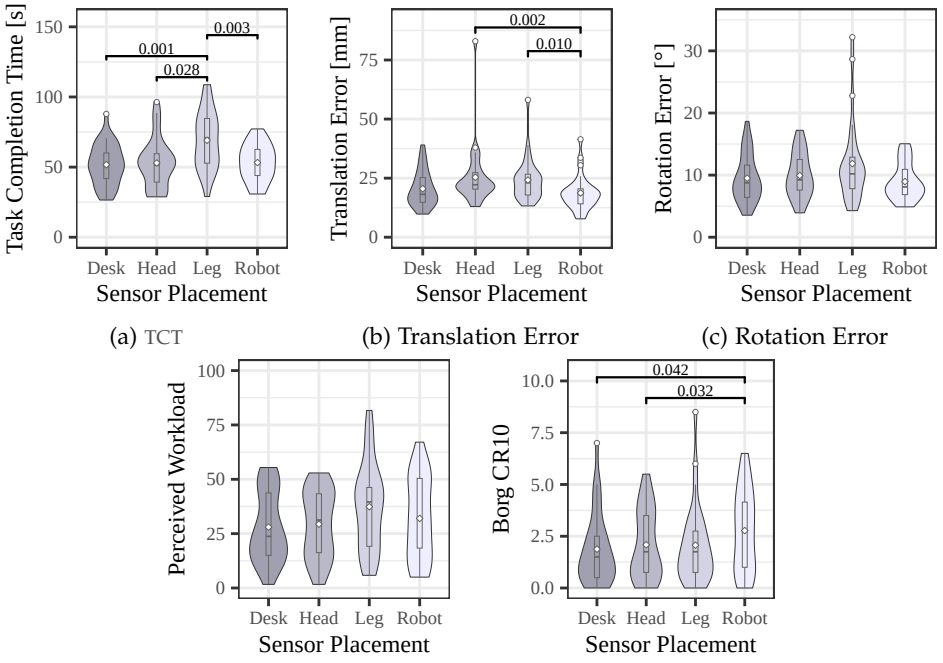
to the reportedly intuitive interaction, where participants felt as if they were directly grasping the robot, making the interaction natural and easy. However, this method also led to significantly higher Borg CR10 scores than head and desk placements (see Figure 9.4e), possibly due to uncomfortable hand positions requiring arm extension to stay within the sensor's FOV. As described in Table 9.3, some participants found the sensor's flexible FOV advantageous, while others faced difficulties with gesture recognition, especially during rotation, as their hands occasionally moved out of view.

Desk and robot placements showed no significant difference in accuracy, but desk placement significantly reduced perceived exertion. This suggests desk placement could be a viable option for long-term interactions where exertion is a priority, provided the workspace allows it. However, robot-mounted sensors remain preferable for short-term, precise tasks or when desk placement is impractical.

Sensor placement at the leg resulted in significantly higher TCT compared to the other placements (see Figure 9.4a). Participants reported that while the interaction felt less intuitive and harder to learn due to reduced hand-eye coordination, the relaxed arm position was effortless. However, this described effortlessness did not reduce Borg CR10 scores or perceived physical demand in the NASA TLX. This may be due to gesture recognition errors when hands

were too close to the sensor, forcing participants to hold their hands in an unnatural position, potentially increasing physical workload.

While significant main effects of sensor placement were found for both TLX scores and rotation errors, post hoc pairwise comparisons did not reveal significant differences between specific placements. This suggests that while there is a general difference between placements, their exact nature remains unclear. Larger sample sizes may clarify this. Descriptively, the leg placement trended towards higher rotation error and TLX scores compared to other placements. This aligns with participant-reported challenges, particularly difficulties in understanding the mapping of hand movements to robot rotations, which potentially increased rotation errors and contributed to a higher perceived workload.




(a) TCT (b) Translation Error (c) Rotation Error

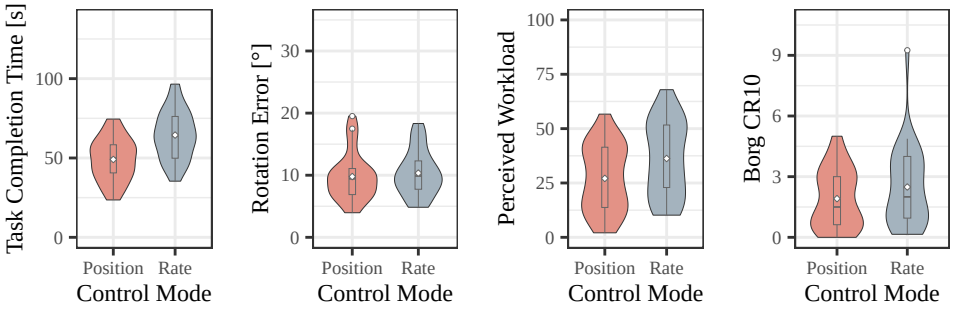
(d) NASA TLX (e) Borg CR10

Figure 9.4: Significant *sensor placement* main effects. Means are indicated by ◇. Significant post hoc pairwise comparisons are indicated with brackets. Adapted from [5], © 2025 IEEE.

ROBOT CONTROL METHOD Statistically significant effects of control mode on TCT, rotation error, NASA TLX scores, and Borg CR10 scores were found, with significantly longer task duration, higher rotation error, greater perceived workload, and higher perceived exertion for rate control compared to position control (see Figure 9.5). Thus, the null hypotheses $H_0^1.1$, $H_0^1.3$, $H_0^1.4$, and $H_0^1.5$ were rejected in favor of their respective alternative hypotheses. These results align with feedback from participant interviews, where position control was described as intuitive, easy, and accurate. Although some participants found rate control to be precise, they also reported overshooting the target and struggling to return to the neutral position, which may have reduced rotational accuracy and increased task duration. Holding their hands in

static, uncomfortable positions may have further contributed to the increased physical exertion, as reflected in the Borg CR10 scores.

No significant effect of control mode on translation accuracy was found, so the null hypothesis $H_0^1.2$ could not be rejected. This is consistent with participant feedback, as both control modes were described as precise. However, as noted above, despite the reported sense of precision, rotational error was significantly higher for rate control compared to position control. This suggests that difficulties with rate control, such as returning to the neutral position or overshooting, may have affected rotation more than translation. One explanation could be that translation is a more familiar interaction, making it easier to correct errors.




(a) TCT (b) Rotation Error (c) NASA TLX (d) Borg CR10

Figure 9.5: Significant *control mode* main effects. Means are indicated by ◇. Adapted from [5], © 2025 IEEE.

## 9.5 DISCUSSION

In this section, the results are discussed in the context of related work, limitations are outlined, and design implications based on the study's findings are presented.

SENSOR PLACEMENT The findings suggest that placing the sensor on the robot significantly improves translation accuracy, though it comes at the cost of higher perceived exertion. In contrast, alternative placements that offset the hand from the interaction site reduce fatigue, consistent with XR interaction research [152]. While previous XR research shows that interacting farther from the body reduces comfort but maintains rotational accuracy [289], this chapter expands on these insights by demonstrating that robot-mounted sensors improve translation precision and seem to have a positive effect on rotation accuracy, likely due to the reported intuitiveness. The results highlight challenges with leg placement, which increased TCT and perceived workload. This aligns with previous findings on 2D XR input [49], where leg-mounted sensors led to higher mental load due to difficulties in understanding the interaction mapping. Participants reported challenges related to the hand being out of sight. Providing feedback on the state of interaction could mitigate this issue and should be explored in future work.

CONTROL MODES The study results indicate position control being more efficient than rate control for gesture-based interaction, consistent with previous findings for haptic input devices [169, 336] and head gestures [275]. However, prior research also suggests that rate control may be advantageous for slow systems and tasks involving large workspaces [169], an argument supported by participants' feedback. While this study chose a precision task requiring only small translations and rotations, because of its association with higher fatigue, a broader range of tasks may have provided more comprehensive insights.

### 9.5.1 *Limitations*

The main limitation of the study is the specificity of the task used. Additionally, to the aforementioned limitation regarding the workspace size, the task allowed participants to remain stationary, whereas tasks requiring more flexibility in operator movement might have highlighted the advantages of body-worn or robot-mounted sensor placements. Therefore, future work should consider using a wider range of tasks, including those requiring greater motion or different viewpoints.

This study aimed to assess the impact of different sensor placements and control modes on user fatigue. However, physical exertion was measured only during short interaction tasks with a duration of approximately one minute. This design choice was motivated by the intended use case of gesture-based interaction as a complement to autonomous robot control, specifically for the manual execution of isolated subtasks rather than full procedural interventions. As a result, the perceived exertion levels may have been lower than in prolonged interaction scenarios, potentially underestimating the extent of fatigue effects. Future studies could extend task duration to better capture cumulative fatigue and its influence on user performance and comfort.

Another limitation of this study relates to the demographics of the participants. The gender distribution was not uniform due to the random sampling method. Additionally, the study was designed for right-handed use, and as a result, only right-handed participants were invited to participate. This limitation may affect the generalizability of the results. Future studies should aim to include a more diverse sample of participants, both in terms of gender and handedness, to increase the applicability of the results.

### 9.5.2 *Implications*

In summary, based on the study results, the following design implications for hand gesture-based HRI are proposed:

**Utilize robot-mounted sensors for short-term precision.** Findings show that placing the sensor on the robot results in intuitive and accurate control. However, this sensor placement increases physical demand and is therefore best suited for shorter interactions.

**Place sensor in the environment for extended interaction.** The results indicate that the sensor placement on the desk offers a balanced approach. It provides good accuracy while minimizing physical strain, making it ideal for prolonged use.

**Employ position control for efficient interaction.** Study results suggest that position control provides faster, more precise, and more intuitive control than rate control; however, rate control may be a suitable alternative for tasks that require larger movements within a limited interaction space.

## 9.6 CONCLUSION

This chapter explored different sensor placements and control methods for hand gesture-based HRI, with the goal of identifying the most effective configurations for interaction accuracy, efficiency, and user comfort. The findings indicate that among the sensor placements, robot-mounted sensors offered precise and intuitive control but were physically demanding. In contrast, desk-mounted sensors provided a balanced trade-off between accuracy and exertion. Leg-mounted sensors, though allowing for a relaxed posture, faced challenges with hand-eye coordination, leading to reduced efficiency. Additionally, position control was found to be more intuitive and precise than rate control, while reducing perceived workload and fatigue. Based on these results, implications were proposed to optimize hand gesture-based HRI.

The second part of this thesis aimed to address specific challenges in hand gesture-based HRI to enhance the overall efficiency of this interaction modality. Based on related work, two major limitations were selected as the focus: the lack of haptic feedback and the issue of user fatigue during interaction.

To address the challenge of missing haptic feedback, two force assistance strategies were investigated. The first experiment examined the effects of force automation. Results showed that automating the contact force can compensate for the absence of haptic feedback in gesture-based robot control. In fact, for general tasks, it enabled hand gesture control to perform comparably to the established hand-guiding method, and even to outperform it in more precise tasks. A second experiment explored sensory substitution through vibrotactile and visual feedback. By also examining feedback design and placement, this study demonstrated that visual feedback, implemented as a color and shape-changing bar in a world-stabilized placement, is effective in conveying force information.

The second identified challenge, user fatigue during interaction, was addressed by exploring the influence of different sensor placements and control modes. Findings from the user study indicated that while robot-mounted sensors improved precision, desk-mounted sensors significantly reduced perceived exertion. Additionally, position control was found to be both more efficient and physically less demanding than rate control.

As the first experiment demonstrated, hand gesture-based robot control has the potential to be comparable or even surpass state-of-the-art interaction methods when its core limitations are mitigated. Given the flexibility and intuitiveness of this input method, improving hand gesture interaction represents a promising path towards enhancing overall HRI efficiency. Therefore, this part of the thesis represents a meaningful step towards more efficient HRI by providing validated strategies to reduce both the impact of missing haptic feedback and the fatigue associated with hand gesture-based interaction.

# Part III

# XR PROTOTYPING FOR HRI EVALUATIONS

**Synopsis:** This part investigates the validity of XRP for simulating HRI user study setups. To this end, two experiments are presented. The first experiment examines the transferability of user study results across different degrees of virtualization for different tasks. The second experiment further explores the transferability of findings from a study on force assistance techniques, conducted in both a real and an XR-based user study setup.

This part contains material adapted from the following publications:

Tonia Mielke, Mareen Allgaier, Danny Schott, Christian Hansen, and Florian Heinrich. "Virtual Studies, Real Results? Assessing the Impact of Virtualization on Human-Robot Interaction." In: *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, Apr. 2025, pp. 1–8. DOI: 10.1145/3706599.3719724 [2].

Tonia Mielke, Mareen Allgaier, Christian Hansen, and Florian Heinrich. "Extended Reality Check: Evaluating XR Prototyping for Human-Robot Interaction in Contact-Intensive Tasks." In: *Transactions on Visualization and Computer Graphics (TVCG)* 31.11 (Oct. 2025), pp. 10035–10044. DOI: 10.1109/TVCG.2025.3616753 [1].

# 11 INTRODUCTION

## 11.1 MOTIVATION

The previous chapters of this thesis employed user studies as a tool for human-centered HRI research. This process of prototyping and evaluating HRI systems is essential to enable more intuitive and efficient user interfaces [270]. However, setting up real-world user studies is often costly [223, 248], with robot locations not always being accessible [67, 252], and technical issues or user misconduct potentially causing physical harm [186]. To address these challenges, XR has emerged as a valuable tool for prototyping HRI interfaces by simulating real-world conditions. In particular, XRP supports early-stage development by reducing experimental costs [223, 249, 338], minimizing risks during physical interaction [63, 244], improving workspace accessibility [252, 316], and accelerating research and development [120]. Using a virtual robot and task also enables the simulation of factors such as object recognition and task environments while offering a robust and consistent experimental setting [349]. Instead of virtualizing the entire system, some approaches incorporate hardware-in-the-loop simulation, embedding a real robot within a virtual environment. This hybrid method allows for fast prototyping, reduced collision risk [63], and more accurate performance assessments [61].

Despite these benefits, the transferability of study results obtained through XRP to real-world scenarios remains a concern. In other areas, such as public displays [206], authentication systems [215], and assembly tasks [254, 337], virtual replicas of real environments have led to similar behavior and objective measures, highlighting the potential of XR as a test bed for human-centered evaluation of real-world scenarios. However, other studies report perceptual differences across varying levels of virtualization [330], which highlights the need to validate the transferability of findings obtained through XRP.

While prior work has investigated perceptual similarities between virtual and physical robots [68, 127, 339] and introduced novel interaction concepts within XR [227, 343], limited attention has been given to the validity of XR-based user studies in HRI, especially for direct robot control.

To address this gap, this part of the thesis aims to answer RQ3:

> **RQ3** | Can XRP produce transferable results for HRI research?

## 11.2 RELATED WORK

According to Plümer and Tatzgern [251], validation of XRP as a performance evaluation tool can follow two main approaches. The *bottom-up approach* isolates specific influencing factors such as display fidelity [48], latency [180], or visual realism [291]. In contrast, the *top-down approach* involves conducting

full user studies in XR and comparing them with their real-world counterparts [204, 215, 254, 330, 337].

In robotic research, the top-down approach has been the more prevalent method. Early top-down studies compared physical and video representations of robots. These works showed that physically present robots often led to more positive interactions than simulated representations [26, 164, 189], with participants displaying higher engagement levels towards real robots [299]. However, Liu et al. [198] found that XR simulations led to better performance compared to traditional on-screen simulations, suggesting potential advantages of XR. To further explore the suitability of XR for simulation in HRI, previous work has examined the perception of simulated robots and the use of XR environments for interactive robot control (see Table 11.1).

Table 11.1: Overview of related work on XRP in HRI. $\approx$ indicates similar findings across prototype environments, and $\neq$ indicates different findings.

| Interaction | | Task | Environment | Findings |
|---|---|---|---|---|
| Perception | [127] | Robot gestures | AR/Real robot + AR/Real task | $\approx$ Perception |
| | [278] | Robot gestures | VR vs. Real robot | $\approx$ Perception |
| Proexemics | [191] | Approaching | VR vs. Real robot | $\neq$ Perception |
| | [151] | Co-presence | VR vs. Real robot | $\approx$ Perception |
| | [339] | Co-presence | VR vs. Real robot | $\neq$ Physical measures |
| Collaboration | [146] | Handover | VR vs. Real robot | $\approx$ Performance |
| | [68] | Handover | AR vs. Real robot | $\neq$ Safety zone preference |
| Social behavior | [347] | Secret keeping | VR vs. Real robot | $\neq$ Perception |
| Robot Control | [227] | Pick-and-place | VR vs. Real robot | $\neq$ Performance[1] |
| | [344] | Manipulation | VR vs. Real robot | $\approx$ Performance[1] |
| | [251] | Drone steering | AR vs. Real robot | $\neq$ Performance, workload |
| | [252] | Steering | AR vs. Real robot | $\neq$ Performance, workload |
| | [63] | Screw removal | VR vs. Real robot + AR task | $\neq$ Performance |
| | [343] | Stacking | VR vs. Real robot | $\neq$ Usability |

[1] Different interaction techniques used in virtual and real conditions.

### 11.2.1 *XRP for Robot Perception*

Han et al. [127] evaluated the comprehensibility of robotic deictic gestures across real, virtual, and hybrid conditions and found no significant differences. Sadka et al. [278] similarly observed consistent interpretations of gestures between real and virtual robots.

Another line of work focuses on the perception of robot proxemics. Li et al. [191] reported that users preferred closer distances with real robots, while virtual robots caused more discomfort. In contrast, Inoue et al. [151] and Weistroffer et al. [339] found similar perceptions and co-presence ratings between physical and virtual robots, although physiological responses like heart rate and skin conductance were higher with real robots. Hsieh and Lu [146] reported no significant performance differences in handover tasks

between real and virtual robots. Cogurcu et al. [68] found that visualization preferences for robotic safety zones varied depending on whether the robot was real or virtual. Wijnen et al. [347] replicated a secret-keeping study in XR and found discrepancies in outcomes, indicating challenges in using XR for some types of HRI studies.

While these works offer insights into robot perception and proxemics, they largely focus on passive observation or perception rather than interactive control of robots. Consequently, they provide limited evidence on how HRI is influenced by the virtualization of the robot or its environment.

### 11.2.2 *XRP for Robot Control*

To explore interactive HRI across real and XR settings, Nenna et al. [227] compared pick-and-place tasks in XR and real environments, finding shorter operation times and lower workload in XR, especially for high-precision tasks. Similarly, Whitney et al. [344] compared different manipulation tasks performed with a physical robot and an XR-simulated robot. They found that physical robots enabled faster task completion, while virtual robots made complex positioning tasks easier. However, both studies employed different control interfaces for the XR and physical settings, which limited the comparability and transferability of interaction methods.

Plümer and Tatzgern [251] investigated XR validity in drone steering tasks using video see-through displays. They identified multiple confounding factors, such as altered depth perception and perceived risk, that affect the absolute validity of XRP. In their follow-up work [252], they assessed visualization techniques for reducing perceived latency. They found that while absolute performance differed between real and simulated environments, the relative ranking of techniques remained consistent, supporting the notion of relative validity. Whitney et al. [343] compared robot control using a physical robot, viewed on a screen, with control of a virtual robot perceived in VR during a stacking task. Their results indicated similar task duration and workload across conditions. However, the XR setup resulted in higher perceived usability. Chen et al. [63] explored not only robot virtualization but also task virtualization, comparing a real robot with a virtual task against an entirely virtual setup. They found that participants perceived the partly simulated condition as more realistic and suitable for producing reliable results. However, task completion was faster in the entirely virtual simulation, suggesting greater caution when interacting with real hardware.

## 11.3 RESEARCH GAP

As summarized in Table 11.1, previous research has investigated the perceptual similarity between virtual and real robot scenarios [127, 339] and introduced novel interaction concepts in XR [227, 343]. However, limited attention has been given to assessing the validity of XR user studies in interactive scenarios, particularly for direct robot control. In the context of HRI, efforts

have been made to compare entirely physical setups with XR-based setups for tasks such as robot steering [251, 252] and pick-and-place operations [68, 343]. Other studies have investigated real robots with simulated tasks in XR, for example, in screw removal tasks [63].

However, these works each focus on a single task, which limits the understanding of which factors influence the transferability of XR study results. Furthermore, a systematic comparison across different levels of virtualization is still missing. This includes setups that use only simulated tasks and those that combine simulated tasks with simulated robots. Addressing this gap is important as perceptual differences between real and simulated robots [68, 339] might affect various task types differently. These effects might be reduced by keeping the real robot in the loop while virtualizing only the task.

Additionally, previous studies primarily focus on tasks involving little to no physical contact between the robot and the environment. In contrast, there is a lack of research on contact-intensive tasks, where the robotic end effector directly interacts with the environment. Understanding how XR translates to such tasks is particularly important, as previous studies have shown that stiffness and force perception can differ significantly in XR [115, 355].

## 11.4 CONTRIBUTION

This part of the thesis investigates the transferability of XRP for HRI user studies, providing the following contributions.

**Investigation of Task-Specific XRP Transferability.** Chapter 12 explores the transferability of XRP study results across three distinct tasks that vary in required precision and complexity.

**Investigation of Different Levels of Virtualization.** To examine how the degree of virtualization affects study outcomes, Chapter 12 compares physical test setups with XRP environments that include a virtual task with a real robot and an entirely simulated setup.

**Investigation of XRP for Contact-Intensive Tasks.** To further explore the suitability of XRP for simulating contact-intensive tasks, Chapter 13 presents a more focused investigation into the transferability of XRP user studies in such scenarios by evaluating different force assistance approaches.

Since the investigation of contact-intensive tasks also includes the comparative evaluation of force assistance strategies, this part additionally contributes to understanding force assistance concepts through the following:

**Comparative Evaluation of Force Assistance Approaches.** While Chapter 8 introduced two force assistance concepts, namely partial automation and sensory substitution, these have not yet been evaluated comparatively. As a result, their relative performance and synergistic effects remain unclear. Chapter 13 addresses this gap by jointly investigating partial automation as presented in Section 8.2 and sensory substitution based on the findings from Section 8.3.

# 12 INVESTIGATION ON XRP VIRTUALIZATION LEVELS

As previously described, prior work has indicated that perceptual differences between real and simulated robots might arise [68, 339]. While some studies have examined whether such differences affect the transferability of XRP user study results to real-world scenarios, a broader understanding of how varying levels of virtualization influence user perception and behavior is still needed. To date, comparisons have primarily focused on fully physical setups with physical robots performing simulated tasks [63], or entirely simulated environments [251, 252]. However, a gap remains in evaluating the transferability across different levels of virtualization.

Addressing this gap is important, as perceptual differences might be reduced when the real robot is kept in the loop while only the task is virtualized. In addition, previous work has focused on specific use cases without analyzing how task characteristics influence transferability. Therefore, this section investigates the transferability of results between three conditions: virtual robots, real robots in virtual tasks, and entirely physical setups, using three practical tasks.

Parts of this section were previously published in Tonia Mielke, Mareen Allgaier, Danny Schott, Christian Hansen, and Florian Heinrich. "Virtual Studies, Real Results? Assessing the Impact of Virtualization on Human-Robot Interaction." In: *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, Apr. 2025, pp. 1–8. DOI: 10.1145/3706599.3719724 [2].

## 12.1 TECHNICAL METHODS

This section assesses the transferability of user study results in HRI across different levels of virtualization. Therefore, three tasks were implemented, each evaluated under three levels of virtualization: a real robot with a real task (*real*), a real robot with a virtual task (*mixed*), and a virtual robot with a virtual task (*virtual*).

*Supplementary video*

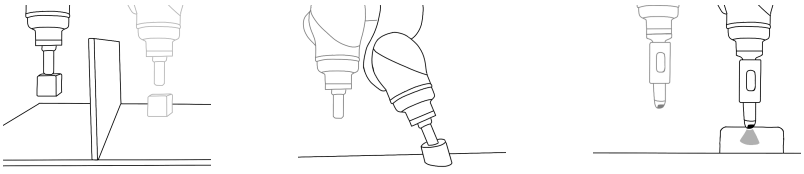
### 12.1.1 *Task Design*

To assess varying ranges of motion, precision, and task complexity, the two most common HRI task domains were explored: pick-and-place and assembly, i.e., peg-in-hole [272]. In addition, robotic ultrasound was investigated as a higher-level cognitive task [133]. While the first two tasks were selected for their general relevance and widespread use, robotic ultrasound was chosen as a specific, practical example, representing a contact-intensive task. An overview of the tasks is illustrated in Figure 12.1.

PICK-AND-PLACE In this task, a cube is picked up from a start location and placed in a target location. To simplify interaction, a magnetic setup was used instead of a gripper, with magnets on the flange and object for pickup and additional magnets underneath for release onto a magnetic surface. Pick-and-Place was chosen as a task to require broad robotic arm movements, further encouraged by obstacles in the workspace.

PEG-IN-HOLE Here, a cylindrical peg attached to the robot must be inserted into different holes. To require both translational and rotational movements, the holes are angled in different directions. This assembly task was chosen for its demand for precise alignment and rotational movement.

ULTRASOUND In robotic ultrasound, an ultrasound probe is attached to the robot, enabling imaging through robot control. The task involves imaging a phantom with spheres of varying sizes and assessing their dimensions. Robotic ultrasound was chosen as it is an emerging field requiring HRI to complement autonomous control [292, 359] and to contrast with the two other industrial tasks. Additionally, it requires users to focus on both the robot and the sonographic images, adding cognitive complexity, in contrast to the other tasks that mainly address motor skills.



(a) Pick-and-Place (b) Peg-in-Hole (c) Ultrasound

Figure 12.1: Schematic overview of tasks investigated.

### 12.1.2 *Real Environment*

The real environment (see top row Figure 12.2) was implemented as described in Section 2.3. Task components, such as the cubes for the pick-and-place task, the holes for the peg-in-hole task, and the corresponding mounts on the robot, were 3D printed. For the ultrasound task, a *Clarius* probe was used, and the phantom and spheres, made of agar-agar with varying concentrations, were custom-built to exhibit different imaging properties (see Section 2.3.5).

### 12.1.3 *Simulated Environment*

Since the *mixed* condition requires perception of the real robot, AR was used to visualize the virtual components within the simulated environment. To minimize confounding factors caused by mediated views through video see-through HMDs [251], an optical see-through HMD enabling direct view of the real environment, the *HoloLens 2*, was employed. The simulated environment, including 3D models of the robot and tasks, was also developed in Unity.

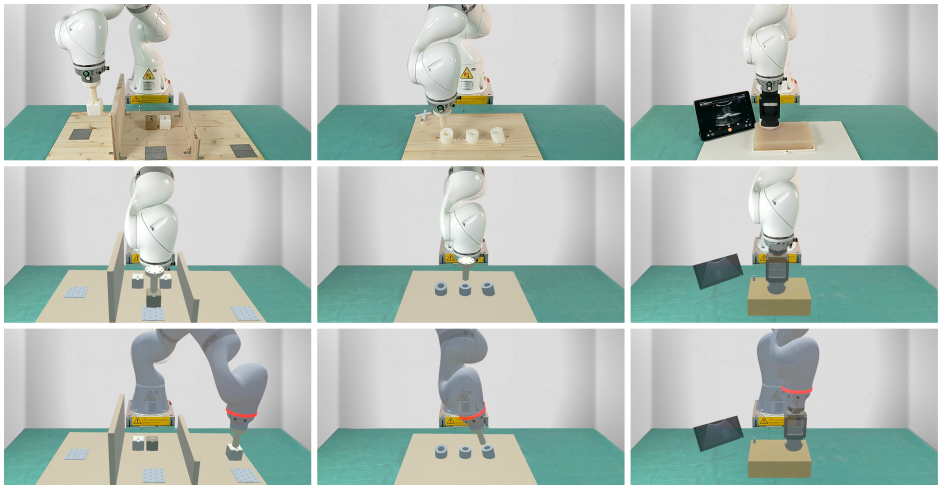


Figure 12.2: Overview of virtualization levels (rows: real, mixed, virtual) and tasks (columns: pick-and-place, peg-in-hole, robotic ultrasound). Reprinted from [2].

Robot simulation was realized with *ROS* (Open Source Robotics Foundation, USA) using the *Unity Robotics Hub*. To replicate the movement sounds of the physical robot, a humming sound, changing pitch and volume depending on the movement velocity, was included. The sound was played by a speaker positioned underneath the robot for accurate spatial sound.

Since a standalone HMD was used for the simulated environment, performance was the primary limiting factor for the ultrasound simulation. Consequently, a fast texture-based approach similar to that of Allgaier et al. [18] was chosen. The simulation used 3D surface meshes as input, which were converted into a 3D render texture using ray casting. To enable this, a CAD model of the real phantom was created, consisting of three material types: the base material, the embedded spheres, and the bracket structure holding the spheres. Each material was assigned an echogenicity value to mimic its appearance in real ultrasound images. The render textures were then used for real-time ultrasound simulation. Custom shaders, based on the method by Allgaier et al. [18], were applied to generate ultrasound-like cross-sections and simulate typical ultrasound characteristics such as attenuation, reflection, blur, and noise. These parameters were empirically tuned to closely resemble real-world imaging. Registration between the virtual and physical environments was performed as described in Section 2.3.4.

### 12.1.4 *Implementation*

The robot control was implemented as described in Section 2.3, using a *Leap Motion Controller 2* positioned in front of the user. A pinch gesture activated control, causing the robot's movements to follow the user's hand while the gesture was maintained. Translational motion was mapped at a 1:1 ratio, while rotational motion was scaled down at a 2:1 ratio. A foot pedal was used to toggle between translation, rotation, or simultaneous translation and

rotation modes. The robot control interface was developed in *Unity* and ran on a desktop PC, sending movement commands via UDP to the physical robot controller or via *ROS TCP* to the simulated robot controller. Inverse kinematics for the commanded robot movements were calculated using the built-in functions of *KUKA Sunrise* for the physical robot and the *Orocos KDL* plugin in *MoveIt* for the simulated robot.

## 12.2 EVALUATION METHODS

To assess the transferability of XRP for user studies, a within-subjects study was conducted. Each participant completed all three tasks in each of the three prototype environments, resulting in nine trials per person (3 prototype environments $\times$ 3 tasks). Additionally, a preliminary study was conducted to evaluate the simulation fidelity and identify areas for improvement.

### 12.2.1 *Preliminary Study for Improvement of Fidelity*

The preliminary study involved five participants who experienced the different prototype environments and tasks, providing qualitative feedback on the simulation fidelity. While participants generally found the XR and real versions to be similar, some specific necessary improvements were identified. Additional sounds, such as the clicking of magnets during pick-and-place and scraping sounds for robot movements on surfaces, were incorporated. Moreover, static lighting conditions were adjusted to better mimic real-world lighting. While the benefits of dynamic shadows were identified, they were not implemented due to computational constraints on the *HoloLens*.

INDEPENDENT VARIABLES The independent variable of the one-factorial investigation was the degree of virtualization (*real*, *mixed*, *virtual*). Each of the three tasks (Pick-and-Place, Peg-in-Hole, Ultrasound) was performed under each of the virtualization levels.

DEPENDENT VARIABLES Commonly used objective and subjective measures for product performance assessment [251] were utilized as dependent variables. The first variable was TCT, measured from the start of the user's interaction to the robot's final movement before task completion was confirmed. A foot pedal press confirmed task completion for all tasks. Task completion for the pick-and-place task involved placing three cubes onto designated metal plates. In the peg-in-hole task, participants had to insert a peg into three holes, with a sound indicating successful insertion. For the ultrasound task, completion was defined by confirming the order of three spheres, ranging from 25 mm to 30 mm in size, by pressing a corresponding pedal when each sphere was visible. To ensure consistent task completion across virtualization levels, such as triggering the sound during peg-in-hole at the same insertion depth or confirming sphere visibility in the ultrasound task, the physical tasks were optically tracked using a *FusionTrack 500* tracking

camera. As a subjective measure, perceived workload was assessed using the raw NASA TLX questionnaire, while subjective task difficulty was measured with the SEQ. Participants were further asked to rate the similarity between virtualization stages for each task using a five-point Likert item (1=very similar; 5=very different).

### 12.2.2 *Hypotheses*

The study was exploratory, with no prior assumptions made regarding the expected outcomes. Consequently, two-sided hypotheses were examined. The null and alternative hypotheses considered were as follows:

$$
\begin{array}{ll}
H_0.x & \text{The mean } x \text{ is equal for all virtualization levels.} \\
H_1.x & \text{The mean } x \text{ differs for at least one virtualization level.}
\end{array}
\qquad x \in \begin{cases} \text{1-TRE} \\ \text{2-TCT} \\ \text{3-TLX} \end{cases}
$$

### 12.2.3 *Sample Design*

As this study investigated two general HRI tasks and included a training session to familiarize participants with the ultrasound task, no task-specific prior knowledge was required for participation. Therefore, participants with a technical background were recruited from the local university via online polls. Each participant received compensation of 15 €.

### 12.2.4 *Procedure*

At the beginning of the study, participants received a brief introduction on the purpose of the study and completed a written consent form. They were then introduced to the robot control through a short training session. The first task block followed, starting with an explanation of the task and its goal. The order of tasks was counterbalanced using a Latin square. In each task block, the three levels of virtualization were presented in a counterbalanced order. Each combination of task and virtualization level was performed twice: once for training and once for evaluation. Training sessions were included to familiarize participants with the task. After each evaluation trial, participants completed the NASA TLX and SEQ. Upon completing each task block, they answered the similarity question and participated in a short semi-structured interview. After all three tasks, a final semi-structured interview was conducted to gather general feedback. Study completion took an average of 60 minutes per participant.

### 12.2.5 *Statistical Analysis*

Several statistical analyses were conducted using R (version 4.4.0) to assess the respective null and alternative hypotheses of the virtualization levels for each dependent measure. The data for the three tasks were analyzed separately.

First, Shapiro-Wilk and Levene's tests were conducted to assess normality and homogeneity assumptions on TCT measures. These assumptions were violated. Hence, robust repeated measures ANOVAs for within-subjects designs based on trimmed means were calculated (see [348]) for TCT data. In addition, robust repeated measures ANOVAs were also conducted on the TLX scale data and the ordinal SEQ data. Post hoc paired sample Yuen's tests based on trimmed means with Holm corrections for multiple testing were conducted as pairwise comparisons in case of significant ANOVA outcomes. For significant effects, effect sizes were calculated using the $\delta_t$ estimate proposed by Algina et al. [16]. For non-significant effects, Bayes factors were calculated using Bayes ANOVAs for repeated measures to further interpret statistical results by estimating how well the null hypotheses could be supported [142].

### 12.2.6 *Simulation Fidelity*

For identifying and accounting for confounding factors prior to the experiment, the XRP validation framework introduced by Plümer and Tatzgern [251] was followed to assess differences between the real and simulated versions.

*Visual Fidelity.* In the XR conditions, participants wore an optical see-through HMD, allowing them to directly perceive the RE, eliminating the need for a virtual representation of the user's hand or environment. To create accurate representations of the task and robot for the virtualized conditions, publicly available mesh files for the robot[1] were used, along with 3D models of the printed components and custom-made CAD models for other task parts.

*Haptic Fidelity.* Since the same touchless control method was used for the robot in all conditions, and no direct contact with the environment was required, the haptic fidelity remained identical across all conditions.

*Audio Fidelity.* To replicate the sounds of the physical robot during movement, a velocity-dependent humming sound was generated for the virtual robot. In the preliminary study, additional sounds were identified for the interaction between the virtual flange and task. These included the magnet clicking when picking up and placing the cube, as well as scratching noises when the flange collided with surfaces.

*Interaction Fidelity.* All conditions used the same interaction technique, hardware, and setup, ensuring consistent interaction fidelity across conditions.

*Functional Fidelity.* The physical and simulated robots were set to have the same maximum velocity and acceleration, as well as similar stiffness behavior. The magnets for the pick-and-place tasks were also tuned to match the real ones' reaching distance. The movement of the robot and its interaction with the environment were validated by simultaneously moving the virtual and real robot. For the ultrasound image, a rendering approach shown to produce realistic images [18] was used, with the visual properties of the imaged structures tuned to mimic the real components.

1 https://github.com/ros-industrial/kuka_experimental

*Data Fidelity.* Objective and subjective data were collected identically across all approaches, as the same automated data collection and questionnaires were used. To ensure consistent measures of task completion across the virtualization stages, the components in the physical setup were optically tracked, allowing the relative position of the task to the robot to be known.

*Simulation Overhead.* During the real conditions, participants did not have to wear the HMD to allow for a realistic assessment of the comparability between the real and virtual conditions, as a study with the real robot would not involve wearing an HMD.

## 12.3 RESULTS

This section outlines and interprets the quantitative and qualitative results obtained from the user study.

### 12.3.1 *Participants*

In total, 31 participants (13 female, 15 male, 3 diverse) took part in the user study, aged between 20 and 36 years (M=25.94, SD=3.60). Among them, 26 were right-handed, and 5 were left-handed. The group comprised 18 university students, 9 PhD students, 3 engineers, and 1 postdoctoral researcher. Participants reported an average familiarity with HRI of 2.77 (SD=1.23) and XR of 3.10 (SD=1.35) on a 5-point Likert item, where 5 indicates high familiarity. They self-reported an average technical affinity of 4.19 (SD=0.65) and motor skills of 3.81 (SD=0.54) on the same scale.

### 12.3.2 *Quantitative Results*

The descriptive results are summarized in Table 12.1, and the statistical results are presented in Table 12.2. Significant pairwise comparisons are illustrated in Figure 12.4. The descriptive results of the similarity scores are shown in Figure 12.3. These scores indicate task-dependent differences across virtualization levels. While the virtualization levels are generally perceived as similar for the pick-and-place and ultrasound tasks, noticeable differences are perceived between the real and virtual scenarios in the peg-in-hole task.

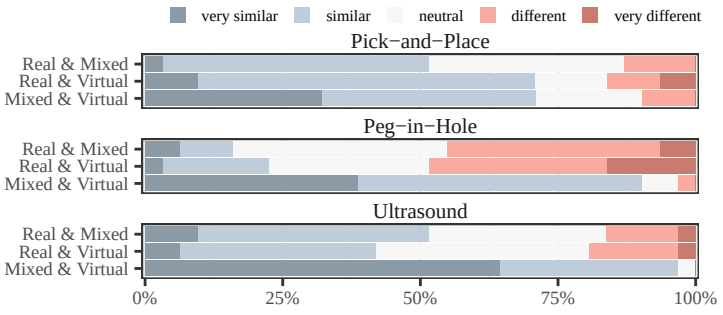



Figure 12.3: Plot of the similarity scores. Adapted from [2].

Table 12.1: Summary of descriptive results for all dependent variables (n=31). Entries are in the format: mean value [standard deviation].

| Variable | TCT (s) | | TLX | | SEQ | |
|---|---|---|---|---|---|---|
| **Pick-and-Place** | 84.50 | [29.72] | 24.53 | [14.61] | 5.68 | [0.99] |
| *Real* | 88.34 | [34.09] | 25.40 | [16.35] | 5.48 | [1.21] |
| *Mixed* | 83.22 | [27.22] | 25.99 | [15.61] | 5.61 | [0.92] |
| *Virtual* | 81.93 | [27.95] | 22.20 | [11.61] | 5.94 | [0.77] |
| **Peg-in-Hole** | 75.35 | [32.46] | 36.94 | [18.99] | 4.63 | [1.27] |
| *Real* | 54.43 | [23.02] | 26.99 | [16.31] | 5.48 | [0.96] |
| *Mixed* | 85.52 | [34.29] | 41.64 | [19.37] | 4.26 | [1.15] |
| *Virtual* | 86.12 | [29.10] | 42.18 | [17.66] | 4.16 | [1.24] |
| **Ultrasound** | 61.78 | [22.43] | 27.05 | [14.41] | 5.58 | [1.06] |
| *Real* | 65.05 | [25.41] | 27.55 | [14.72] | 5.55 | [0.96] |
| *Mixed* | 60.78 | [22.40] | 26.75 | [13.68] | 5.68 | [1.08] |
| *Virtual* | 59.50 | [19.44] | 26.85 | [15.25] | 5.52 | [1.15] |

Table 12.2: Summary of robust ANOVAs results ($\alpha < .05$), reporting the test statistic $F$ and Bayes factor $BF_{10}$ for non-significant effects, and effect size $\delta_t$ for statistically significant effects.

| Variable | F | p | Sig. | $BF_{10}$ | $\delta_t$ | Interpretation | Fig. |
|---|---|---|---|---|---|---|---|
| **Pick-and-Place** | | | | | | | |
| TCT | 1.09 | 0.348 | | 0.51 | - | Anecdotal | 12.4a |
| TLX | 1.12 | 0.332 | | 1.41 | - | Anecdotal | 12.4b |
| SEQ | 2.73 | 0.079 | | 1.56 | - | Anecdotal | 12.4c |
| **Peg-in-Hole** | | | | | | | |
| TCT | 25.36 | <0.001 | * | - | -1.02 | Large | 12.4a |
| TLX | 25.93 | <0.001 | * | - | -0.75 | Medium | 12.4b |
| SEQ | 20.94 | <0.001 | * | - | 1.03 | Large | 12.4c |
| **Ultrasound** | | | | | | | |
| TCT | 1.81 | 0.181 | | 0.34 | - | Anecdotal | 12.4a |
| TLX | 0.23 | 0.770 | | 0.11 | - | Moderate | 12.4b |
| SEQ | 0.04 | 0.951 | | 0.16 | - | Moderate | 12.4c |

### 12.3.3 *Qualitative Results*

During the semi-structured interviews conducted in the study, a total of 349 individual statements were recorded. After summarizing statements within participants, 203 statements were identified that were stated by at least two participants. These were then consolidated into the 34 statements presented in Table 12.3.

### 12.3.4 *Interpretation of Results*

This section interprets the quantitative results for the three different tasks and contextualizes them with the qualitative feedback obtained.

Table 12.3: Summary and frequency of statements (#) received during the semi-structured interview. Adapted from [2].

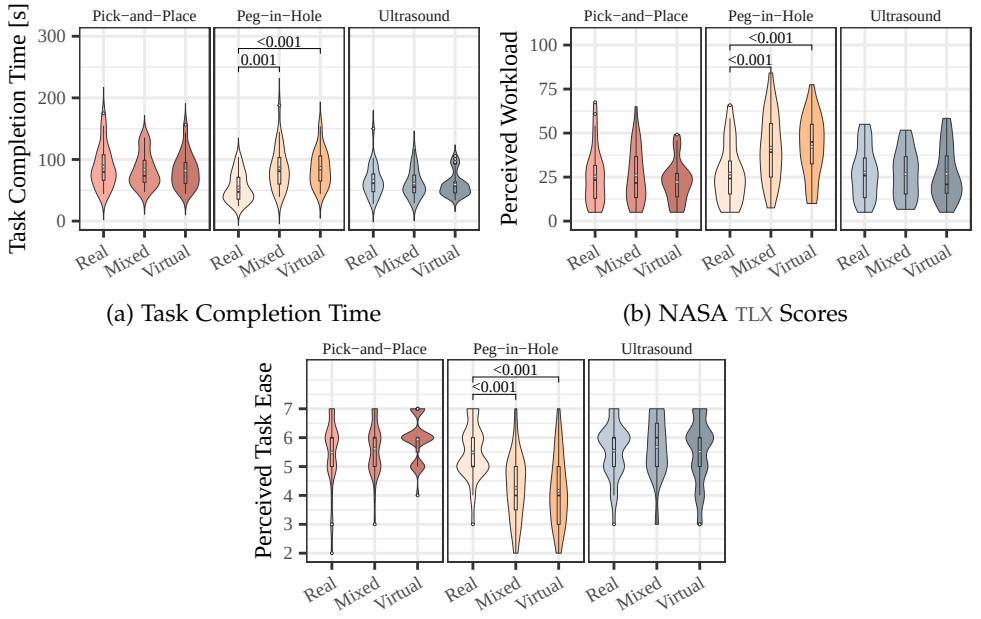
| General | Pick-and-Place |
|---|---|
| • Precision in virtualized difficult (4)<br>• Perceived dissonance in *mixed* (4)<br>• FOV of HoloLens too small (3)<br>• FOV of HoloLens sufficient (3)<br>• Sounds in *mixed* and *virtual* helpful (3)<br>• One has to get used to the HoloLens (3)<br>• Looking at *real* more comfortable (2) | • More scared about collisions (13) and careful in *real* (7)<br>• Virtualization levels very similar (12)<br>• Task felt easier in *virtual* and *mixed* (5)<br>• Depth perception challenging in *virtual* and *mixed* (2)<br>• Sounds were helpful (2) |
| **Peg-in-Hole** | **Ultrasound** |
| • Perception of depth (16), peg orientation (16), hole orientation (7), and distances (4) harder in virtualized settings, affecting precision (11)<br>• Task perceived easier in *real* (10)<br>• Perception *mixed* easier than *virtual* (4)<br>• *Mixed* and *virtual* felt similar (3)<br>• Virtualization levels (5), broad movements (2), and control (3) similar<br>• *Mixed* felt confusing (3)<br>• More fear of collisions in *real* (2) | • Virtualization levels similar (16)<br>• Contact force perception harder in *mixed* and *virtual* (9), attributed to phantom deformation missing (8)<br>• More afraid of applying excessive contact force in *real* (9)<br>• Task easier in *mixed* and *virtual* (4)<br>• FOV of HoloLens too small for simultaneous view of robot and US image (3)<br>• Task easier in *real* (3)<br>• Friction in virtualized realistic (2) |

PICK-AND-PLACE No significant main effects were observed on TCT, TLX, or SEQ for the pick-and-place task. Thus, the null hypotheses $H_0.1$-$H_0.3$ for the pick-and-place task could not be rejected. Similarity scores suggest that most participants perceived the task as similar across conditions, and Bayes factor analysis provided anecdotal evidence supporting the null hypothesis for TCT, indicating no effect. However, while again no significant effects were found for TLX and SEQ, Bayes factor analysis provided anecdotal evidence supporting the alternative hypotheses. This may be attributed to differences reported during interviews, where 13 participants noted a greater fear of collision in the *real* scenario, and seven described being more cautious. Participants emphasized that virtualizing the task was the primary factor in reducing perceived risk.

PEG-IN-HOLE For the peg-in-hole task, significant main effects were observed on TCT, TLX, and SEQ, with effect sizes $\delta_t$ indicating medium to large effects. Consequently, the null hypotheses $H_0.1$-$H_0.3$ were rejected in favor of the respective alternative hypotheses $H_1.1$-$H_1.3$ for the peg-in-hole task. Post hoc comparisons showed significantly higher TCT and TLX scores, and lower SEQ scores, in the *mixed* and *virtual* conditions compared to the *real* condition. These differences likely stem from perceptual challenges in the virtual tasks. Sixteen participants reported difficulties with the perception of depth and the peg's rotation, which made aligning the peg with the hole more challenging. Consequently, eleven participants described achieving precise motions as difficult. Although four participants described the perception of the real robot in the *mixed* condition as helpful, this did not result in significant differences between the *mixed* and *virtual* conditions, which was also reflected in the high

perceived similarity between the *mixed* and *virtual* conditions. This suggests that the use of a virtual task is the primary factor contributing to perceptual differences that result in increased task difficulty, longer completion times, increased workload, and decreased task ease.

ULTRASOUND The results for the ultrasound task show no significant main effects of the virtualization approach on all measures. Therefore, the null hypotheses $H_0.1$-$H_0.3$ for the ultrasound task could not be rejected. Post hoc Bayes analysis indicates anecdotal evidence supporting the absence of an effect for TCT and moderate evidence for TLX and SEQ. Differences between virtualization approaches may stem from variations in the ultrasound images, with 11 participants noting that the real image had more noise. Additionally, nine participants reported that contact was harder to perceive in the *mixed* and *virtual* scenarios, with eight attributing this to less complex feedback, such as reduced deformation of the phantom. Conversely, nine participants noted being afraid of applying too much contact force in the real scenario. Similarity scores emphasize the resemblance between the *mixed* and *virtual* scenarios.




(a) Task Completion Time

(b) NASA TLX Scores

(c) Single Ease Question

Figure 12.4: Violin Plots of experimental results of dependent variables. Means are indicated by ◇. Brackets indicate statistically significant post hoc pairwise comparison results. Adapted from [2].

## 12.4 DISCUSSION

This section discusses the findings related to virtualization level and task dependency, outlines the study's limitations, and summarizes the results into design implications.

VIRTUALIZATION LEVEL Previous research indicates that participants are more cautious with real robots due to collision concerns, while virtual robots are associated with lower perceived risk [63, 68, 251, 252]. This work confirms and extends these findings by highlighting the role of task virtualization. Specifically, it was found that the fear of collision is reduced even when only the task is virtual, while still using a real robot. This highlights the importance of both the robot's and the obstacles' virtuality. Additional perceptual differences were reported during the contact-intensive ultrasound task, where participants noted difficulty perceiving contact force between the end effector and the environment.

TASK DEPENDENCY However, despite participants' subjective feedback reporting perceptual differences, no significant impact was observed on the recorded measures for the pick-and-place and the ultrasound task. In contrast, for the peg-in-hole task, the *real* condition resulted in the best TCT, TLX, and SEQ. This suggests that the slower pace, potentially linked to more cautious behavior, was overshadowed by other confounding factors. Participant feedback suggested that perception, especially depth perception, played a crucial role, aligning with previous findings [251]. While perceptual challenges in XR are a relevant topic with ongoing research efforts to address them [12, 177], the findings highlight the task dependence of this limitation. As no significant differences were found between virtualizations for broader tasks, and they were perceived as similar, the employed fidelity seems sufficient for these tasks. However, for precise tasks in virtual settings, special attention should be given to visual fidelity.

GENERAL In this study, AR was used to facilitate XRP by integrating virtual components into the RE. Thus, in both the *mixed* and *virtual* conditions, everything except the task-relevant components remained real, including the testing environment and the user's hands. However, as discussed in Section 11.2, XRP can also be realized using VR, which simulates the entire environment. Utilizing VR enables the simulation of environmental factors that may influence task performance. While the current experiment investigated isolated tasks without considering external factors, VR offers potential for exploring more embedded tasks. Especially since previous work has shown that HCI experiments conducted in AR, VR, lab settings, and in situ can yield differing results [330], future research should examine virtualization levels more comprehensively, including VR and in situ environments.

### 12.4.1 *Limitations*

One limitation of this study is the use of a single fidelity level, which limits the ability to isolate specific factors influencing transferability between conditions. Although the implementation achieved a promising fidelity level based on the taxonomy by Plümer and Tatzgern [251], the use of the *HoloLens* constrained visual fidelity. While participants described the limited FOV as sufficient, potential concerns remain regarding shadowing and lighting effects. Previous

research suggests that certain performance measures may be independent of visual fidelity [291]. However, future studies should investigate the impact of varying fidelity levels on the transferability of performance between real and virtual tasks.

Another limitation stems from the study design itself, focusing solely on absolute validity, assessing whether the real and virtualized use cases yielded equivalent performance. However, related work suggests that even when absolute validity is not achieved, relative order effects may still align [252]. However, assessing such relative validity, i.e., whether the comparative outcomes between conditions remain consistent across environments, requires a study design that includes multiple factor levels. Therefore, future research should investigate transferability in comparative studies, examining whether relative validity holds, especially for precision tasks where absolute validity could not be achieved. In addition, the participant sample consisted primarily of individuals with an academic background, which may limit the generalizability of the findings. Future studies should consider including target user groups relevant to the tasks, such as industry professionals or medical practitioners, to identify challenges across different user groups and improve the external validity of the findings.

### 12.4.2 *Implications*

The findings of this study can be condensed into the following implications for XRP in HRI:

**The transferability of XRP results is task-dependent.** The results of the user study indicate that, while no significant differences were found between XRP environments for coarse tasks such as pick-and-place and ultrasound, significant differences did emerge for the more precise peg-in-hole task. This suggests that when employing XRP, it is essential to consider whether the specific task investigated is suitable for virtualization.

**Perceptual differences must be accounted for in all tasks.** Qualitative feedback revealed that perceptual differences were noted across all tasks, including reduced perception of depth, risk, and contact forces. These limitations should be taken into account when designing XRP studies, particularly those investigating perception or user experience.

**The primary factor influencing transferability is task virtualization.** The study results show that the level of task virtualization is the main factor influencing transferability, with no significant differences observed between entirely virtualized setups and those involving a real robot. This implies that, for tasks suitable for virtualization, entirely virtual setups can be used confidently for prototyping and experimentation, reducing the need for physical hardware and enabling more accessible and scalable research.

## 12.5 CONCLUSION

The goal of this work was to assess the transferability of HRI study results between real and XR setups. The results highlight that the XRP validity depends on the task being investigated. For broader movement tasks, such as the pick-and-place task, and complex tasks such as robotic ultrasound, evidence suggests that conducting the task in a real versus partially virtualized setting yields comparable results. However, significant differences between the real and the XR setups were found for the peg-in-hole task, which requires higher precision. The primary factor behind this difference seems to be task virtualization, as using a real robot with a virtual task still resulted in significantly different outcomes, with this setup being rated similarly to the virtual robot condition. This suggests that partial simulation of study setups in HRI research can yield valid results regardless of the level of virtualization, though perceptual challenges must be considered, particularly for tasks requiring precise perception and interaction. However, this study focused solely on absolute validity. For comparative user studies, relative validity, i.e., consistent comparative effects across environments, may also be relevant. To examine this in more detail, the next chapter will investigate a specific task, analyzing comparative effects more thoroughly.

# 13 INVESTIGATION ON XRP FOR FORCE ASSISTANCE

While Chapter 12 demonstrated the potential of XRP for yielding transferable HRI user study outcomes, qualitative feedback revealed perceptual differences, such as difficulties in assessing contact force and variations in perceived risk between real and simulated environments. Notably, the reported differences in force perception align with previous findings that stiffness and force cues can vary in XR [115, 355]. However, since prior work on XRP transferability has primarily examined tasks involving little to no contact [251], it remains unclear how differences in force perception between real and virtual environments affect transferability. To address this, the following section investigates XRP transferability in more detail for contact-intensive tasks, focusing on the robotic ultrasound task introduced in Chapter 12. Specifically, it examines not only absolute validity but also relative validity by applying the force assistance approaches introduced in Chapter 8 to determine whether the effects of partial automation and sensory substitution remain consistent across different prototype environments.

By comparatively evaluating both the prototype environments and the force assistance approaches, this chapter also contributes to RQ2.1 *Can force assistance improve efficiency in HRI?*. While Chapter 8 introduced partial automation and sensory substitution as techniques for assisting force perception, it did not compare them directly. This chapter fills that gap by assessing their relative performance and potential synergistic effects, thereby supporting a deeper understanding of force assistance techniques.

Parts this section were previously published in Tonia Mielke, Mareen Allgaier, Christian Hansen, and Florian Heinrich. "Extended Reality Check: Evaluating XR Prototyping for Human-Robot Interaction in Contact-Intensive Tasks." In: *Transactions on Visualization and Computer Graphics (TVCG)* 31.11 (Oct. 2025), pp. 10035–10044. DOI: 10.1109/TVCG.2025.3616753 [1].

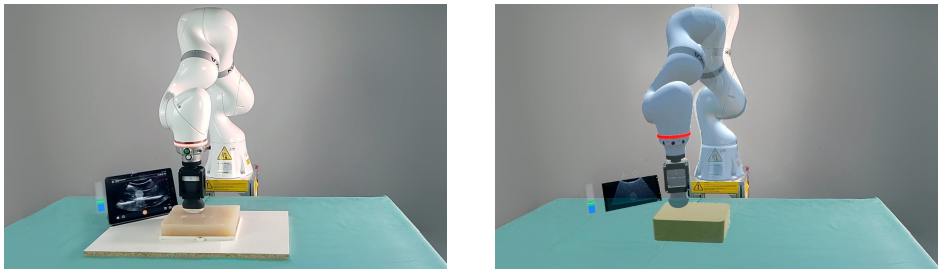
## 13.1 TECHNICAL METHODS

*Supplementary video*

This section outlines the technical implementation used to evaluate the transferability of XRP for contact-intensive tasks. Building on the force assistance concepts introduced in Chapter 8, it describes the two prototype environments and the implementation of partial automation and sensory substitution.

### 13.1.1 *Prototype Environments*

As the results presented in Chapter 12 indicated that no significant differences arise between virtualized setups that include the real robot and those that are entirely virtual, this chapter focuses on two extremes: an entirely real environment and one where the robot and the task are simulated.



(a) Real Prototype Environment (b) Simulated Prototype Environment

Figure 13.1: Overview of prototypes explored in the study. A bar for visual force feedback is displayed adjacent to the ultrasound image. Reprinted from [1].

#### 13.1.1.1 *Real Environment*

The real environment consisted of the components described in Section 2.3. A custom-built cuboid made of agar-agar was used as a phantom (for more details, see Section 2.3.5). It contained a 3D-printed bracket on which three spheres, made of agar-agar with a higher concentration, were placed. The visual force feedback, as well as visual feedback about the active interaction mode, was displayed beside the ultrasound screen (see Figure 13.1a).

#### 13.1.1.2 *Simulated Environment*

The simulated environment is implemented as described in Section 12.1.3. It was developed in *Unity* and displayed using a *HoloLens 2* as an optical see-through display, allowing the physical surroundings to remain visible while augmenting the task-relevant components (see Figure 13.1b). The robot, ultrasound probe, phantom, and screen were visualized using previously described 3D models (Section 12.1.3). The robot simulation and ultrasound rendering methods followed the same principles and performance considerations outlined earlier. Alignment between virtual and physical components was again ensured through the registration approach described in Section 2.3.4.

#### 13.1.1.3 *Robot Control*

The robot control was implemented as described in Section 12.1.4, using the same architecture based on a *Leap Motion Controller 2* for gesture input. As before, translation was mapped 1:1 and rotation at a 2:1 scale, with a foot pedal used to toggle between control modes. The Unity-based interface communicated with either the physical or simulated robot controller, and inverse kinematics were handled using the same pipeline described previously.

### 13.1.2 *Force Assistance Concepts*

In this section, the force assistance concepts previously investigated in Chapter 8, namely sensory substitution and partial automation, are employed. The specific implementation of the force assistance approaches used in this section is detailed below.

#### 13.1.2.1 *Visual Force Feedback*

As shown in Section 8.3, a bar visualization allows for more precise perception of contact force compared to an arrow visualization. Therefore, a bar is implemented that conveys force information through changes in height and color. The visualization follows the implementation described in Section 8.3, representing the absolute force applied along the probe's vertical axis. The color indicates the relation to the optimal force $F_d$: blue signals insufficient force, which may compromise acoustic coupling and image quality; red indicates excessive force, posing a risk of tissue deformation or injury; and green represents forces within the optimal range, defined as $F_d \pm 2.5\,\mathrm{N}$ (see Figure 13.2). In line with earlier results, the bar is positioned next to the sonographic image, as this placement has been shown to reduce perceived difficulty and support continuous perception.

The force data is again measured using the robot's force/torque sensors. As the forces involved are relatively small, the measurements are sensitive to digital noise, which is reduced using an empirically tuned Kalman filter.

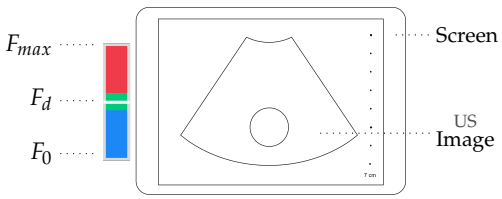



Figure 13.2: Design and placement of the visual force feedback. Reprinted from [1].

#### 13.1.2.2 *Contact Force Automation*

To automate contact force, the hybrid position/force control approach previously described in Section 8.2 is employed. This decouples the interactive position control from force regulation by controlling only the motion along the force direction using force control. The commanded joint torques, $\tau$, are thus determined through a combination of 5 DOF controlled by position control during interaction and 1 DOF controlled by force control (see Figure 13.3). An impedance control scheme is implemented to balance the external force, $F_{\text{ext}}$, with the desired force, $F_d$.

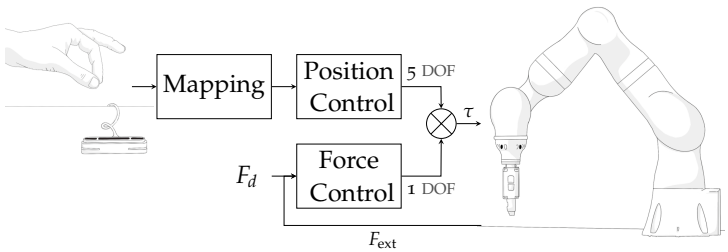



Figure 13.3: Loop diagram on force automation. Reprinted from [1].

## 13.2 EVALUATION METHODS

To evaluate the differences between force assistance methods in both real and simulated environments, a three-factorial within-subjects user study was conducted.

### 13.2.1 *Task*

For the evaluation, a task necessitating exploratory movement of the robotic end effector on a surface was required. Therefore, an ultrasound inspection task was selected. To facilitate this, an ultrasound phantom with three embedded spheres was developed. The task was to localize the three spheres and identify the one that differed in size from the others. To this end, spheres varying in size between 25 mm and 30 mm were embedded in the phantom. To minimize the influence of differing path lengths on study outcomes, the spheres were always placed in fixed positions (see Figure 13.4). Participants were instructed to maintain a force level as close as possible to the desired force $F_d = 10\,\mathrm{N}$ during movement. To minimize learning effects, three phantoms were provided for both the real and simulated environments. The phantom size, sphere dimensions, and sphere placement were identical in both the real and simulated environments to ensure consistent task difficulty.

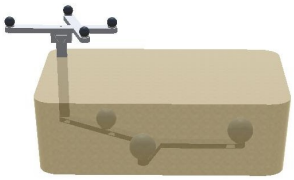


Figure 13.4: Setup of the phantom used in the user study. Three spheres are mounted on a bracket connected to a marker shield extending from the phantom to enable tracking.

### 13.2.2 *Variables*

INDEPENDENT VARIABLES The independent variables of the three-factorial study were the *environment* (real vs. simulated), the *visual feedback* (with or without visual force feedback), and the *force automation* (manual control vs. partial automation).

DEPENDENT VARIABLES The general purpose of this study was to assess the validity of XRP in evaluating force assistance techniques for HRI in contact-intensive tasks. To this end, commonly used objective and subjective measures for performance assessment were employed. The first objective variable was the TCT. To prevent the duration of moving the probe to the phantom from affecting this time, measurement started upon first reaching $F_d$, indicated

by the LED ring on the robot blinking blue. The time recording stopped when participants confirmed task completion by pressing a foot pedal. To quantify the consistency of the applied force, the second variable was the Average Absolute Deviation (AAD), calculated as the mean absolute difference between the applied contact force and $F_d$ at each time point.

$$\text{AAD} = \frac{1}{N}\sum_{i=1}^{N}|F_i - F_d| \tag{13.1}$$

The force was continuously recorded during interaction, excluding periods of inactivity, when the pinch gesture was not held, to prevent static probe positions from influencing the results. Data points for the applied force were acquired every 50 ms. A Kalman filter was employed to reduce measurement noise. To mitigate the effects of latency in force measurements and ensure uniform sampling across devices, the data were interpolated to an equidistant sampling rate of 1 kHz. Additionally, the *perceived workload* was assessed as a subjective measure, using the standardized raw NASA TLX.

As previous work has found differing perceptions of simulated robots compared to their real-world counterparts [191, 339], as well as the potential impact of partial automation on attitudes towards robots [217], *trust* was additionally assessed to identify differences in how participants perceive the robot across prototype environments and assistance approaches. Trust was measured using the 14-item version of the trust scale introduced by Schäfer et al. [285]. Following Eilers et al. [92], responses were recorded on a 5-point Likert item instead of using percentages.

### 13.2.3 *Hypotheses*

While the other studies described in this thesis were exploratory, with no prior assumptions about the expected outcomes, this study builds on the experiments reported in Chapter 8 and Chapter 12. The insights from those earlier investigations allowed for the formulation of specific hypotheses ($H$) regarding the influence of visual force feedback, force automation, and prototype environment.

**H1.** Previous work has found that visual feedback enables consistent force application [314]. Accordingly, it is hypothesized that force deviation will be reduced by visual feedback ($H1.1$). Since visual feedback, particularly in automated scenarios, has been associated with improved transparency that can foster trust [239], it is hypothesized that trust will be increased by visual feedback ($H1.2$). As prior studies have shown that task duration tends to increase when visual feedback is provided (see Section 8.3), task completion time is expected to increase with visual feedback ($H1.3$). This increase in duration is also expected to lead to greater perceived workload when visual feedback is available ($H1.4$).

**H2.** In related work, force automation has been shown to improve force stability [100, 370]. Based on these findings, it is hypothesized that force deviation will be decreased by force automation ($H2.1$). Related studies [267]

and Section 8.2 have also reported reductions in task completion time through force automation, so it is expected that task completion time will be reduced by force automation (*H*2.2). Additionally, perceived workload is expected to be reduced by force automation (*H*2.3), as indicated in prior work [267] and Section 8.2. Since automation has been found to decrease the sense of agency [320], which may in turn reduce trust [366], it is anticipated that trust will be reduced by force automation (*H*2.4).

**H3.** Prior studies have reported that stiffness and force perception can differ significantly between simulated and physical environments [115, 355]. These perceptual differences may influence task performance. Therefore, it is hypothesized that force deviation (*H*3.1), task completion time (*H*3.2), and perceived workload (*H*3.3) will be higher in the simulated environment compared to the physical environment. Since earlier work has found that people tend to respond differently to simulated robots [191, 339], it is also expected that trust will be lower in the simulated environment (*H*3.4).

**H4.** Previous research has shown that even when absolute measurements differ between real and simulated environments, relative trends may remain consistent [251, 252]. Therefore, despite the differences anticipated in *H*3, relative validity is expected, meaning that the general trends associated with force visualization and automation, as described in *H*1 and *H*2, are expected to be consistent in both the real and the simulated environment.

### 13.2.4 *Sample Design*

As outlined in Chapter 12, involving domain experts in user studies can provide more nuanced insights. Since the task in this study was robotic ultrasound, participants with knowledge of ultrasound imaging were required. Since no procedural expertise was necessary, medical students were considered suitable. Participants were recruited from the local university via online polls and were compensated with 30 €.

### 13.2.5 *Procedure*

Upon arrival, participants were given a brief introduction to the study's objectives. They were then asked to complete a written consent form and a demographic questionnaire. Next, they received an explanation of the robot control method and its safety features, including limitations on reach and velocity, as well as automatic stopping when force limits were exceeded. Participants then began the first block in the first environment, with the order of environments counterbalanced across participants. Each block consisted of four trials, one for each force assistance approach (none, visual feedback, force automation, and visual feedback + force automation). The order of these approaches was counterbalanced across participants using a Latin Square. For each force assistance approach, the trial began with an explanation of the assistance method, followed by a training trial in which participants performed

the task on a phantom. This training was intended to increase participants' proficiency in robot control and ensure they could successfully complete the task regardless of prior experience. Once they confirmed they were ready to proceed, two evaluation trials followed, conducted on different phantoms selected randomly. After completing the evaluation trials, participants were asked to complete the NASA TLX and Trust questionnaires. After each environment block, a short semi-structured interview was conducted to gather in-depth feedback on the force assistance concepts. Upon completing both blocks, a final semi-structured interview was conducted to explore potential differences between the real and simulated environments. The study took an average of 60 minutes.

### 13.2.6 *Statistical Analysis*

Since two trials were conducted for each combination of the three factors (*environment*, *visual feedback*, and *force automation*), TCT and AAD were averaged across identical experimental conditions for each participant. To evaluate the effects of the three investigated factors, the data for the dependent variables (TCT, AAD, TLX, and Trust) were tested for normality using the Shapiro-Wilk test and for homogeneity using Levene's test. If the assumptions were met, three-way repeated measures ANOVAs were conducted. When they were violated, robust three-way ANOVAs for within-subjects designs based on trimmed means were used (see [348]). For ANOVAs, the test statistic $F$ and effect size $\eta^2$ were reported. For robust ANOVAs, test statistic $Q$ was reported, and effect sizes for main effects were estimated using $\delta_t$, as proposed by Algina et al. [16]. For interaction effects, the more conventional effect size $\eta^2$ was used. Additionally, Bayes Factor analysis between the *environment* factor levels for all combinations of visual feedback and force automation was performed using Bayes paired t-tests [142]. All statistical analyses were conducted using R (version 4.4.0).

### 13.2.7 *Simulation Fidelity*

As in Section 12.2.6, the XRP validation framework proposed by Plümer and Tatzgern [251] was used to assess and correct potential confounding factors prior to the experiment.

*Visual Fidelity.* Since the simulated setup included the same virtual robot and ultrasound task as described in Chapter 12, and participants reported perceiving the real and the simulated environments for the ultrasound tasks as similar, the visual fidelity can be considered sufficient.

*Haptic Fidelity.* No direct physical contact with the robot or the environment was required for task completion. Therefore, haptic fidelity was identical across *environment* conditions.

*Audio Fidelity.* The primary audio source during the task was the sound of the robot, which was replicated in the simulated environment using a

velocity-dependent humming sound played through a speaker positioned beneath the real robot.

*Interaction Fidelity.* The same gesture-based interaction technique, using the same sensors, was employed in both the real and simulated conditions, resulting in consistent interaction fidelity.

*Functional Fidelity.* As in Section 12.2.6 consistent behavior between the physical and simulated robot was ensured by configuring both with the same maximum velocity, acceleration, and stiffness. The stiffness of the ultrasound phantom, as well as the interaction between the robot and phantom, were determined through palpation. To verify similar behavior between the real and simulated robot, they were again moved simultaneously, with the simulated robot overlapping the real one to identify potential offsets and enable empirical fine-tuning.

*Data Fidelity.* The same automated approaches and questionnaires were used for data collection, ensuring data fidelity.

*Simulation Overhead.* Since the HMD was necessary for force visualization and to eliminate simulation overhead caused by wearing it, participants were required to wear the HMD in all conditions.

## 13.3 RESULTS

This section presents the user study results and their interpretation.

### 13.3.1 *Participants*

In total, 26 medical students participated (18 identified as female, 8 as male), aged between 20 and 33 years (M=24.73, SD=4.08), and in their second to sixth year of study (M=4.00, SD=1.61). 13 participants reported no visual impairment and 13 reported having corrected-to-normal vision. They conducted the study with their dominant hand, with 24 being right-handed and two left-handed. Participants rated their task-relevant skills using a 5-point Likert item, where 1 indicated low and 5 indicated high. They reported their technical affinity as (M=3.27, SD=0.87) and their motor skills as (M=3.77, SD=0.71). Additionally, they rated their task-relevant experience on a scale from 1 (no familiarity) to 5 (high familiarity) for HRI (M=2.12, SD=0.99), MR (M=2.08, SD=0.82), and ultrasound (M=3.00, SD=1.06).

### 13.3.2 *Quantitative Results*

The following section outlines the results of the statistical tests, with a focus on significant findings. For a more detailed overview of the descriptive results and statistics, refer to Table A.5 and Table A.6. The descriptive results are presented in Figure 13.5, and the results of the Bayes factor analysis are shown in Table 13.1.

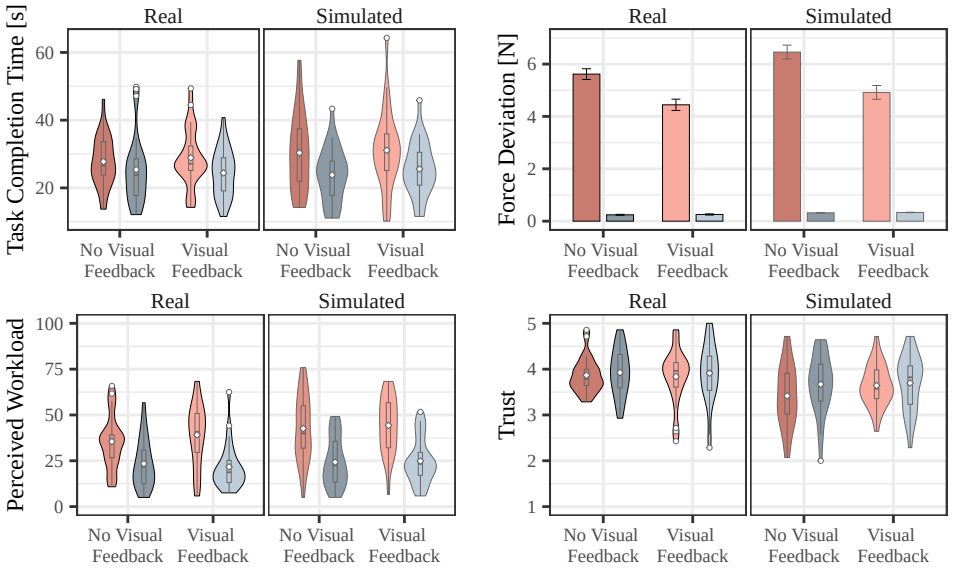



Figure 13.5: Descriptive results of the dependent variables. ■ & ■ show 'manual', while ■ & ■ show 'automated' *force automation* levels. Lighter colors represent the absence, while darker colors show the presence of *visual feedback*. Means are indicated by ◇. Adapted from [1].

Since TCT results were not normally distributed, a robust ANOVA was conducted. The test revealed a significant main effect of *force automation* on TCT ($Q = 19.31$, $p < 0.001$, $\delta_t = 0.561$ (medium), see Figure 13.6). All other main effects were not statistically significant, and no significant two-way or three-way interactions were found.

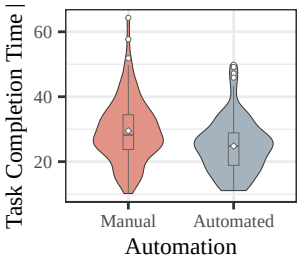



Figure 13.6: *Force automation* main effects on TCT. Means are indicated by ◇. Adapted from [1].

The force deviation results were also not normally distributed. Therefore, again, a robust ANOVA was conducted. Significant main effects of *environment* ($Q = 25.13$, $p < 0.001$, $\delta_t = -0.063$ (small)), *visual feedback* ($Q = 56.94$, $p < 0.001$, $\delta_t = -0.168$ (small)) and *force automation* ($Q = 1364.48$, $p < 0.001$, $\delta_t = 4.896$ (large)) were found (see Figure 13.7). Additionally, significant two-way interaction for *environment* × *force automation* ($Q = 17.78$, $p < 0.001$, $\eta^2 = 0.003$ (small)) and *visual feedback* × *force automation* ($Q = 62.24$, $p < 0.01$, $\eta^2 = 0.02$ (small)) were found (see Figure 13.8). No significant three-way interaction effects were observed.

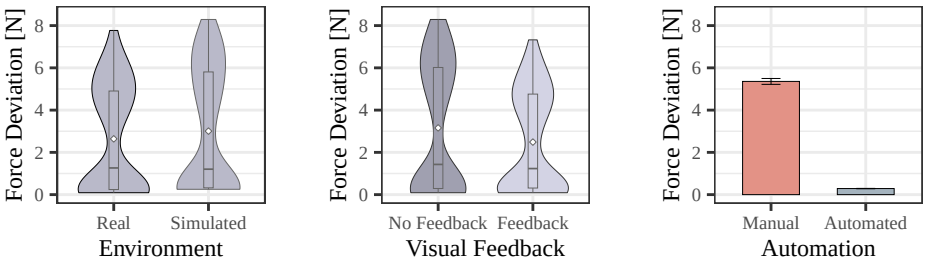



Figure 13.7: *Environment*, *visual feedback*, and *force automation* main effects on AAD. Means are indicated by ◇. Adapted from [1].

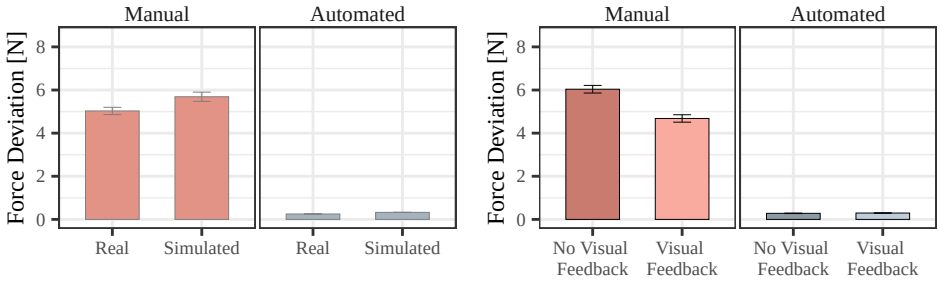



Figure 13.8: Two-way interaction effects on AAD. Bars represent means. Error bars represent standard errors. Adapted from [1].

The NASA TLX results were normally distributed; therefore, a repeated measures ANOVA was conducted. The test revealed significant main effects of *environment* ($F = 7.04$, $p = 0.014$, $\eta^2 = 0.01$ (small)) and *force automation* ($F = 59.71$, $p < 0.001$, $\eta^2 = 0.025$ (large), see Figure 13.9). No other main effects were found, and no significant two-way or three-way interactions were observed.

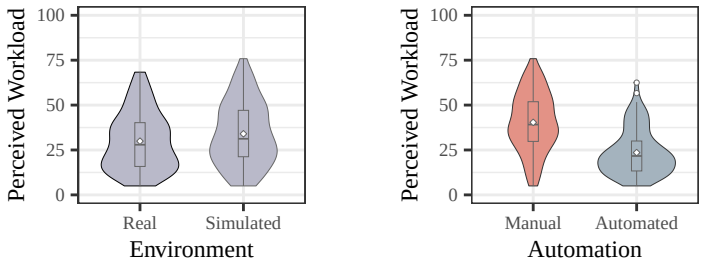



Figure 13.9: *Environment* and *force automation* main effects on TLX. Means are indicated by ◇. Adapted from [1].

As trust results were normally distributed, repeated measures ANOVA was conducted, revealing significant *environment* ($F = 10.54$, $p = 0.003$, $\eta^2 = 0.06$ (medium)) and *force automation* ($F = 6.44$, $p = 0.017$, $\eta^2 = 0.009$ (small)) main effects (see Figure 13.10). No other main effects or significant two-way or three-way interactions were found.

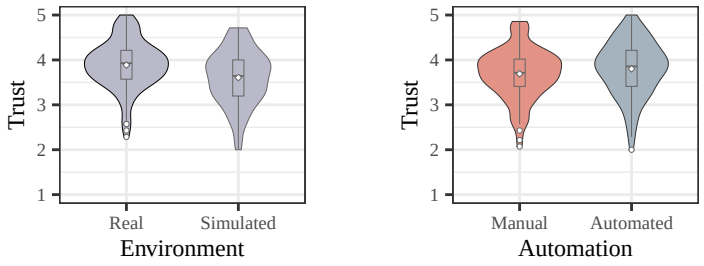



Figure 13.10: *Environment* and *force automation* main effects on trust. Means are indicated by ◇. Adapted from [1].

Table 13.1: Summary of Bayes factor analyses results between *Environment* levels. For each measure, $BF_{10}$ is listed. **Color Legend:** Moderate evidence for $H_0$, Anecdotal evidence for $H_0$, Anecdotal evidence for $H_1$, Strong evidence for $H_1$, Very strong evidence for $H_1$, Extreme evidence for $H_1$. Adapted from [1].

| Visual Feedback | Force Automation | TCT | AAD | TLX | Trust |
|---|---|---|---|---|---|
| Absence | Manual | 0.43 | 2.23 | 1.86 | 668.90 |
| Absence | Automated | 0.32 | 47.77 | 0.22 | 2.33 |
| Presence | Manual | 0.35 | 1.52 | 0.68 | 0.55 |
| Presence | Automated | 0.31 | 16.43 | 1.37 | 0.56 |

### 13.3.3 *Qualitative Results*

The qualitative feedback provided by participants during the semi-structured interview was paraphrased and clustered. Statements that were consistent across at least two participants were considered for clustering. In total, 239 individual statements were recorded, of which 168 were repeated by at least two participants. These repeated statements were then clustered into 37 summarizing categories. The clustered statements are presented in Table 13.2.

### 13.3.4 *Interpretation of Results*

**H1.** An interaction effect was observed regarding force deviation, indicating that force deviation is independent of visual force feedback when contact force is automated, since during force automation the applied force is independent of user input. However, for manual control, the presence of visual force feedback was found to improve force deviation, as hypothesized in *H*1.1. This finding is consistent with previous work [314], and as participant feedback suggests, the result may be explained by the general difficulty in perceiving contact force. The visual feedback was reportedly perceived as helpful for assessing and applying the desired contact force more accurately.

An increase in trust was additionally expected as a result of visual feedback (*H*1.2). However, no significant effects of visual feedback on trust were observed. One possible explanation is that the overall performance of the

Table 13.2: Clustered feedback of participants during the semi-structured interview with the frequency (#). Adapted from [1].

| **Force Assistance** | **Prototype Environment** |
|---|---|
| *Visual Force Feedback:*<br>+ Visualization helpful in manual force control (9) and in general (6)<br>+ Without visual feedback, assessing contact force is difficult (2)<br>- Distracting (5), induces insecurity (4), requires concentration (4), and is overall not helpful (3) and overwhelming (2)<br>- Interpreting visual feedback difficult (3)<br>- LED provides sufficient feedback (3)<br>*Contact Force Automation:*<br>+ Automation helpful (11), eases interaction (7), and requires less concentration (7)<br>+ Manually controlling force more difficult (2)<br>+ Automation makes interaction faster (3) and less stressful (2)<br>- Feeling less in control (5)<br>*Combination Feedback and Automation:*<br>+ Using visualization to verify autonomous control (3)<br>- Giving no attention to the visualization when force automated (13) | • Prototype environments similar (11).<br>• Perception of contact force (9) and depth (6) more difficult in the simulated environment<br>• Visual feedback especially helpful in the simulated environment (6)<br>• Robot behavior perceived as similar in both environments (4)<br>• Less fear of causing damage (3) and lower perceived risk (2) in simulated environment<br>• Real US easier to read (4) and clearer (2)<br>• Simulated US image easier to read (3)<br>• More feedback from the surroundings in the real environment (4)<br>• Real environment feels more familiar (2)<br>• Interaction more difficult in the simulated environment (2)<br>• Sounds in simulated environment helpful (2)<br>• Autonomous control more helpful in the simulated environment (2)<br>• Perception of the robot is easier in the real environment (2)<br>• Feeling of control stronger with real robot (2) |

robot, which has been identified as a major factor influencing human-robot trust [128], remained similar across all conditions. As a result, the visual feedback may not have produced a distinct impact. Furthermore, some participants indicated that the feedback provided when exceeding the force threshold was sufficient, suggesting that additional visual feedback was not perceived as necessary to enhance trust further.

A negative effect of visual feedback on task completion time ($H1.3$) and perceived workload ($H1.4$) had also been hypothesized. However, no significant effects on either measure were observed. This outcome suggests that the advantages of the visual feedback in supporting force perception may have outweighed its cognitive demands. Although some participants mentioned that the feedback was distracting, required increased concentration, or felt overwhelming, these negative aspects appeared to have been mitigated by the overall usefulness of the visual force feedback in improving task performance. Thus, $H1.1$ is accepted, while $H1.2$–$H1.4$ are rejected.

**H2.** In previous work, positive effects of contact force automation on force stability [100, 370], task completion time [267], and perceived workload [267] were reported. In line with these findings and as hypothesized, force automation was found to improve force deviation ($H2.1$), task completion time ($H2.2$), and perceived workload ($H2.3$), regardless of the prototype environment and the presence of visual feedback. Participant feedback further supports these results, with several participants describing the automation as helpful and noting that it eased the interaction.

A reduction in trust caused by automation was also hypothesized ($H2.4$). Although some participants reported feeling less in control during partial au-

tomation, significantly higher trust was observed in the automated condition compared to manual control. This result may be attributed to system performance influencing trust [128], particularly in automated systems, where better performance has been shown to increase trust [217]. Therefore, *H*2.1–*H*2.3 are accepted, while *H*2.4 is rejected.

**H3.** It was hypothesized that force deviation (*H*3.1) and perceived workload (*H*3.3) would be higher in the simulated environment. The data supported these hypotheses. An interaction effect showed that force deviation was unaffected by the prototype environment when force was automated, but during manual control, significantly higher deviation was observed in the simulated environment. In addition, a higher perceived workload was reported in the simulated environment. Qualitative feedback further supports these results, with participants indicating greater difficulty in perceiving both contact force and depth in the simulated version.

An increase in task completion time was also expected (*H*3.2), but no significant main effects of the *environment* factor on task completion time were found. Bayes factor analysis revealed anecdotal to moderate evidence in favor of the null hypothesis. A possible explanation could be a time-accuracy trade-off: although more effort was required to achieve adequate contact force due to perceptual challenges, this effort was reflected more in differences in force accuracy than in the time taken to complete the task.

As expected, trust was significantly lower in the simulated environment (*H*3.4). Bayes factor analysis between *environment* conditions revealed extreme evidence for the alternative hypothesis in the manual control condition without visual feedback, indicating that trust significantly differed between the simulated and real settings. This outcome may be attributed to the limited visual fidelity of the simulated robot, which might have reduced participants' perception of the robot and its interaction with the environment. However, this effect appears to have been mitigated by force assistance, particularly by visual feedback. When visual feedback was present, the results provided anecdotal evidence for the null hypothesis, suggesting no significant difference in trust between the real and simulated conditions. This indicates that visual feedback may help overcome perceptual differences and enhance trust in simulated environments. Based on these findings, *H*3.1, *H*3.3, and *H*3.4 are accepted, while *H*3.2 is rejected.

**H4.** As demonstrated in *H*3, most independent variables differed between environments. However, the concept of relative validity suggests that while absolute measures may differ, relative trends in comparative studies tend to remain consistent [251, 252]. In this context, no interaction effects involving the *environment* factor were found for task completion time, perceived workload, or trust. This indicates that the general trends observed for the force assistance concepts described in *H*1 and *H*2 were consistent across both physical and simulated environments.

For average absolute deviation, an interaction effect involving the *environment* factor was found. However, a closer examination of the descriptive results revealed that this effect stemmed from a difference in average abso-

lute deviation during manual control, which was eliminated when partial automation was applied. Despite this, the relative difference between the presence and absence of automation, namely that automation improves average absolute deviation, remained consistent across both environments. Based on this, $H4$ is accepted.

## 13.4 DISCUSSION

In the following, the results of the conducted studies are discussed, the limitations are addressed, and implications are outlined.

**XRP VALIDITY** While this study demonstrated that conducting a user study in a simulated environment using XR can yield transferable comparative results, significant absolute differences between the real and simulated environments were observed. One factor contributing to the lack of absolute validity could be visual fidelity. Participants reported difficulties in assessing contact force and issues with depth perception in the simulated scenario. The visual realism was constrained by the *HoloLens 2* as hardware, which has computational limitations. This restricted the use of features like shadows or deformations that could have aided in conveying contact force and depth. However, since this study only explored one level of realism, identifying the specific factors influencing perception was not possible. Future work could explore varying levels of fidelity, as demonstrated by Schott et al. [291], to identify visual factors that enhance task performance and help achieve absolute validity. In addition to examining how visual fidelity affects performance measures, future research should investigate how it influences participants' attitudes towards simulated robots. The results, particularly the differences in trust and qualitative feedback, suggest varying perceptions between the real and simulated robots. Investigating whether these factors can be mitigated by improving visual fidelity could be a crucial step towards achieving absolute XR study validity.

**FORCE ASSISTANCE APPROACHES** The force assistance approaches were previously investigated individually in Chapter 8. The experiment in Section 8.2 revealed that automating the contact force significantly reduces task duration and perceived workload. The present study expands these findings by showing that force automation also significantly improves force deviation and trust, two measures not evaluated before.

Regarding sensory substitution via visual feedback, the experiment presented in Section 8.3 found that visual feedback improved subjective measures such as perceived difficulty and perceived continuity but increased task duration. In contrast, the experiment in this chapter did not find such negative effect. Instead, it showed that the presence of visual feedback decreased force deviation. The absence of negative effects on completion time, combined with the improvement in force deviation, may be attributed to this study's use of the optimal visual feedback design and placement identified in Section 8.3. The comparison of visual feedback to the baseline with no feedback in the pre-

vious study included aggregated values across all visual feedback conditions. Therefore, this study extends the findings of Section 8.3 by demonstrating that optimizing visual feedback design and placement can mitigate negative effects such as prolonged task duration, while achieving positive effects such as reduced force deviation.

This study also enabled a direct comparison between the force assistance approaches. The results show that force automation reduces force deviation more than visual feedback and additionally shortens completion time, lowers workload, and increases trust. Finally, this study examined the potential synergy between the two methods. However, the results indicate that combining them does not yield any specific advantages. Together, these findings indicate that employing partial automation alone is the most effective way to enhance efficiency in HRI for contact-intensive tasks.

### 13.4.1 *Limitations*

One limitation of this study lies in the simplicity of the force assistance approaches employed. Both techniques were limited to visualizing and automating a force in one single direction. This was based on related work showing the effectiveness of one-dimensional force automation [155, 190] and the utility of visualizations without directional cues [308]. While the probe could be both translated and rotated during the study, the task itself did not require substantial tilting. Tasks involving more pronounced tilt and other use cases, such as those involving irregular surface geometries or varying stiffness, may require more sophisticated assistance strategies. For example, this could include multidimensional visualizations or automation of both translational and rotational movements [370]. Nevertheless, even with these basic assistance approaches, these findings suggest that XR simulation can effectively investigate the comparative performance of assistance strategies. When functional fidelity is controlled, similar performance benefits of automation and visualization were observed in both the real and simulated environments, which highlights the potential of XRP for force assistance approaches in contact-intensive tasks.

Another limitation of this study arises from its specific context. The exemplary use case of robotic ultrasound was investigated, which is unique in that, besides observing the end effector, the ultrasound image is also used to assess contact force and to navigate. Participants noted differences in the ultrasound images between the real and simulated environments, which may have influenced the study results. While this task presents unique challenges and benefits, it is believed that the findings could be transferable to other contact-intensive robotic tasks, such as polishing, grinding, or medical procedures. This is because such tasks share key characteristics with robotic ultrasound, including the need for precise contact control and continuous surface interaction. The overall conclusion that XR setups, even with perceptual challenges, can yield valid comparative findings on force assistance techniques is, therefore, likely generalizable to similar contact-intensive tasks. The force assistance techniques themselves are also believed to be applicable

beyond robotic ultrasound. For visual force feedback, adjustments to the visualization placement and the scaling of the bar may be necessary. As for automation, more sophisticated approaches may be required for complex scenarios, but it is assumed that the findings on the utility of force assistance are transferable, provided the methods are adjusted for the specific application.

Using a medical task for the study additionally influenced the choice of study population, with medical students selected as participants. Since medical students might be less familiar with robotics compared to individuals in engineering fields, differences might arise. For example, prior experience with robots has been shown to influence trust [281]. Furthermore, the study population consisted primarily of relatively young individuals who reported some familiarity with HRI and MR, which may not fully represent the typical user population in medical settings. The gender distribution in the study was not uniform, with a higher proportion of female participants. This imbalance resulted from the random sampling approach, which recruited medical students without gender-specific criteria. Therefore, future research should include a more diverse population, with participants from different backgrounds and a more balanced gender distribution, to improve external validity and ensure inclusion and diversity.

### 13.4.2 *Implications*

The insights on XRP for HRI in contact-intensive tasks, as well as on force assistance through visual force feedback and force automation, gained from this user study can be summarized in the following implications:

**Simulated user studies can yield valid comparative results for contact-intensive tasks.** The results show that while absolute measures (e.g., task completion time or force deviation) may differ between real and simulated environments, the relative differences between conditions, such as the effects of automation or visualization, are consistent. This supports and extends the concept of relative validity proposed in prior work [251, 252], and demonstrates that it also applies to more complex tasks involving physical contact, in which perception may be affected by reduced visual fidelity.

**Visual force feedback can enhance force perception during manual robot control.** It was found that visualizing contact force reduces force deviation during manual control. Since no further significant effects on task duration or perceived workload were observed, it can be concluded that visual feedback meaningfully enhances performance by improving force perception.

**Partial automation can improve performance in contact-intensive tasks.** Consistent with previous work [100, 267, 370], it was found that partial automation, specifically, the automation of contact force, results in improved force accuracy, reduced task completion time, and lower perceived workload. These findings are expanded by the additional observation of increased trust when contact force was automated. As a result, support is provided for the use of partial automation in tasks that require physical interaction.

**Partial automation alone is sufficient for improving efficiency.** The findings further indicate that partial automation reduces force deviation more than visual feedback, while also providing other benefits such as reduced completion time, lower perceived workload, and increased trust. Additionally, no synergistic effects between the two force assistance approaches were found, indicating that force automation alone is sufficient to support users efficiently.

## 13.5 CONCLUSION

In this work, the validity of XRP for user studies in contact-intensive robotic tasks was investigated. To this end, a comparative study on force assistance techniques was conducted using the exemplary use case of robotic ultrasound. By evaluating both visual force feedback and contact force automation, the aim was to gain insights into the transferability of simulated user studies in XR to their real-world counterparts. The results indicate that even in complex robotic scenarios, relative validity can be achieved, despite absolute differences between real and simulated environments. Absolute differences include variations in force deviation, perceived workload, and trust. Regarding force assistance techniques, visual force feedback was found to reduce force deviation during manual robot control, while contact force automation led to improvements in task completion time, force deviation, perceived workload, and user trust. This study highlights the potential of XR-based simulations to provide valid comparative results, even for tasks where the robot interacts physically with its environment.

The final part of this thesis focused on exploring XRP as a tool for conducting HRI user studies. The goal was to investigate how well results from user studies performed in partially simulated XR test environments transfer to real-world settings. To this end, two experiments were conducted.

The first experiment examined the task-specific transferability of XRP across different levels of virtualization and a range of HRI tasks. With regard to the level of virtualization, no significant differences were found between setups that included a virtual task with a physical robot and those that were entirely virtual. This suggests that the main factor influencing transferability is the virtualization of the task itself. The results also showed that transferability is task-dependent. While precise applications, such as peg-in-hole, revealed significant differences between real and virtual setups, broader movements, such as pick-and-place, and more complex procedures, such as ultrasound, showed no significant differences. Nonetheless, qualitative feedback highlighted perceptual differences that affected interaction across all scenarios.

To further investigate these perceptual differences, especially regarding the perception of contact forces between the robotic end effector and the environment, a second experiment was conducted. This experiment investigated the transferability of XRP for contact-intensive tasks by comparing the two force assistance approaches introduced in Chapter 8 in both a real prototype and a simulated XR environment. The findings indicate that although absolute performance differed, relative validity was achieved. Both prototype environments produced consistent comparative results between the force assistance methods, highlighting the potential of XRP for comparative evaluations.

Together, these findings contribute to a more systematic understanding of how XR can be applied as an evaluation tool in HRI research. The results show that the transferability of XRP study outcomes is task-dependent and primarily influenced by the virtualization of the task rather than the presence of a physical robot. Furthermore, even in contact-intensive scenarios where absolute validity may be limited due to perceptual differences, relative validity can still be achieved. These insights lay the groundwork for developing practical implications for the use of XRP in HRI, supporting its role as a reliable and efficient tool for early-stage prototyping and evaluation. This, in turn, can enable more accessible, scalable, and cost-effective HRI research.

# CLOSING

# 15 CONCLUSION

## 15.1 THESIS SUMMARY

The general objective of this thesis was to investigate different factors influencing the efficiency of HRI for robot control. To address this, three main research questions were formulated.

> **RQ1** | How can efficient and accurate XR-to-robot registration be achieved?

This question was motivated by the need for accurate alignment of the coordinate spaces of the interaction and the robotic workspace. To this end, two experiments were conducted. The first aimed to identify accurate and efficient registration methods by comparing manual and point-based approaches. While point-based registration was found to be more efficient than manual alignment, the resulting registration accuracy was not sufficient for tasks requiring high precision. To address this, a technical evaluation of different influencing factors, including tracking method, XR device placement, and registration point configuration, was conducted. Additionally, a refinement approach was investigated, using points recorded between the registration points to perform point cloud registration. Results indicate that registration accuracy can be significantly improved by carefully selecting the tracking technique, marker size, and registration point characteristics, and employing the proposed refinement approach. These findings demonstrate the potential of point-based registration for efficient and precise alignment.

> **RQ2** | How can efficient mid-air gesture-based HRI be designed?

Previous work on hand gesture-based robot control has identified two key challenges: the lack of haptic perception and the fatigue caused by mid-air interaction. Therefore, the second RQ was addressed by formulating two sub-questions:

> **RQ2.1** | Can force assistance improve efficiency in HRI?

This investigation into methods assisting the force perception during HRI included three experiments. The first experiment aimed to evaluate partial automation as a support mechanism. In a user study, the impact of hybrid position-force control, automating contact force, was compared across hand-guiding and hand gesture interaction for two distinct tasks. The results indicate that hand gesture interaction without automation is outperformed by the other methods. However, partial automation can mitigate the negative effects of missing haptic feedback, enabling gesture-based interaction

with automation to perform comparably to hand-guiding for broad tasks. Furthermore, it was shown that for tasks requiring precision, gesture-based control with automation can result in faster interaction than hand-guiding, highlighting the potential of hand gestures for precise robot control.

A second experiment investigated sensory substitution as an alternative strategy to address the lack of haptic feedback. Focusing on visual and vibrotactile feedback as modalities and design variables, a user study revealed the potential of visual feedback and identified important factors, including feedback design and placement, for effective force assistance in HRI.

The third experiment, conducted as part of the investigation into the transferability of XRP, comparatively evaluated the two previously described force assistance techniques. The results showed that while sensory substitution improves force deviation, automation reduces force deviation even further and additionally improves task duration, perceived workload, and trust.

**RQ2.2** | How can fatigue be reduced in hand gesture-based HRI?

The second sub-question addressed the design of hand gesture control itself, particularly concerning fatigue caused by mid-air interaction. Therefore, both position and rate control were investigated, as well as five different sensor placements for hand tracking. The results indicate that position control is more suitable for precise, efficient interactions than rate control. While placing the sensor on the robot enabled intuitive and accurate control, this setup led to physical strain. In contrast, placing the sensor on a desk offered a more balanced solution, enabling good accuracy while significantly reducing fatigue.

**RQ3** | Can XRP produce transferable results for HRI research?

The final research question aimed to assess the use of XR as a prototyping environment for HRI, motivated by XR's potential to enable inexpensive, rapid, and accessible development. To evaluate whether user studies conducted in simulated XR setups produce valid results, two experiments were carried out. The first experiment explored the impact of different levels of virtualization, including simulating only the task and simulating the entire setup, as well as the influence of task type. The study found no significant differences between using partially or fully simulated setups. However, the validity of the simulated environments compared to the real setup was task-dependent. For broader or more complex tasks, no significant differences were found between real and XR setups. In contrast, for tasks requiring high precision, results differed significantly between conditions. Participant feedback indicated noticeable perceptual differences between the real and simulated setups, particularly regarding depth perception, perceived risk, and contact force. These findings suggest the need for further investigation, especially for evaluations where the perception of contact force is essential.

As a result, a second experiment was conducted to assess the transferability of user studies for contact-intensive tasks. Leveraging the force assistance con-

cepts developed in RQ2, this experiment examined whether the effectiveness of these assistance approaches is transferable between real and simulated setups. The results indicate that although absolute validity (i.e., identical performance and user experience) was not achieved, relative validity was observed. This suggests that while outcomes differ across environments, comparative trends remain consistent. Therefore, XR setups may offer a promising path for efficient comparative evaluations, even in contact-intensive scenarios.

## 15.2 CONTRIBUTION

By answering the three posed research questions, this work represents an important step toward enabling efficient HRI. It establishes the following design implications, specifically relevant to scenarios comparable to those investigated in this work.

*How can efficient XR-to-robot registration be enabled?*
To enable efficient registration, point-based registration should be employed. The tracking method and marker size should be carefully selected to achieve optimal registration accuracy. Additionally, choosing appropriate registration point characteristics in alignment with the selected tracking technique can further improve accuracy. Finally, refinement using point cloud registration, based on points recorded while moving between registration points, should be applied.

*How can efficient mid-air gesture-based HRI be designed?*
To address the absence of haptic feedback, automation of contact force should be applied. If this is not feasible, sensory substitution through visual feedback can improve force perception. To translate gestures into robot motion, position control should be used, as it is more intuitive and efficient than rate control. Sensor placement should also be optimized depending on the use case. Sensors mounted on the robot allow for intuitive control but may increase physical strain. In contrast, placing sensors in front of the user offers a balanced trade-off between accuracy and physical comfort.

*Can XRP produce transferable results for HRI research?*
The transferability of XRP depends on the specific task. For precise tasks, perceptual differences may lead to variation between results obtained in physical and virtual setups. However, even though such differences can influence absolute performance outcomes, it was shown that XRP can still provide valid comparative results. Therefore, while perceptual differences should be taken into account, XRP has potential for generating transferable comparative findings.

*Use Case: Robotic Ultrasound*
As outlined in Section 1.2.4 the research in this thesis was partially centered on the specific use case of robotic ultrasound. Hence, the resulting guidelines and implications primarily apply to this task. However, to support transferability to other use cases, several parts of the work deliberately addressed a range of tasks: broad and precise tasks in Section 8.2, a general precision task in Chapter 9, and different practical tasks in Chapter 12. Consequently, the findings are believed to be applicable beyond the presented use case. In particular, the research on XR-to-robot registration is relevant for any application integrating XR into robotic workspaces with robot manipulators.

## 15.3 LIMITATIONS AND FUTURE WORK

The findings of this thesis are subject to several limitations. While the limitations of each study are discussed in their respective sections, this section summarizes the main technical and methodological limitations of the thesis as a whole.

### 15.3.1 *Technical Limitations*

One major limitation of this thesis is that each RQ was investigated using only a single hardware setup. As a result, the findings might have been influenced by the specific devices used in the individual experiments.

For example, the tracking accuracy of the *KUKA LBR iiwa* and the *HoloLens* may have influenced the registration results discussed in Part I. The intrinsic sensors of the *KUKA LBR iiwa* may have impacted the hand-guiding baseline condition in Section 8.2. The characteristics of the *Leap Motion* sensor could have affected the evaluation of different sensor placements in Chapter 9, while the visual fidelity of the *HoloLens* influenced the realism of the XR prototyping approaches in Part III. In addition, this thesis focused on HMDs as the XR device of choice, based on their widespread use in the field [205]. However, other XR devices, such as projectors or handheld displays, can also be used to augment robotic workspaces [311].

While these limitations may have influenced absolute performance outcomes, the relative trends observed across conditions may be platform-independent. For example, the comparative effectiveness of different registration methods or force assistance techniques, as well as the relative validity of XRP, may not depend on specific hardware. Thus, the core findings of this thesis are expected to generalize beyond the particular hardware used. Nonetheless, future work should validate these results using different robotic platforms and XR devices, including projectors or handheld displays, to confirm their broader applicability.

### 15.3.2 *Methodological Limitations*

In addition to technical limitations, the findings are also shaped by methodological constraints related to the study tasks, selected variables, participant sample, the focus on hand gestures, and the isolation of specific variables.

TASKS The studies conducted in this thesis focused on isolated tasks in controlled laboratory environments. As a result, task durations were relatively short, which may have limited the ability to observe effects, such as fatigue. This design choice was based on the assumption that hand gesture-based HRI would be used to supplement autonomous control, for example, by correcting or guiding individual subtasks. However, future work should integrate the investigated approaches into full workflows to assess how real-world conditions, such as multitasking or environmental distractions, influence the results.

In the studies related to force perception, robotic ultrasound was selected as a representative use case for contact-intensive tasks. While maintaining a continuous contact force is a requirement shared with industrial tasks, such as polishing or grinding, ultrasound imaging includes a secondary task that involves interpreting the ultrasound image for navigation. This additional component may have influenced specific outcomes, such as the placement of visual feedback. Although the general principles of HRI explored here are expected to be transferable, future research should examine their relevance across different domains and task contexts.

VARIABLES The choice of independent and dependent variables shaped the scope of this thesis. Each study investigated different independent variables, which were made up of several aspects. For example, feedback designs in Section 8.3 and control methods in Chapter 9 each included different characteristics. While this approach reveals general trends, examining individual characteristics in more detail could provide deeper insights.

While this thesis included semi-structured interviews to gather qualitative feedback, the core definition of efficiency relied primarily on quantitative measures, such as task duration, accuracy, and workload. As a result, other important aspects, such as usability, learnability, or user satisfaction, were not explicitly evaluated. This limits the scope of the findings to task-based efficiency and may overlook broader dimensions of interaction quality. Future work should incorporate established usability metrics and user experience evaluations to provide a more holistic assessment of interaction methods.

STUDY SAMPLE The sample sizes in the individual experiments were relatively small, primarily due to challenges in recruiting suitable participants. Larger sample sizes could have increased the statistical power of the analyses, thereby reducing the likelihood of missing true effects. However, the sample sizes used in this thesis are comparable to those reported in related studies in the field of HRI [20] and, more broadly, HCI [55], and are thus consistent with established research practices.

The participants in the studies were primarily students with engineering or medical backgrounds, depending on the task domain. While their expertise was considered sufficient to complete the study tasks and yield valid results, they do not fully represent the intended user groups, such as medical professionals for robotic ultrasound or industrial workers for manufacturing tasks. Including domain experts could offer deeper insights into the practical integration of the methods into real-world workflows. Future studies should therefore involve professionals from the relevant fields.

FOCUS ON HAND GESTURES While the main RQ of this thesis concerns general efficiency in HRI, the experiments focused specifically on hand gestures as the primary interaction modality. This choice was motivated by promising findings in related literature highlighting the potential of hand gestures for intuitive robot control. However, other input modalities or multimodal approaches may also support efficient HRI. Section 8.2 compares hand gestures combined with contact force automation to a state-of-the-art hand-guiding approach and shows promising results, but further alternatives should be explored. Although such alternatives were examined in a co-authored publication by Schreiter et al. [9], this investigation falls outside the scope of the present thesis. Future work should systematically explore different interaction modalities, including multimodal input, to further identify effective strategies for enhancing HRI.

ISOLATION OF DIFFERENT FACTORS In this thesis, key aspects of HRI were isolated and optimized. For example, while registration efficiency was improved, the practical benefits of increased registration accuracy were not evaluated in the broader context of HRI. Similarly, although different elements of hand gesture-based interaction were enhanced individually, the fully optimized prototype was not compared against alternative interaction methods. Future work should therefore embed the findings of this thesis within a broader range of HRI approaches to assess their general applicability and practical impact.

### 15.3.3 *Future Work*

This thesis focused on HRI for the direct control of robot manipulators. While the control of robot manipulators is essential in industrial and medical robotics, future work could explore the transferability of the findings to other HRI task domains. In particular, this includes investigating the applicability of the interaction designs and the findings on XRP to other types of robots, such as mobile or social robots, as well as to the control of drones. Testing applicability in these contexts would also require identifying missing or platform-specific functionalities, since different robot types may necessitate additional or adapted capabilities or interaction mechanisms [50, 58].

Additionally, future work could explore HRI tasks beyond the direct control of the robot's DOF. This task domain was selected because it is meaningful to teach robots new tasks or to control them in situations that are too

risky or too complex for autonomous operation. However, especially in HRC, interaction could benefit from more implicit forms of communication, understanding the human collaborator's verbal and non-verbal behavior, and adjusting the robot's control accordingly, without requiring direct movement commands [333]. While such interaction at higher LORA might still benefit from this thesis's findings regarding registration and XRP, future work could further explore interaction concepts for higher-level HRI.

## 15.4 GENERAL CONTRIBUTION

The goal of this thesis was to explore the main RQ: *How can efficiency be improved in HRI?* To answer this question, the thesis investigated three key research areas: registration, the design of the interaction, and the use of XRP for evaluation. Based on seven experiments, evidence-based recommendations were developed for registration methods and parameters, strategies to address missing haptic feedback and physical fatigue, and the use of XRP in user studies for HRI.

With these findings, this thesis contributes to enabling more efficient HRI for robot control. The proposed registration methods support accurate alignment of XR and robotic workspaces. The insights on hand gesture-based interaction design offer guidance on how to integrate human expertise into robot control efficiently. Finally, the findings on XRP demonstrate its potential as a more accessible and valid approach for conducting HRI research. Overall, the results support the development of more human-centered HRI by providing practical tools to assist users during interaction with robot manipulators.

# A APPENDIX

## A.1 SUPPLEMENTARY RESULTS

### A.1.1 *Investigation on Registration Parameters*

Table A.1: Summary of descriptive results reporting the TRE in mm for each combination of *Tracking Method* and *Condition*. All values are reported as mean [standard deviation].

| Variable | ARToolKit | Retroreflective | Vuforia |
|---|---|---|---|
| **Viewing Angle** | | | |
| *0°* | 2.14 [0.64] | 3.87 [1.19] | 3.23 [1.20] |
| *15°* | 2.30 [0.48] | 4.62 [1.93] | 2.57 [0.69] |
| *30°* | 4.46 [6.16] | 3.94 [1.47] | 3.76 [1.92] |
| **Viewing Distance** | | | |
| *66 cm* | 2.26 [0.63] | 4.41 [1.74] | 1.82 [0.48] |
| *100 cm* | 2.14 [0.64] | 3.87 [1.19] | 3.23[1.20] |
| *133 cm* | 2.22 [0.64] | - | 5.37 [10.8] |
| **Marker Size** | | | |
| *50 mm* | 3.04 [1.36] | 14.49 [6.07] | 11.18 [24.6] |
| *75 mm* | 2.14 [0.64] | 3.87 [1.19] | 3.23 [1.20] |
| *100 mm* | 2.65 [1.17] | 6.39 [2.28] | 2.07 [0.69] |
| **Point Distance** | | | |
| *3.75 cm* | 2.86 [1.21] | 6.2 [1.50] | 3.37 [2.57] |
| *7.5 cm* | 2.14 [0.64] | 3.87 [1.19] | 3.23[1.20] |
| *15 cm* | 3.79 [1.65] | 19.58 [8.97] | 1.85 [0.64] |
| **Point Distribution** | | | |
| *Equidistant* | 2.14 [0.64] | 3.87 [1.19] | 3.23 [1.20] |
| *Planar* | 1.26 [0.45] | 4.9 [1.84] | 3.13 [1.18] |
| *Random* | 1.98 [1.08] | 6.88 [1.72] | 3.16 [1.53] |
| **Amount of Points** | | | |
| *4* | 2.14 [0.64] | 3.87 [1.19] | 3.23 [1.20] |
| *8* | 1.37 [0.26] | 5.23 [1.33] | 3.04 [2.86] |
| **Refinement Approaches** | | | |
| *Point-based* | 2.14 [0.64] | 3.87 [1.19] | 3.23 [1.20] |
| *Point-based + Path Points* | 2.06 [0.44] | 3.57 [1.11] | 3.12 [0.87] |
| *ICP + Path Points* | 1.76 [0.40] | 3.2 [1.23] | 2.82 [0.86] |

Table A.2: Summary of the ANOVAs' results ($\alpha < .05$). Test statistic $F$ and effect size $\eta^2$ are reported. For the factor *Viewing Distance* the test statistic $\chi^2$ and the effect size $R^2$ is reported.

| Effect type | Factor | F | p | Sig. | $\eta^2$ | Effect | Fig. |
|---|---|---|---|---|---|---|---|
| **Tracking Accuracy** | | | | | | | |
| Main | *Tracking* | 259 | <0.001 | * | -0.054 | Medium | 5.3 |
| **Viewing Angle** | | | | | | | |
| Main | *Tracking* | 6.33 | <0.001 | * | 0.045 | Small | |
| | *Angle* | 3.43 | 0.068 | | - | - | 5.4 |
| Interaction | *Tracking* × *Angle* | 3.30 | 0.057 | | - | - | 5.4 |
| **Viewing Distance** | | | | | | | |
| Main | *Tracking* | 6.71 | 0.034 | * | 0.085 | Small | |
| | *Distance* | 0.01 | 0.994 | | - | - | 5.5 |
| Interaction | *Tracking* × *Distance* | 6.49 | 0.090 | | - | - | 5.5 |
| **Marker Size** | | | | | | | |
| Main | *Tracking* | 8.60 | 0.004 | * | 0.081 | Medium | 5.6 |
| | *Size* | 14.49 | <0.001 | * | 0.109 | Medium | 5.6 |
| Interaction | *Tracking* × *Size* | 3.21 | 0.081 | | - | - | |
| **Point Distance** | | | | | | | |
| Main | *Tracking* | 114.60 | <0.001 | * | 0.513 | Big | |
| | *Distance* | 62.43 | <0.001 | * | 0.338 | Big | 5.7 |
| Interaction | *Tracking* × *Distance* | 62.38 | <0.001 | * | 0.514 | Big | 5.7 |
| **Point Distribution** | | | | | | | |
| Main | *Tracking* | 144.20 | <0.001 | * | 0.557 | Big | |
| | *Distribution* | 16.63 | <0.001 | * | 0.106 | Medium | 5.8 |
| Interaction | *Tracking* × *Distribution* | 12.72 | <0.001 | * | 0.196 | Big | 5.8 |
| **Amount of Points** | | | | | | | |
| Main | *Tracking* | 51.26 | <0.001 | * | 0.379 | Big | |
| | *Amount* | 0.51 | 0.483 | | - | - | 5.9 |
| Interaction | *Tracking* × *Amount* | 7.16 | 0.007 | * | 0.087 | Medium | 5.9 |
| **Refinement Approaches** | | | | | | | |
| Main | *Tracking* | 22.66 | <0.001 | * | 0.3224 | Big | 5.10 |
| | *Refinement* | 23.84 | <0.001 | * | 0.046 | Small | 5.10 |
| Interaction | *Tracking* × *Refinement* | 1.25 | 0.297 | | | - | |

### A.1.2 *Investigations on Sensory Substitution*

Table A.3: Summary of descriptive results for all dependent variables ($n = 21$). All techniques are composed of: Type, Position and Vibrotactile. All entries are in the format: mean value [standard deviation]. Adapted from [6].

| Technique | | | TCT [s] | AMax [N] | AAD [N] | Perceived Difficulty | Perceived Continuity |
|---|---|---|---|---|---|---|---|
| Arrow | FOV | Absent | 33.34 [16.90] | 13.26 [2.80] | 4.66 [1.52] | 1.71 [1.35] | 2.86 [1.59] |
| Arrow | FOV | Present | 33.62 [19.52] | 13.39 [2.52] | 4.43 [1.23] | 1.52 [1.25] | 3.43 [1.47] |
| Arrow | Probe | Absent | 31.61 [14.32] | 13.99 [1.92] | 4.59 [1.15] | 1.90 [1.09] | 2.95 [1.50] |
| Arrow | Probe | Present | 34.21 [22.33] | 13.19 [3.28] | 4.50 [1.05] | 1.62 [1.32] | 3.00 [1.41] |
| Arrow | Screen | Absent | 26.28 [11.67] | 14.09 [3.12] | 4.21 [1.24] | 1.29 [1.01] | 3.57 [1.16] |
| Arrow | Screen | Present | 36.83 [23.76] | 14.32 [2.45] | 4.96 [1.44] | 1.43 [0.93] | 3.90 [0.83] |
| Bar | FOV | Absent | 32.86 [16.60] | 13.11 [2.30] | 4.05 [1.05] | 2.14 [1.11] | 2.52 [1.21] |
| Bar | FOV | Present | 34.39 [18.76] | 13.31 [2.06] | 4.32 [1.47] | 1.81 [1.12] | 2.90 [1.41] |
| Bar | Probe | Absent | 32.91 [11.99] | 13.47 [2.83] | 4.16 [1.22] | 2.24 [1.14] | 2.76 [1.04] |
| Bar | Probe | Present | 36.73 [17.13] | 14.24 [2.92] | 4.39 [1.40] | 2.29 [1.23] | 2.90 [1.26] |
| Bar | Screen | Absent | 31.49 [16.54] | 15.08 [2.47] | 4.43 [1.60] | 1.57 [1.08] | 3.62 [1.16] |
| Bar | Screen | Present | 36.54 [12.24] | 12.54 [2.64] | 4.02 [1.35] | 1.43 [1.12] | 3.48 [1.12] |
| None | None | Present | 29.09 [13.89] | 14.45 [3.33] | 5.68 [1.34] | 2.81 [1.63] | 2.24 [1.64] |
| None | None | Absent | 28.54 [15.82] | 14.11 [4.02] | 5.44 [1.62] | 4.24 [1.00] | 0.76 [0.83] |

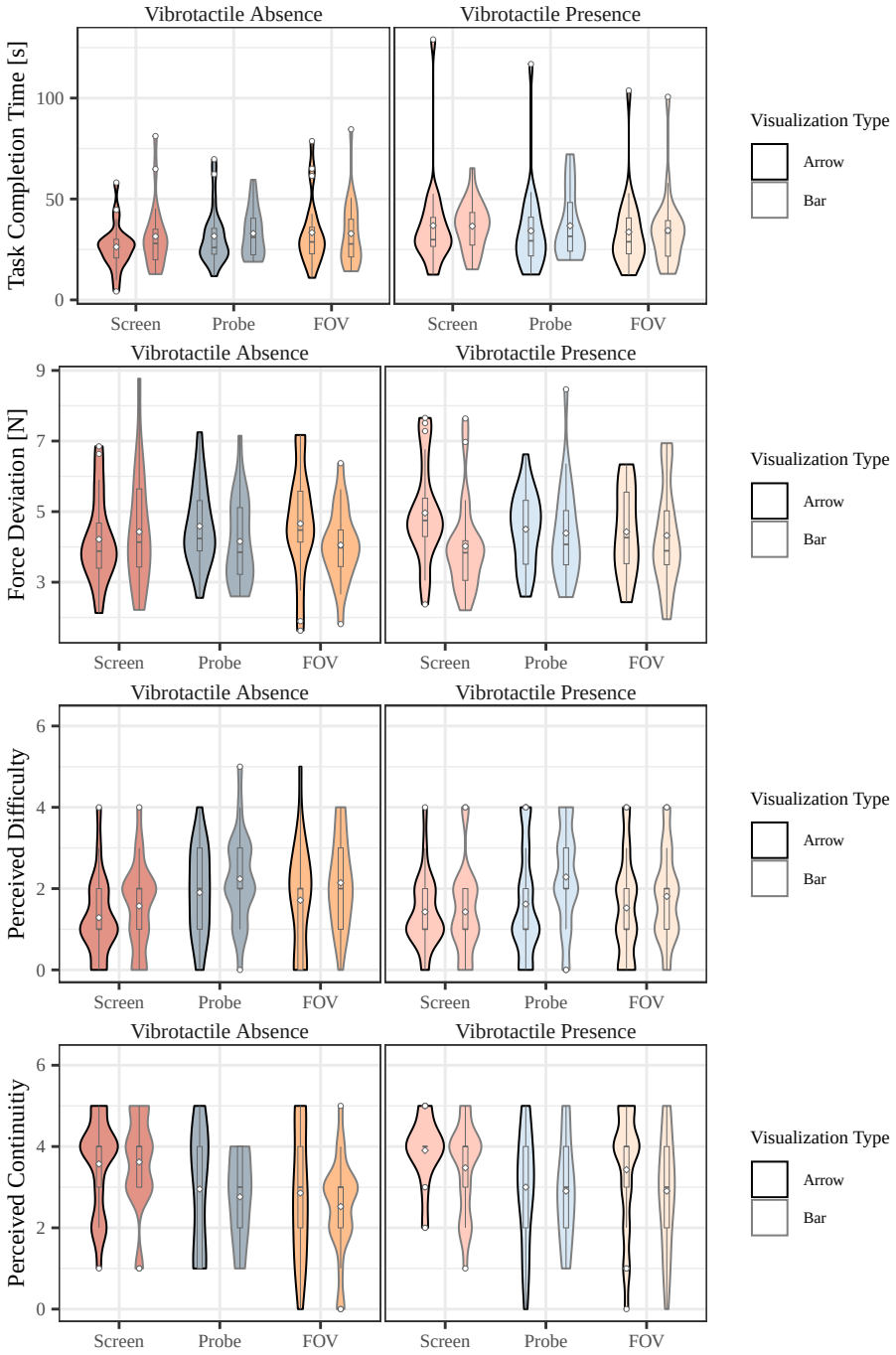



Figure A.1: Additional descriptive results. Means are indicated by ◇. Black outlines represent the *arrow* visualization and gray outlines the *bar* visualization. Adapted from [6].

Table A.4: Summary of all statistical analyses ($\alpha < .05$). For repeated measures ANOVAs (AoV), test statistic F and effect size $\eta^2$ are reported. For robust repeated measures ANOVAs (RAoV), test statistic Q and effect size $\delta_t$ are given. $\chi^2$ and Kendall's W are reported for Friedman tests (Frie) and V and r are given for Wilcoxon tests (Wilc), respectively. Adapted from [6].

| Variable / Effect type | Factor | Test | Test Statistic | p | Sig. | Effect Size |
|---|---|---|---|---|---|---|
| **TCT** | | | | | | |
| Main effects | *Type* | RAoV | 2.995 | 0.084 | | 0.180 |
| | *Position* | RAoV | 0.009 | 0.991 | | 0.022 |
| | *Vibrotactile* | RAoV | 8.869 | 0.003 | * | -0.275 |
| Two-way | *Type* × *Position* | RAoV | 0.195 | 0.823 | | - |
| | *Type* × *Vibrotactile* | RAoV | 0.053 | 0.818 | | - |
| | *Position* × *Vibrotactile* | RAoV | 1.682 | 0.186 | | - |
| Three-way | | RAoV | 0.015 | 0.985 | | - |
| Baseline | *Modality* | RAoV | 4.804 | 0.016 | * | -0.090 |
| **AMax** | | | | | | |
| Main effects | *Type* | AoV | 0.054 | 0.819 | | 0.000 |
| | *Position* | AoV | 2.479 | 0.097 | | 0.014 |
| | *Vibrotactile* | AoV | 1.461 | 0.241 | | 0.004 |
| Two-way | *Type* × *Position* | AoV | 0.339 | 0.647 | | 0.003 |
| | *Type* × *Vibrotactile* | AoV | 0.263 | 0.614 | | 0.001 |
| | *Position* × *Vibrotactile* | AoV | 2.455 | 0.099 | | 0.013 |
| Three-way | | AoV | 3.902 | 0.028 | * | 0.030 |
| Baseline | *Modality* | AoV | 0.605 | 0.614 | | 0.016 |
| **AAD** | | | | | | |
| Main effects | *Type* | RAoV | 8.068 | 0.005 | * | -0.348 |
| | *Position* | RAoV | 0.128 | 0.880 | | -0.079 |
| | *Vibrotactile* | RAoV | 0.162 | 0.688 | | -0.055 |
| Two-way | *Type* × *Position* | RAoV | 0.028 | 0.972 | | - |
| | *Type* × *Vibrotactile* | RAoV | 0.761 | 0.383 | | - |
| | *Position* × *Vibrotactile* | RAoV | 0.124 | 0.883 | | - |
| Three-way | | RAoV | 2.287 | 0.102 | | - |
| Baseline | *Modality* | AoV | 7.362 | <0.001 | * | 0.197 |
| **Perceived Difficulty** | | | | | | |
| Main effects | *Type* | Wilc | 143.500 | 0.053 | | 0.452 |
| | *Position* | Frie | 6.237 | 0.044 | * | 0.148 |
| | *Vibrotactile* | Wilc | 115.000 | 0.432 | | 0.175 |
| Two-way | *Type* × *Position* | RAoV | 0.418 | 0.658 | | - |
| | *Type* × *Vibrotactile* | RAoV | 0.048 | 0.826 | | - |
| | *Position* × *Vibrotactile* | RAoV | 0.214 | 0.807 | | - |
| Three-way | | RAoV | 0.747 | 0.474 | | - |
| Baseline | *Modality* | Frie | 30.429 | <0.001 | * | 0.483 |
| **Perceived Continuity** | | | | | | |
| Main effects | *Type* | Wilc | 63.500 | 0.073 | | 0.395 |
| | *Position* | Frie | 7.630 | 0.022 | * | 0.182 |
| | *Vibrotactile* | Wilc | 50.500 | 0.226 | | 0.305 |
| Two-way | *Type* × *Position* | RAoV | 0.166 | 0.847 | | - |
| | *Type* × *Vibrotactile* | RAoV | 1.032 | 0.310 | | - |
| | *Position* × *Vibrotactile* | RAoV | 0.929 | 0.395 | | - |
| Three-way | | RAoV | 0.112 | 0.894 | | - |
| Baseline | *Modality* | Frie | 40.215 | <0.001 | * | 0.638 |

### A.1.3 *Investigation on Sensor Placement and Control Modes*

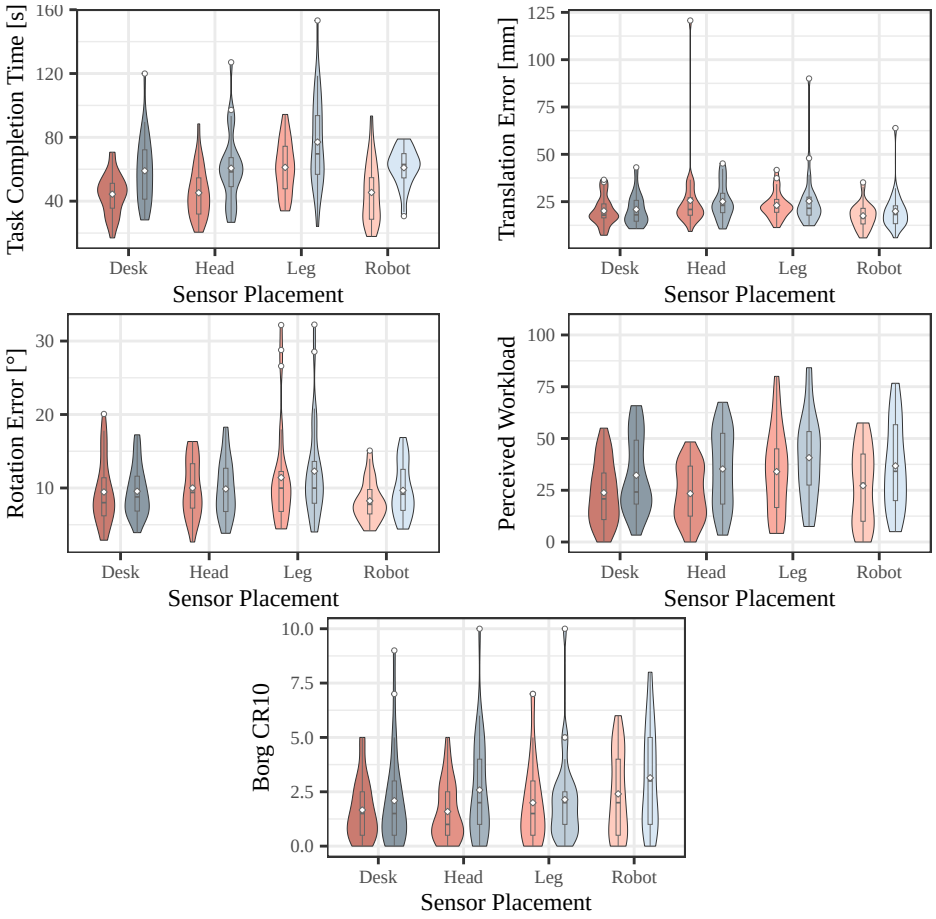



Figure A.2: Descriptive results of the dependent variables. ■ show *position control*, while ■ show *rate control*. Means are indicated by ◇.

### A.1.4 *Investigations on XRP for Contact Intensive Tasks*

Table A.5: Summary of descriptive results for all dependent variables. Adapted from [1].

| Environment | Visual Feedback | Force Automation | TCT [s] | AAD [N] | TLX | Trust |
|---|---|---|---|---|---|---|
| Physical | Absent | Manual | 27.75 [7.32] | 5.62 [1.05] | 35.48 [14.57] | 3.87 [0.38] |
| Physical | Absent | Automated | 25.40 [10.36] | 0.24 [0.09] | 23.43 [13.23] | 3.93 [0.51] |
| Physical | Present | Manual | 28.90 [8.54] | 4.44 [1.10] | 39.13 [16.99] | 3.84 [0.60] |
| Physical | Present | Automated | 24.41 [7.06] | 0.26 [0.11] | 21.67 [12.23] | 3.91 [0.65] |
| Simulated | Absent | Manual | 30.35 [11.50] | 6.46 [1.36] | 42.60 [18.05] | 3.42 [0.69] |
| Simulated | Absent | Automated | 23.83 [8.01] | 0.32 [0.03] | 24.26 [13.28] | 3.67 [0.65] |
| Simulated | Present | Manual | 31.11 [11.66] | 4.92 [1.34] | 44.29 [15.38] | 3.64 [0.46] |
| Simulated | Present | Automated | 25.57 [7.95] | 0.33 [0.04] | 24.87 [12.38] | 3.70 [0.61] |

Table A.6: Summary of all statistical analyses ($\alpha < .05$). For repeated measures ANOVAs (AoV), test statistic F and effect size $\eta^2$ are reported. For robust repeated measures ANOVAs (RAoV), test statistic Q and effect size $\delta_t$ are given. Adapted from [1].

| Variable / Effect type | Factor | H | Test | Test Statistic | p | Sig. | Effect Size |
|---|---|---|---|---|---|---|---|
| TCT | | | | | | | |
| Main | *Environment* | 3.2 | RAoV | 1.945 | 0.163 | | - |
| | *Feedback* | 1.3 | RAoV | 0.980 | 0.323 | | - |
| | *Automation* | 2.2 | RAoV | 19.309 | <0.001 | * | 0.561 |
| Two-way | *Environment* × *Feedback* | | RAoV | 0.150 | 0.699 | | - |
| | *Environment* × *Automation* | | RAoV | 0.625 | 0.431 | | - |
| | *Feedback* × *Automation* | | RAoV | 0.010 | 0.921 | | - |
| Three-way | | | RAoV | 0.015 | 0.796 | | - |
| AAD | | | | | | | |
| Main | *Environment* | 3.1 | RAoV | 25.126 | <0.001 | * | -0.063 |
| | *Feedback* | 1.1 | RAoV | 56.943 | <0.001 | * | -0.168 |
| | *Automation* | 2.1 | RAoV | 1364.489 | <0.001 | * | 4.896 |
| Two-way | *Environment* × *Feedback* | | RAoV | 2.571 | 0.109 | | - |
| | *Environment* × *Automation* | | RAoV | 17.775 | <0.001 | * | 0.003 |
| | *Feedback* × *Automation* | | RAoV | 62.236 | <0.001 | * | 0.02 |
| Three-way | | | RAoV | 2.822 | 0.093 | | - |

Table A.6 (continued): Summary of all statistical analyses ($\alpha < .05$). For repeated measures ANOVAs (AoV), test statistic F and effect size $\eta^2$ are reported. For robust repeated measures ANOVAs (RAoV), test statistic Q and effect size $\delta_t$ are given. Adapted from [1].

| **Variable / Effect type** | **Factor** | H | **Test** | **Test Statistic** | **p** | **Sig.** | **Effect Size** |
|---|---|---|---|---|---|---|---|
| **TLX** | | | | | | | |
| Main | *Environment* | 3.3 | AoV | 7.036 | 0.014 | * | 0.01 |
| | *Feedback* | 1.4 | AoV | 0.822 | 0.373 | | - |
| | *Automation* | 2.3 | AoV | 59.71 | <0.001 | * | 0.25 |
| Two-way | *Environment* × *Feedback* | | AoV | 0.007 | 0.932 | | - |
| | *Environment* × *Automation* | | AoV | 2.279 | 0.144 | | - |
| | *Feedback* × *Automation* | | AoV | 2.035 | 0.166 | | - |
| Three-way | | | AoV | 1.005 | 0.326 | | - |
| **Trust** | | | | | | | |
| Main | *Environment* | 3.4 | AoV | 10.538 | 0.003 | * | 0.06 |
| | *Feedback* | 1.2 | AoV | 0.556 | 0.463 | | - |
| | *Automation* | 2.4 | AoV | 6.445 | 0.018 | * | 0.009 |
| Two-way | *Environment* × *Feedback* | | AoV | 3.137 | 0.089 | | - |
| | *Environment* × *Automation* | | AoV | 0.55 | 0.465 | | - |
| | *Feedback* × *Automation* | | AoV | 0.917 | 0.347 | | - |
| Three-way | | | AoV | 1.876 | 0.183 | | - |

# PUBLICATIONS

This thesis is based on the following seven publications. Some illustrations and phrasings have previously appeared in these publications.


[1] Tonia Mielke, Mareen Allgaier, Christian Hansen, and Florian Heinrich. "Extended Reality Check: Evaluating XR Prototyping for Human-Robot Interaction in Contact-Intensive Tasks." In: *Transactions on Visualization and Computer Graphics (TVCG)* 31.11 (Oct. 2025), pp. 10035–10044. DOI: 10.1109/TVCG.2025.3616753 (cit. on pp. 4, 130, 148–150, 156–159, 181, 182).

[2] Tonia Mielke, Mareen Allgaier, Danny Schott, Christian Hansen, and Florian Heinrich. "Virtual Studies, Real Results? Assessing the Impact of Virtualization on Human-Robot Interaction." In: *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, Apr. 2025, pp. 1–8. DOI: 10.1145/3706599.3719724 (cit. on pp. 4, 130, 135, 137, 141, 143, 144).

[3] Tonia Mielke, Marilena Georgiades, Oliver S. Großer, Maciej Pech, Christian Hansen, and Florian Heinrich. "Enhancing Gesture-Based Human-Robot Interaction: Investigating the Role of Force Automation." In: *Proceedings of the 21st ACM IEEE International Conference on Human-Robot Interaction (HRI '26)*. Edinburgh, Scotland, UK: Association for Computing Machinery, 2026, pp. 1–10. DOI: 10.1145/3757279.3785568 (cit. on pp. 4, 70, 80–83, 85–90).

[4] Tonia Mielke, Florian Heinrich, and Christian Hansen. "Enhancing AR-to-Robot Registration Accuracy: A Comparative Study of Marker Detection Algorithms and Registration Parameters." In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 19–23. DOI: 10.1109/ICRA55743.2025.11128039 (cit. on pp. 4, 36, 52, 53, 55, 57–62).

[5] Tonia Mielke, Florian Heinrich, and Christian Hansen. "Gesturing Towards Efficient Robot Control: Exploring Sensor Placement and Control Modes for Mid-Air Human-Robot Interaction." In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 19–23. DOI: 10.1109/ICRA55743.2025.11127519 (cit. on pp. 4, 70, 111, 114, 116, 118, 121–124).

[6] Tonia Mielke, Florian Heinrich, and Christian Hansen. "SensARy Substitution: Augmented Reality Techniques to Enhance Force Perception in Touchless Robot Control." In: *Transactions on Visualization and Computer Graphics (TVCG)* 31.5 (Mar. 2025), pp. 3235–3244. DOI: 10.1109/TVCG.2025.3549856 (cit. on pp. 4, 70, 93–95, 102–106, 177–179).

[7] Tonia Mielke, Fabian Joeres, Danny Schott, and Christian Hansen. "Interactive Registration Methods for Augmented Reality in Robotics: A Comparative Evaluation." In: *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2023, pp. 501–506. DOI: 10.1109/ISMAR-Adjunct60411.2023.00109 (cit. on pp. 4, 36, 41–43, 48, 49).

In addition, two further articles were co-authored during this dissertation:

[8] Josefine Schreiter, Tonia Mielke, Marilena Georgiades, Maciej Pech, Christian Hansen, and Florian Heinrich. "Exploring Interaction Concepts for the Manipulation of a Collaborative Robot: A Comparative Study." In: *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 04–06. DOI: 10.1109/HRI61500.2025.10974113 (cit. on pp. 1, 2, 28, 72, 73).

[9] Josefine Schreiter, Tonia Mielke, Danny Schott, Maximilian Thormann, Jazan Omari, Maciej Pech, and Christian Hansen. "A multimodal user interface for touchless control of robotic ultrasound." In: *Int. J. CARS* 18.8 (Aug. 2023), pp. 1429–1436. ISSN: 1861-6429. DOI: 10.1007/s11548-022-02810-0 (cit. on pp. 28, 29, 72, 73, 75, 79, 80, 173).

[10] Fikri M. Abu-Zidan, Ashraf F. Hefny, and Peter Corr. "Clinical ultrasound physics." In: *Journal of Emergencies, Trauma, and Shock* 4.4 (Oct. 2011), p. 501. ISSN: 0974-2700. DOI: 10.4103/0974-2700.86646 (cit. on p. 20).

[11] George Adamides, Georgios Christou, Christos Katsanos, Michalis Xenos, and Thanasis Hadzilacos. "Usability Guidelines for the Design of Robot Teleoperation: A Taxonomy." In: *IEEE Trans. Hum.-Mach. Syst.* 45.2 (Dec. 2014), pp. 256–262. DOI: 10.1109/THMS.2014.2371048 (cit. on p. 22).

[12] Haley Adams, Jeanine Stefanucci, Sarah Creem-Regehr, and Bobby Bodenheimer. "Depth Perception in Augmented Reality: The Effects of Display, Shadow, and Position." In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. New York, USA: IEEE, 2022, pp. 12–16. DOI: 10.1109/VR51125.2022.00101 (cit. on p. 145).

[13] Sung Joon Ahn, Wolfgang Rauh, and Hans-Jürgen Warnecke. "Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola." In: *Pattern Recognit.* 34.12 (Dec. 2001), pp. 2283–2303. ISSN: 0031-3203. DOI: 10.1016/S0031-3203(00)00152-7 (cit. on p. 17).

[14] Takintope Akinbiyi, Carol E. Reiley, Sunipa Saha, Darius Burschka, Christopher J. Hasser, David D. Yuh, and Allison M. Okamura. "Dynamic Augmented Reality for Sensory Substitution in Robot-Assisted Surgical Systems." In: *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2006, pp. 2006–03. ISBN: 978-1-4244-0032. DOI: 10.1109/IEMBS.2006.259707 (cit. on pp. 77, 78).

[15] Gorkem Anil Al, Pedro Estrela, and Uriel Martinez-Hernandez. "Towards an intuitive human-robot interaction based on hand gesture recognition and proximity sensors." In: *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2020, pp. 14–16. DOI: 10.1109/MFI49285.2020.9235264 (cit. on pp. 111, 113, 114).

[16] James Algina, H. J. Keselman, and Randall D. Penfield. "An Alternative to Cohen's Standardized Mean Difference Effect Size: A Robust Parameter and Confidence Interval in the Two Independent Groups Case." In: *Psychological Methods* 10.3 (2005), pp. 317–328. DOI: 10.1037/1082-989x.10.3.317 (cit. on pp. 26, 46, 85, 100, 119, 140, 154).

[17] Daniel R. Allen, Terry M. Peters, and Elvis C. S. Chen. "A co-calibration framework for the accuracy assessment of vision-based tracking systems." In: *Proceedings Volume 12034, Medical Imaging 2022:*

*Image-Guided Procedures, Robotic Interventions, and Modeling*. Vol. 12034. SPIE, Apr. 2022, pp. 598–608. DOI: 10.1117/12.2606815 (cit. on p. 50).

[18] Mareen Allgaier, Florentine Huettl, Laura Isabel Hanke, Hauke Lang, Tobias Huber, Bernhard Preim, Sylvia Saalfeld, and Christian Hansen. "LiVRSono - Virtual Reality Training with Haptics for Intraoperative Ultrasound." In: *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. New York, USA: IEEE, Oct. 2023, 980–989. DOI: 10.1109/ismar59233.2023.00114 (cit. on pp. 137, 140).

[19] Rasmus S. Andersen, Ole Madsen, Thomas B. Moeslund, and Heni Ben Amor. "Projecting robot intentions into human environments." In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 26–31. DOI: 10.1109/ROMAN.2016.7745145 (cit. on p. 37).

[20] Ainhoa Apraiz, Ganix Lasa, and Maitane Mazmela. "Evaluation of User Experience in Human–Robot Interaction: A Systematic Literature Review." In: *Int. J. Social Rob.* 15.2 (Feb. 2023), pp. 187–210. ISSN: 1875-4805. DOI: 10.1007/s12369-022-00957-z (cit. on pp. 22, 23, 172).

[21] Carmelo Ardito, Paolo Buono, Maria Francesca Costabile, and Giuseppe Desolda. "Interaction with Large Displays: A Survey." In: *ACM Comput. Surv.* 47.3 (Feb. 2015), pp. 1–38. ISSN: 0360-0300. DOI: 10.1145/2682623 (cit. on p. 71).

[22] Stephanie Arevalo Arboleda, Franziska Rücker, Tim Dierks, and Jens Gerken. "Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues." In: *ACM Conferences*. New York, NY, USA: Association for Computing Machinery, May 2021, pp. 1–14. DOI: 10.1145/3411764.3445398 (cit. on p. 37).

[23] K Somani Arun, Thomas S Huang, and Steven D Blostein. "Least-squares fitting of two 3-D point sets." In: *IEEE Transactions on pattern analysis and machine intelligence* 5 (1987), pp. 698–700 (cit. on pp. 18, 31, 53).

[24] Angelica I. Aviles-Rivero, Samar M. Alsaleh, John Philbeck, Stella P. Raventos, Naji Younes, James K. Hahn, and Alicia Casals. "Sensory Substitution for Force Feedback Recovery: A Perception Experimental Study." In: *ACM Trans. Appl. Percept.* 15.3 (Apr. 2018), pp. 1–19. ISSN: 1544-3558. DOI: 10.1145/3176642 (cit. on pp. 4, 77–79, 93, 94).

[25] Ronald T. Azuma. "A Survey of Augmented Reality." In: *Presence: Teleoperators and Virtual Environments* 6.4 (Aug. 1997), pp. 355–385. DOI: 10.1162/pres.1997.6.4.355 (cit. on pp. 13, 37).

[26] Wilma A Bainbridge, Justin W Hart, Elizabeth S Kim, and Brian Scassellati. "The benefits of interactions with physically present robots over video-displayed agents." In: *International Journal of Social Robotics* 3 (2011), pp. 41–52. DOI: 10.1007/s12369-010-0082-7 (cit. on p. 132).

[27] X. Bao, S. Wang, R. Housden, J. Hajnal, and K. Rhode. "A Constant-Force End-Effector With Online Force Adjustment for Robotic Ultrasonography." In: *IEEE Rob. Autom. Lett.* 6.2 (Feb. 2021), pp. 2547–2554. DOI: 10.1109/LRA.2021.3061329 (cit. on p. 23).

[28] Christoph Bartneck, Tony Belpaeme, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Sabanovic. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. en. CRC Press, Mar. 2005. DOI: 10.1201/b10624 (cit. on pp. 14, 15).

[29] D. Bassily, C. Georgoulas, J. Guettler, T. Linner, and T. Bock. "Intuitive and Adaptive Robotic Arm Manipulation using the Leap Motion Controller." In: *ISR/Robotik 2014; 41st International Symposium on Robotics*. Berlin, Germany: VDE, pp. 02–03. ISBN: 978-3-8007-3601-0 (cit. on pp. 111, 114).

[30] Jenay M. Beer, Arthur D. Fisk, and Wendy A. Rogers. "Toward a framework for levels of robot autonomy in human-robot interaction." In: *Journal of human-robot interaction* 3.2 (July 2014), p. 74. DOI: 10.5898/JHRI.3.2.Beer (cit. on pp. 8, 9).

[31] Michael Van den Bergh, Daniel Carton, Roderick De Nijs, Nikos Mitsou, Christian Landsiedel, and Kolja Kuehnlenz. "Real-time 3D hand gesture interaction with a robot for understanding directions from humans." In: *2011 RO-MAN*. IEEE, pp. 2011–03. DOI: 10.1109/ROMAN.2011.6005195 (cit. on p. 112).

[32] Sylvain Bernhardt, Stéphane A. Nicolau, Luc Soler, and Christophe Doignon. "The status of augmented reality in laparoscopic surgery as of 2016." In: *Med. Image Anal.* 37 (Apr. 2017), pp. 66–90. ISSN: 1361-8415. DOI: 10.1016/j.media.2017.01.007 (cit. on p. 18).

[33] P. J. Besl and Neil D. McKay. "A method for registration of 3-D shapes." In: *IEEE Trans. Pattern Anal. Mach. Intell.* 14.2 (Aug. 2002), pp. 239–256. DOI: 10.1109/34.121791 (cit. on pp. 18, 54).

[34] Cindy L. Bethel and Robin R. Murphy. "Review of Human Studies Methods in HRI and Recommendations." In: *Int. J. Social Rob.* 2.4 (Dec. 2010), pp. 347–359. ISSN: 1875-4805. DOI: 10.1007/s12369-010-0064-9 (cit. on p. 22).

[35] Geoffrey Biggs and Bruce MacDonald. "A survey of robot programming systems." In: *Proceedings of the Australasian conference on robotics and automation*. Vol. 1. 2003, pp. 1–3 (cit. on p. 71).

[36] M. Billinghurst, J. Bowskill, N. Dyer, and J. Morphett. "An evaluation of wearable information spaces." In: *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No.98CB36180)*. IEEE, pp. 14–18. ISBN: 978-0-8186-8362. DOI: 10.1109/VRAIS.1998.658418 (cit. on pp. 78, 94).

[37] W. Birkfellner, F. Watzinger, F. Wanschitz, R. Ewers, and H. Bergmann. "Calibration of tracking systems in a surgical environment." In: *IEEE Trans. Med. Imaging* 17.5 (Aug. 2002), pp. 737–742. DOI: 10.1109/42.736028 (cit. on p. 15).

[38] Gary Bishop, B. Danette Allen, and Greg Welch. *Tracking: Beyond 15 minutes of thought*. Chapel Hill, NC, USA, 2001 (cit. on p. 16).

[39] David Black and Septimiu Salcudean. "Human-as-a-robot performance in mixed reality teleultrasound." In: *Int. J. CARS* 18.10 (Oct. 2023), pp. 1811–1818. ISSN: 1861-6429. DOI: 10.1007/s11548-023-02896-0 (cit. on pp. 77, 93).

[40] Jose-Luis Blanco. "A tutorial on se (3) transformation parameterizations and on-manifold optimization." In: *University of Malaga, Tech. Rep* 3.6 (2010), pp. 1–58 (cit. on p. 31).

[41] Sebastian Blankemeyer, Rolf Wiemann, Lukas Posniak, Christoph Pregizer, and Annika Raatz. "Intuitive Robot Programming Using Augmented Reality." In: *Procedia CIRP* 76 (Jan. 2018), pp. 155–160. ISSN: 2212-8271. DOI: 10.1016/j.procir.2018.02.028 (cit. on pp. 38, 39, 41).

[42] Richard A. Bolt. ""Put-that-there": Voice and gesture at the graphics interface." In: vol. 14. 3. Association for Computing Machinery, July 1980, pp. 262–270. DOI: 10.1145/965105.807503 (cit. on p. 11).

[43] Gunnar Borg. *Borg's Perceived Exertion And Pain Scales*. July 1998. ISBN: 0-88011-623-4 (cit. on pp. 24, 118).

[44] Leonardo Borgioli et al. "Sensory Glove-Based Surgical Robot User Interface." In: *arXiv* (Mar. 2024). DOI: 10.48550/arXiv.2403.13941. eprint: 2403.13941 (cit. on pp. 2, 71–73, 111).

[45] Sebastian Boring, Marko Jurmu, and Andreas Butz. "Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays." In: *ACM Other conferences*. New York, NY, USA: Association for Computing Machinery, Nov. 2009, pp. 161–168. DOI: 10.1145/1738826.1738853 (cit. on p. 111).

[46] Pierre Boulanger. *Applications of Augmented Reality - Current State of the Art*. IntechOpen, 2024. ISBN: 978-1-83769335-1 (cit. on p. 18).

[47] Doug A Bowman, Ernst Kruijff, Joseph J LaViola, and Ivan Poupyrev. *3D User Interfaces*. en. Boston, MA: Addison Wesley, July 2004. ISBN: 978-0-20175867-2 (cit. on pp. 11, 12, 14).

[48] Doug A Bowman, Cheryl Stinson, Eric D Ragan, Siroberto Scerbo, Tobias Höllerer, Cha Lee, Ryan P McMahan, and Regis Kopper. "Evaluating effectiveness in virtual environments with MR simulation." In: *Interservice/Industry Training, Simulation, and Education Conference*. Vol. 4. New York, USA: National Training and Simulation Association (NTSA), 2012, p. 44 (cit. on p. 131).

[49] Eugenie Brasier, Olivier Chapuis, Nicolas Ferey, Jeanne Vezien, and Caroline Appert. "ARPads: Mid-air Indirect Input for Augmented Reality." In: *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2020, pp. 09–13. DOI: 10.1109/ISMAR50242.2020.00060 (cit. on pp. 112–114, 124).

[50] C. Breazeal. "Social interactions in HRI: the robot view." In: *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 34.2 (May 2004), pp. 181–186. DOI: 10.1109/TSMCC.2004.826268 (cit. on p. 173).

[51] L Britt Wilson, Richard A Hoppmann, Floyd E Bell, and Victor V Rao. *Understanding physiology with ultrasound*. en. Cham, Switzerland: Springer Nature, Jan. 2023. ISBN: 978-1-0716-1862-2. DOI: https://doi.org/10.1007/978-1-0716-1863-9 (cit. on p. 20).

[52] Torgny Brogårdh. "Present and future robot control development—An industrial perspective." In: *Annual Reviews in Control* 31.1 (Jan. 2007), pp. 69–79. ISSN: 1367-5788. DOI: 10.1016/j.arcontrol.2007.01.002 (cit. on p. 71).

[53] M. R. Burcher, J. A. Noble, Lianghao Man, and M. Gooding. "A system for simultaneously measuring contact force, ultrasound, and position information for use in force-based correction of freehand scanning." In: *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 52.8 (Aug. 2005), pp. 1330–1342. DOI: 10.1109/TUFFC.2005.1509791 (cit. on pp. 4, 20).

[54] Yi Cai, Yi Wang, and Morice Burnett. "Using augmented reality to build digital twin for reconfigurable additive manufacturing system." In: *J. Manuf. Syst.* 56 (), pp. 598–604. ISSN: 0278-6125. DOI: 10.1016/j.jmsy.2020.04.005 (cit. on pp. 38–41).

[55] Kelly Caine. "Local Standards for Sample Size at CHI." In: *ACM Conferences*. New York, NY, USA: Association for Computing Machinery, May 2016, pp. 981–992. DOI: 10.1145/2858036.2858498 (cit. on p. 172).

[56] Giulio Campagna, Christoph Frommel, Tobias Haase, Alberto Gottardi, Enrico Villagrossi, Dimitrios Chrysostomou, and Matthias Rehm. "Fostering Trust through Gesture and Voice-Controlled Robot Trajectories in Industrial Human-Robot Collaboration." In: *2025 IEEE International Conference on Robotics and Automation*. IEEE. 2025 (cit. on p. 112).

[57] Julie Carmigniani, Borko Furht, Marco Anisetti, Paolo Ceravolo, Ernesto Damiani, and Misa Ivkovic. "Augmented reality technologies, systems and applications." In: *Multimed. Tools Appl.* 51.1 (Jan. 2011), pp. 341–377. ISSN: 1573-7721. DOI: 10.1007/s11042-010-0660-6 (cit. on p. 14).

[58] Jessica R Cauchard, Mohamed Khamis, Jérémie Garcia, Matjaž Kljun, and Anke M Brock. "Toward a roadmap for human-drone interaction." en. In: *Interactions* 28.2 (Mar. 2021), pp. 76–81 (cit. on p. 173).

[59] Sonia Mary Chacko, Armando Granado, and Vikram Kapila. "An Augmented Reality Framework for Robotic Tool-path Teaching." In: *Procedia CIRP* 93 (Jan. 2020), pp. 1218–1223. ISSN: 2212-8271. DOI: 10.1016/j.procir.2020.03.143 (cit. on p. 37).

[60] Wesley P. Chan, Camilo Perez Quintero, Matthew K. X. J. Pan, Maram Sakr, H. F. Machiel Van der Loos, and Elizabeth Croft. "A Multimodal System Using Augmented Reality, Gestures, and Tactile Feedback for Robot Trajectory Programming and Execution." In: *Virtual Reality*. Upper Saddle River, NJ, USA: River Publishers, Sept. 2022, pp. 142–158. ISBN: 978-1-00334000-3. DOI: 10.1201/9781003340003-8 (cit. on pp. 75–79, 93, 95, 96, 107).

[61] An Chi Chen, Muhammad Hadi, Peter Kazanzides, and Ehsan Azimi. "Mixed Reality Based Teleoperation of Surgical Robotics." In: *2023 International Symposium on Medical Robotics (ISMR)*. IEEE, 2023, pp. 19–21. DOI: 10.1109/ISMR57123.2023.10130178 (cit. on pp. 28, 37–39, 72, 73, 111, 112, 114, 131).

[62] Chang Chen, Liang Chen, Xuefeng Zhou, and Wu Yan. "Controlling a robot using leap motion." In: *Published in: 2017 2nd International Conference on Robotics and Automation Engineering (ICRAE)* (), pp. 29–31. DOI: 10.1109/ICRAE.2017.8291351 (cit. on pp. 2, 111, 114).

[63] Ian Yen-Hung Chen, Bruce MacDonald, Burkhard Wünsche, Geoffrey Biggs, and Tetsuo Kotoku. "Analysing mixed reality simulation for industrial applications: A case study in the development of a robotic screw remover system." In: *Simulation, Modeling, and Programming for Autonomous Robots*. Lecture notes in computer science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 350–361. DOI: 10.1007/978-3-642-17319-6_33 (cit. on pp. 37, 131–135, 145).

[64] Shihang Chen, Fang Wang, Yanping Lin, Qiusheng Shi, and Yanli Wang. "Ultrasound-guided needle insertion robotic system for percutaneous puncture." In: *Int. J. CARS* 16.3 (Mar. 2021), pp. 475–484. ISSN: 1861-6429. DOI: 10.1007/s11548-020-02300-1 (cit. on pp. 21, 80, 83).

[65] Soon Hau Chua, Simon T. Perrault, Denys J. C. Matthies, and Shengdong Zhao. "Positioning Glass: Investigating Display Positions of Monocular Optical See-Through Head-Mounted Display." In: *ChineseCHI2016: Proceedings of the Fourth International Symposium on Chinese CHI*. New York, NY, USA: Association for Computing Machinery, May 2016, pp. 1–6. ISBN: 978-1-45034760-0. DOI: 10.1145/2948708.2948713 (cit. on pp. 79, 95).

[66] A. Cirillo, F. Ficuciello, C. Natale, S. Pirozzi, and L. Villani. "A Conformable Force/Tactile Skin for Physical Human–Robot Interaction." In: *IEEE Rob. Autom. Lett.* 1.1 (Dec. 2015), pp. 41–48. DOI: 10.1109/LRA.2015.2505061 (cit. on pp. 81, 91).

[67] Andre Cleaver, Darren Tang, Victoria Chen, and Jivko Sinapov. "HAVEN: A uNity-based virtual robot environment to showcase HRI-based Augmented Reality." In: (2020). DOI: 10.48550/arXiv.2011.03464. eprint: 2011.03464 (cs.RO) (cit. on p. 131).

[68] Yunus Emre Cogurcu, James A Douthwaite, and Steve Maddock. "A comparative study of safety zone visualisations for virtual and physical robot arms using augmented reality." In: *Computers* 12.4 (2023), p. 75. DOI: 10.3390/computers12040075 (cit. on pp. 131–135, 145).

[69] Jacob Cohen. *The earth is round (p<. 05).* 1994. DOI: 10.1037/0003-066X.49.12.997 (cit. on p. 26).

[70] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Andover, England, UK: Taylor & Francis, May 2013. ISBN: 978-0-20377158-7. DOI: 10.4324/9780203771587 (cit. on p. 26).

[71] Douglas M. Considine and Glenn D. Considine. "Robot Technology Fundamentals." In: *Standard Handbook of Industrial Automation*. Boston, MA, USA: Springer, Boston, MA, 1986, pp. 262–320. ISBN: 978-1-4613-1963-4. DOI: 10.1007/978-1-4613-1963-4_17 (cit. on p. 7).

[72] Enrique Coronado, Shunki Itadera, and Ixchel G. Ramirez-Alpizar. "Integrating Virtual, Mixed, and Augmented Reality to Human–Robot Interaction Applications Using Game Engines: A Brief Review of Accessible Software Tools and Frameworks." In: *Appl. Sci.* 13.3 (Jan. 2023), p. 1292. ISSN: 2076-3417. DOI: 10.3390/app13031292 (cit. on p. 29).

[73] Enrique Coronado, Jessica Villalobos, Barbara Bruno, and Fulvio Mastrogiovanni. "Gesture-based robot control: Design challenges and evaluation with humans." In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2017–03. DOI: 10.1109/ICRA.2017.7989321 (cit. on p. 113).

[74] Gabriel M. Costa, Marcelo R. Petry, João G. Martins, and António Paulo G. M. Moreira. "Assessment of Multiple Fiducial Marker Trackers on Hololens 2." In: *IEEE Access* 12 (Jan. 2024), pp. 14211–14226. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3356722 (cit. on pp. 39, 52, 63).

[75] Christopher Michael Jason Cox, Ben Hicks, James Gopsill, and Chris Snider. "From Haptic Interaction to Design Insight: An Empirical Comparison of Commertial Hand-Tracking Technology." In: *Proceedings of the Design Society* 3 (July 2023), pp. 1965–1974. ISSN: 2732-527X. DOI: 10.1017/pds.2023.197 (cit. on p. 29).

[76] Seán Cronin and Gavin Doherty. "Touchless computer interfaces in hospitals: A review." In: *Health Informatics J.* 25.4 (Feb. 2018), pp. 1325–1342. ISSN: 1460-4582. DOI: 10.1177/1460458217748342 (cit. on p. 13).

[77] Andrew Crossan, Mark McGill, Stephen Brewster, and Roderick Murray-Smith. "Head tilting for interaction in mobile contexts." In: *ACM Other conferences*. New York, NY, USA: Association for Computing Machinery, Sept. 2009, pp. 1–10. DOI: 10.1145/1613858.1613866 (cit. on p. 113).

[78] Carolina Cruz-Neira, Daniel J. Sandin, Thomas A. DeFanti, Robert V. Kenyon, and John C. Hart. "The CAVE: audio visual experience automatic virtual environment." In: *Commun. ACM* 35.6 (June 1992), pp. 64–72. ISSN: 0001-0782. DOI: 10.1145/129888.129892 (cit. on p. 15).

[79] Graham Currell. *Scientific Data Analysis*. Oxford, England, UK: Oxford University Press. ISBN: 978-0-19871254-1 (cit. on p. 25).

[80] Jeremy Dalton. *Reality Check: How Immersive Technologies Can Transform Your Business*. Philadelphia, PA, USA: Kogan Page, Jan. 2021. ISBN: 978-1-78966634-2. URL: https://books.google.de/books/about/Reality_Check.html?id=gzQPEAAAQBAJ&redir_esc=y (cit. on p. 13).

[81] E. R. Davies. *Machine Vision*. Morgan Kaufmann, 2005. ISBN: 978-0-12-206093-9. DOI: 10.1016/B978-0-12-206093-9.X5000-X (cit. on p. 17).

[82] Alessandro De Franco, Edoardo Lamon, Pietro Balatti, Elena De Momi, and Arash Ajoudani. "An Intuitive Augmented Reality Interface for Task Scheduling, Monitoring, and Work Performance Improvement in Human-Robot Collaboration." In: *2019 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*. IEEE, 2019, pp. 03–05. DOI: 10.1109/IWOBI47054.2019.9114472 (cit. on pp. 4, 23, 77–79, 93, 95, 108).

[83] Federico De Lorenzis, Marina Nadalin, Massimo Migliorini, Francesca Scarrone, Fabrizio Lamberti, and Jacopo Fiorenza. "A Study on Affordable Manipulation in Virtual Reality Simulations: Hand-Tracking versus Controller-Based Interaction." In: *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, pp. 25–29. DOI: 10.1109/VRW58643.2023.00027 (cit. on p. 29).

[84] Konstantinos G Derpanis. "Overview of the RANSAC Algorithm." In: *Image Rochester NY* 4.1 (2010), pp. 2–3 (cit. on p. 19).

[85] Manish Dhyani, Shawn C. Roll, Matthew W. Gilbertson, Melanie Orlowski, Arash Anvari, Qian Li, Brian Anthony, and Anthony E. Samir. "A pilot study to precisely quantify forces applied by sonographers while scanning: A step toward reducing ergonomic injury." In: *Work* 58.2 (Jan. 2017), pp. 241–247. ISSN: 1051-9815. DOI: 10.3233/WOR-172611 (cit. on p. 81).

[86] Manish Dhyani, Shawn C. Roll, Matthew W. Gilbertson, Melanie Orlowski, Arash Anvari, Qian Li, Brian Anthony, and Anthony E. Samir. "A pilot study to precisely quantify forces applied by sonographers while scanning: A step toward reducing ergonomic injury." In: *Work* 58.2 (2017), pp. 241–247. ISSN: 1875-9270. DOI: 10.3233/WOR-172611. eprint: 28922185 (cit. on p. 97).

[87] Ralf Doerner, Wolfgang Broll, Paul Grimm, and Bernhard Jung, eds. *Virtual Und Augmented Reality (VR / AR)*. de. 2013th ed. eXamen.Press. Berlin, Germany: Springer, Feb. 2014. ISBN: 978-3642289026 (cit. on pp. 14–16).

[88] Ralf Doerner, Wolfgang Broll, Paul Grimm, and Bernhard Jung. *Virtual and Augmented Reality (VR/AR): Foundations and methods of extended realities (XR)*. en. Cham, Switzerland: Springer Nature, Jan. 2022. ISBN: 978-3-03079061-5 (cit. on pp. 13–16, 30).

[89] Zachary J. Dougherty and Ryder C. Winck. "Comparison Between Position and Rate Control Using a Foot Interface." In: American Society of Mechanical Engineers Digital Collection, Nov. 2018, V003T32A011. DOI: 10.1115/DSCC2018-9115 (cit. on p. 113).

[90] Mitchell Doughty and Nilesh R. Ghugre. "Head-Mounted Display-Based Augmented Reality for Image-Guided Media Delivery to the Heart: A Preliminary Investigation of Perceptual Accuracy." In: *Journal of Imaging* 8.2 (2022). ISSN: 2313-433X. DOI: 10.3390/jimaging8020033 (cit. on p. 53).

[91] Andrius Dzedzickis, Jurga Subačiūtė-Žemaitienė, Ernestas Šutinys, Urtė Samukaitė-Bubnienė, and Vytautas Bučinskas. "Advanced applications of industrial robotics: New trends and possibilities." In: *Appl. Sci.* 12.1 (Dec. 2021), p. 135. ISSN: 2076-3417. DOI: 10.3390/app12010135 (cit. on p. 1).

[92] Christine Eilers, Rob van Kemenade, Benjamin Busam, and Nassir Navab. "On the importance of patient acceptance for medical robotic imaging." In: *Int. J. CARS* 18.7 (July 2023), pp. 1261–1267. ISSN: 1861-6429. DOI: 10.1007/s11548-023-02948-5 (cit. on pp. 25, 152).

[93] Shirine El Zaatari, Mohamed Marei, Weidong Li, and Zahid Usman. "Cobot programming for collaborative industrial tasks: An overview." In: *Rob. Auton. Syst.* 116 (June 2019), pp. 162–180. ISSN: 0921-8890. DOI: 10.1016/j.robot.2019.03.003 (cit. on p. 71).

[94] 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)*. IEEE, pp. 21–25. DOI: 10.1109/ICRA.2018.8461179 (cit. on p. 62).

[95] Nima Enayati, Elena De Momi, and Giancarlo Ferrigno. "Haptics in Robot-Assisted Surgery: Challenges and Benefits." In: *IEEE Rev. Biomed. Eng.* 9 (Mar. 2016), pp. 49–65. DOI: 10.1109/RBME.2016.2538080 (cit. on p. 75).

[96] Marzieh Ershad, Alireza Ahmadian, Nassim Dadashi Serej, Hooshang Saberi, and Keyvan Amini Khoiy. "Minimization of target registration error for vertebra in image-guided spine surgery." In: *Int. J. CARS* 9.1 (Jan. 2014), pp. 29–38. ISSN: 1861-6429. DOI: 10.1007/s11548-013-0914-7 (cit. on pp. 39, 40, 52, 64).

[97] João Marcelo Evangelista Belo, Anna Maria Feit, Tiare Feuchtner, and Kaj Grønbæk. "XRgonomics: Facilitating the Creation of Ergonomic 3D Interfaces." In: *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, May 2021, pp. 1–11. ISBN: 978-1-45038096-6. DOI: 10.1145/3411764.3445349 (cit. on p. 112).

[98] Kevin Evans, Shawn Roll, and Joan Baker. "Work-Related Musculoskeletal Disorders (WRMSD) Among Registered Diagnostic Medical Sonographers and Vascular Technologists: A Representative Sample." In: *Journal of Diagnostic Medical Sonography* 25.6 (Nov. 2009), pp. 287–299. ISSN: 8756-4793. DOI: 10.1177/8756479309351748 (cit. on p. 20).

[99] Farbod Fahimi. "Autonomous Robots: Modeling, Path Planning, and Control." In: New York, NY, USA: Springer, 2049. ISBN: 978-0-38709537-0 (cit. on p. 7).

[100] Ting-Yun Fang, Haichong K. Zhang, Rodolfo Finocchi, Russell H. Taylor, and Emad M. Boctor. "Force-assisted ultrasound imaging system through dual force sensing and admittance robot control." In: *Int. J. CARS* 12.6 (June 2017), pp. 983–991. ISSN: 1861-6429. DOI: 10.1007/s11548-017-1566-9 (cit. on pp. 75, 76, 79, 80, 152, 159, 163).

[101] Mana Farshid, Jeannette Paschen, Theresa Eriksson, and Jan Kietzmann. "Go boldly!: Explore augmented reality (AR), virtual reality (VR), and mixed reality (MR) for business." In: *Bus. Horiz.* 61.5 (Sept. 2018), pp. 657–663. ISSN: 0007-6813. DOI: 10.1016/j.bushor.2018.05.009 (cit. on p. 13).

[102] Ahmad'Athif Mohd Faudzi, Muaammar Hadi Kuzman Ali, M. Asyraf Azman, and Zool Hilmi Ismail. "Real-time Hand Gestures System for Mobile Robots Control." In: *Procedia Eng.* 41 (Jan. 2012), pp. 798–804. ISSN: 1877-7058. DOI: 10.1016/j.proeng.2012.07.246 (cit. on p. 112).

[103] Tiare Feuchtner and Jörg Müller. "Ownershift: Facilitating Overhead Interaction in Virtual Reality with an Ownership-Preserving Hand Space Shift." In: *ACM Conferences*. New York, NY, USA: Association for Computing Machinery, Oct. 2018, pp. 31–43. DOI: 10.1145/3242587.3242594 (cit. on p. 112).

[104] Andy Field, Jeremy Miles, and Zoë Field. *Discovering Statistics Using R*. Thousand Oaks, CA, USA: SAGE Publications Ltd, June 2025. ISBN: 978-1-44625846-0 (cit. on pp. 25, 26).

[105] Martin A. Fischler and Robert C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." In: *Commun. ACM* 24.6 (June 1981), pp. 381–395. ISSN: 0001-0782. DOI: 10.1145/358669.358692 (cit. on pp. 19, 44, 53).

[106] R. A. Fisher. In: *Statistical Methods for Research Workers*. Vol. 5. Edinburgh, UK: Oliver and Boyd, 1934. ISBN: 978-9-35128658-5 (cit. on p. 25).

[107] J. M. Fitzpatrick, J. B. West, and C. R. Maurer. "Predicting error in rigid-body point-based registration." In: *IEEE Trans. Med. Imaging* 17.5 (Aug. 2002), pp. 694–702. DOI: 10.1109/42.736021 (cit. on p. 38).

[108] J. Michael Fitzpatrick. "Fiducial registration error and target registration error are uncorrelated." In: *Proceedings Volume 7261, Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling*. Vol. 7261. SPIE, Mar. 2009, pp. 21–32. DOI: 10.1117/12.813601 (cit. on p. 23).

[109] Andrea Fontana and James Frey. "The art of science." In: *The handbook of qualitative research* 361376 (1994), pp. 361–376 (cit. on p. 24).

[110] Javad Fotouhi et al. "Reflective-AR Display: An Interaction Methodology for Virtual-to-Real Alignment in Medical Robotics." In: *IEEE Rob. Autom. Lett.* 5.2 (Feb. 2020), pp. 2722–2729. DOI: 10.1109/LRA.2020.2972831 (cit. on pp. 38–41, 62).

[111] Marek Franaszek, Geraldine S Cheok, Karl van Wyk, and Jeremy A Marvel. *Improving 3D vision-robot registration for assembly tasks*. Apr. 2020. DOI: 10.6028/nist.ir.8300. URL: http://dx.doi.org/10.6028/NIST.IR.8300 (cit. on p. 65).

[112] Junling Fu, Alberto Rota, Shufei Li, Jianzhuang Zhao, Qingsheng Liu, Elisa Iovene, Giancarlo Ferrigno, and Elena De Momi. "Recent Advancements in Augmented Reality for Robotic Applications: A Survey." In: *Actuators* 12.8 (Aug. 2023), p. 323. ISSN: 2076-0825. DOI: 10.3390/act12080323 (cit. on p. 29).

[113] Samir Yitzhak Gadre, Eric Rosen, Gary Chien, Elizabeth Phillips, Stefanie Tellex, and George Konidaris. "End-User Robot Programming Using Mixed Reality." In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 20–24. DOI: 10.1109/ICRA.2019.8793988 (cit. on p. 1).

[114] Samir Yitzhak Gadre, Eric Rosen, Gary Chien, Elizabeth Phillips, Stefanie Tellex, and George Konidaris. "End-User Robot Programming Using Mixed Reality." In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 20–24. DOI: 10.1109/ICRA.2019.8793988 (cit. on p. 37).

[115] Yoren Gaffary, Benoît Le Gouis, Maud Marchal, Ferran Argelaguet, Bruno Arnaldi, and Anatole Lécuyer. "AR Feels "Softer" than VR: Haptic Perception of Stiffness in Augmented versus Virtual Reality." In: *Transactions on Visualization and Computer Graphics (TVCG)* 23.11 (Aug. 2017), pp. 2372–2377. DOI: 10.1109/TVCG.2017.2735078 (cit. on pp. 134, 148, 153).

[116] Andre Gaschler, Maximilian Springer, Markus Rickert, and Alois Knoll. "Intuitive robot tasks with augmented reality and virtual obstacles." In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2014–07. ISBN: 978-1-4799-3685-4. DOI: 10.1109/ICRA.2014.6907747 (cit. on pp. 38, 39, 42).

[117] Matthew W. Gilbertson and Brian W. Anthony. "Force and Position Control System for Freehand Ultrasound." In: *IEEE Trans. Rob.* 31.4 (June 2015), pp. 835–849. DOI: 10.1109/TRO.2015.2429051 (cit. on pp. 4, 20, 77, 78, 94).

[118] Gauthier Gras, Hani J. Marcus, Christopher J. Payne, Philip Pratt, and Guang-Zhong Yang. "Visual Force Feedback for Hand-Held Microsurgical Instruments." In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham, Switzerland: Springer, Nov. 2015,

pp. 480–487. ISBN: 978-3-319-24553-9. DOI: 10.1007/978-3-319-24553-9_59 (cit. on pp. 77, 78).

[119] Uwe Gruenefeld, Lars Prädel, Jannike Illing, Tim Stratmann, Sandra Drolshagen, and Max Pfingsthorn. "Mind the ARm: realtime visualization of robot motion intent in head-mounted augmented reality." In: *ACM Other conferences*. New York, NY, USA: Association for Computing Machinery, Sept. 2020, pp. 259–266. DOI: 10.1145/3404983.3405509 (cit. on p. 37).

[120] Enrico Guerra and Michael Prilla. *Virtually Real Robots: XR as a Proxy for Physical Human-Robot Interaction*. Mensch und Computer 2024 - Workshopband. 2024. DOI: 10.18420/muc2024-mci-ws06-204 (cit. on pp. 2, 131).

[121] Tibor Guzsvinecz, Veronika Szucs, and Cecilia Sik-Lanyi. "Suitability of the Kinect Sensor and Leap Motion Controller—A Literature Review." In: *Sensors* 19.5 (Mar. 2019), p. 1072. ISSN: 1424-8220. DOI: 10.3390/s19051072 (cit. on p. 29).

[122] James C. Gwilliam, Mohsen Mahvash, Balazs Vagvolgyi, Alexander Vacharat, David D. Yuh, and Allison M. Okamura. "Effects of haptic and graphical force feedback on teleoperated palpation." In: *2009 IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 12–17. DOI: 10.1109/ROBOT.2009.5152705 (cit. on pp. 4, 75, 77, 78).

[123] Mark A Haidekker. *Medical imaging technology*. en. 2013th ed. SpringerBriefs in Physics. New York, NY: Springer, Apr. 2013. ISBN: 978-1-46147072-4 (cit. on p. 20).

[124] Hill Hajnal. *Medical Image Registration*. Andover, England, UK: Taylor & Francis, June 2001. ISBN: 978-0-42911499-1. DOI: 10.1201/9781420042474 (cit. on p. 23).

[125] Kimberly Hambuchen, Jessica Marquez, and Terrence Fong. "A Review of NASA Human-Robot Interaction in Space." In: *Curr. Robot. Rep.* 2.3 (Sept. 2021), pp. 265–272. ISSN: 2662-4087. DOI: 10.1007/s43154-021-00062-5 (cit. on p. 71).

[126] Nathaniel M. Hamming, Michael J. Daly, Jonathan C. Irish, and Jeffrey H. Siewerdsen. "Effect of fiducial configuration on target registration error in intraoperative cone-beam CT guidance of head and neck surgery." In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 20–25. DOI: 10.1109/IEMBS.2008.4649997 (cit. on pp. 39, 40, 52, 63, 64).

[127] Zhao Han, Yifei Zhu, Albert Phan, Fernando Sandoval Garza, Amia Castro, and Tom Williams. "Crossing reality: Comparing physical and virtual robot deixis." In: *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. Stockholm Sweden: ACM, Mar. 2023, 152–161. DOI: 10.1145/3568162.3576972 (cit. on pp. 131–133).

[128] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction." In: *Hum. Factors* 53.5 (Sept. 2011), pp. 517–527. ISSN: 0018-7208. DOI: 10.1177/0018720811417254 (cit. on pp. 159, 160).

[129] Gill Harrison and Allison Harris. "Work-related musculoskeletal disorders in ultrasound: Can you reduce risk?" In: *Ultrasound* 23.4 (June 2015), pp. 224–230. ISSN: 1742-271X. DOI: 10.1177/1742271X15593575 (cit. on p. 20).

[130] Sandra G. Hart. "Nasa-Task Load Index (NASA-TLX); 20 Years Later." In: vol. 50. 9. SAGE Publications Inc, Oct. 2006, pp. 904–908. DOI: 10.1177/154193120605000909 (cit. on p. 24).

[131] Sandra G. Hart and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." In: *Advances in Psychology*. Vol. 52. North-Holland, Jan. 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9 (cit. on pp. 24, 45, 83, 118).

[132] Masaki Haruna, Noboru Kawaguchi, Masaki Ogino, and Toshiaki Koike-Akino. "Comparison of Three Feedback Modalities for Haptics Sensation in Remote Machine Manipulation." In: *IEEE Rob. Autom. Lett.* 6.3 (Mar. 2021), pp. 5040–5047. DOI: 10.1109/LRA.2021.3070301 (cit. on pp. 76, 79, 93).

[133] Felix von Haxthausen, Sven Böttger, Daniel Wulff, Jannis Hagenah, Verónica García-Vázquez, and Svenja Ipsen. "Medical Robotics for Ultrasound Imaging: Current Systems and Future Trends." In: *Curr. Robot. Rep.* 2.1 (Mar. 2021), pp. 55–71. ISSN: 2662-4087. DOI: 10.1007/s43154-020-00037-y (cit. on pp. 21, 135).

[134] Avijit Hazra and Nithya Gogtay. "Biostatistics Series Module 3: Comparing Groups: Numerical Variables." In: *Indian J. Dermatol.* 61.3 (May 2016), pp. 251–260. ISSN: 1998-3611. DOI: 10.4103/0019-5154.182416. eprint: 27293244 (cit. on p. 26).

[135] Florian Heinrich, Fabian Joeres, Kai Lawonn, and Christian Hansen. "Comparison of Projective Augmented Reality Concepts to Support Medical Needle Insertion." In: *Transactions on Visualization and Computer Graphics (TVCG)* 25.6 (Mar. 2019), pp. 2157–2167. DOI: 10.1109/TVCG.2019.2903942 (cit. on pp. 107, 108).

[136] Abdelfetah Hentout, Mustapha Aouache, Abderraouf Maoudj, and Isma Akli. "Human–robot interaction in industrial collaborative robotics: a literature review of the decade 2008–2017." In: *Advanced Robotics* 33.15-16 (2019), pp. 764–799. DOI: 10.1080/01691864.2019.1636714 (cit. on pp. 1, 9, 10).

[137] J. Hettig, A. Mewes, O. Riabikin, M. Skalej, B. Preim, and C. Hansen. "Exploration of 3d medical image data for interventional radiology using myoelectric gesture control." In: *Guide Proceedings*. Eurographics Association, Sept. 2015, pp. 177–185. DOI: 10.5555/2853955.2853979 (cit. on p. 71).

[138] Julian Hettig, Patrick Saalfeld, Maria Luz, Mathias Becker, Martin Skalej, and Christian Hansen. "Comparison of gesture and conventional interaction techniques for interventional neuroradiology." In: *Int. J. CARS* 12.9 (Sept. 2017), pp. 1643–1653. ISSN: 1861-6429. DOI: 10.1007/s11548-017-1523-7 (cit. on p. 71).

[139] Derek L. G. Hill, Philipp G. Batchelor, Mark Holden, and David J. Hawkes. "Medical image registration." In: *Phys. Med. Biol.* 46.3 (Mar. 2001), R1. ISSN: 0031-9155. DOI: 10.1088/0031-9155/46/3/201 (cit. on p. 19).

[140] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. "Consumed endurance: a metric to quantify arm fatigue of mid-air interactions." In: *CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, Apr. 2014, pp. 1063–1072. ISBN: 978-1-45032473-1. DOI: 10.1145/2556288.2557130 (cit. on pp. 2, 111–113).

[141] Tzu-Hsuan Ho and Kai-Tai Song. "Supervised Control for Robot-Assisted Surgery Using Augmented Reality." In: *2020 20th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, pp. 13–16. DOI: 10.23919/ICCAS50221.2020.9268278 (cit. on pp. 38, 39, 42).

[142] Herbert Hoijtink, Irene Klugkist, and Paul A. Boelen, eds. *Bayesian Evaluation of Informative Hypotheses*. Springer New York, 2008. DOI: 10.1007/978-0-387-09612-4 (cit. on pp. 26, 89, 140, 154).

[143] Tim Horeman, Sharon P. Rodrigues, John J. van den Dobbelsteen, Frank-Willem Jansen, and Jenny Dankelman. "Visual force feedback in laparoscopic training." In: *Surg. Endosc.* 26.1 (Jan. 2012), pp. 242–248. ISSN: 1432-2218. DOI: 10.1007/s00464-011-1861-4 (cit. on pp. 77, 78).

[144] Kasper Hornbæk. "Current practice in measuring usability: Challenges to usability studies and research." In: *Int. J. Hum. Comput. Stud.* 64.2 (Feb. 2006), pp. 79–102. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2005.06.002 (cit. on p. 22).

[145] Thomas Howard and Jérôme Szewczyk. "Assisting Control of Forces in Laparoscopy Using Tactile and Visual Sensory Substitution." In: *New Trends in Medical and Service Robots*. Cham, Switzerland: Springer, Apr. 2016, pp. 151–164. ISBN: 978-3-319-30674-2. DOI: 10.1007/978-3-319-30674-2_12 (cit. on pp. 75–79, 93, 96, 106).

[146] Shang-Ying Hsieh and Jun-Ming Lu. "Feasibility evaluation for immersive virtual reality simulation of human-machine collaboration: a case study of hand-over tasks." In: *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018) Volume V: Human Simulation and Virtual Environments, Work With Computing Systems (WWCS), Process Control 20*. Springer. Florence, Italy: Springer International Publishing, 2019, pp. 364–369. DOI: 10.1007/978-3-319-96077-7_38 (cit. on p. 132).

[147] Hui-Min Huang, Elena Messina, Robert Wade, Ralph English, Brian Novak, and James Albus. "Autonomy Measures for Robots." In: American Society of Mechanical Engineers Digital Collection, Mar. 2008, pp. 1241–1247. DOI: 10.1115/IMECE2004-61812 (cit. on p. 9).

[148] Johannes Hügle, Jens Lambrecht, and Jörg Krüger. "An integrated approach for industrial robot control and programming combining haptic and non-haptic gestures." In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 2017–01. DOI: 10.1109/ROMAN.2017.8172402 (cit. on p. 112).

[149] *ISO 9000:2005: Quality management systems — Fundamentals and vocabulary*. Tech. rep. Geneva, CH: International Organization for Standardization, 2005 (cit. on p. 22).

[150] *ISO 9241-210:2019: Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems*. Tech. rep. Geneva, CH: International Organization for Standardization, 2019 (cit. on p. 21).

[151] Kenji Inoue, Seri Nonaka, Yoshihiro Ujiie, Tomohito Takubo, and Tatsuo Arai. "Comparison of human psychology for real and virtual mobile manipulators." In: *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* New York, USA: IEEE, 2005, pp. 73–78. DOI: 10.1109/roman.2005.1513759 (cit. on p. 132).

[152] Hasan Iqbal, Seemab Latif, Yukang Yan, Chun Yu, and Yuanchun Shi. "Reducing Arm Fatigue in Virtual Reality by Introducing 3D-Spatial Offset." In: *IEEE Access* 9 (Apr. 2021), pp. 64085–64104. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3075769 (cit. on p. 124).

[153] Andrés Jaramillo-Yánez, Marco E. Benalcázar, and Elisa Mena-Maldonado. "Real-Time Hand Gesture Recognition Using Surface Electromyography and Machine Learning: A Systematic Literature Review." In: *Sensors* 20.9 (Apr. 2020), p. 2467. ISSN: 1424-8220. DOI: 10.3390/s20092467 (cit. on p. 12).

[154] Reza N Jazar. *Theory of Applied Robotics: Kinematics, dynamics, and Control*. en. Cham, Switzerland: Springer Nature, May 2022. ISBN: 978-1-4419-1749-2. DOI: 10.1007/978-1-4419-1750-8 (cit. on p. 7).

[155] Zhongliang Jiang, Matthias Grimm, Mingchuan Zhou, Ying Hu, and Nassir Navab. "Automatic Force-Based Probe Positioning for Precise Robotic Ultrasound Acquisition." In: *IEEE Trans. Ind. Electron.* PP.99 (Nov. 2020), p. 1. ISSN: 0278-0046. DOI: 10.1109/TIE.2020.3036215 (cit. on p. 162).

[156] Zhongliang Jiang, Septimiu E. Salcudean, and Nassir Navab. "Robotic ultrasound imaging: State-of-the-art and future perspectives." In: *Medical Image Analysis* 89 (2023), p. 102878. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2023.102878 (cit. on pp. 21, 27).

[157] Yannick Jonetzko, Judith Hartfill, Niklas Fiedler, Fangwei Zhong, Frank Steinicke, and Jianwei Zhang. "Evaluating Visual and Auditory Substitution of Tactile Feedback During Mixed Reality Teleoperation." In: *Cognitive Computation and Systems*. Singapore: Springer, May 2023, pp. 331–345. ISBN: 978-981-99-2789-0. DOI: 10.1007/978-981-99-2789-0_28 (cit. on pp. 75–78, 93, 95, 107).

[158] Tonke L. de Jong, Adriaan Moelker, Jenny Dankelman, and John J. van den Dobbelsteen. "Designing and validating a PVA liver phantom with respiratory motion for needle-based interventions." In: *Int. J. CARS* 14.12 (Dec. 2019), pp. 2177–2186. ISSN: 1861-6429. DOI: 10.1007/s11548-019-02029-6 (cit. on p. 84).

[159] KUKA AG. *KUKA Sunrise.OS 1.11: Operating and Programming Instructions for System Integrators*. Operating and Programming Instructions for System Integrators. KUKA AG. Germany, 2016 (cit. on p. 38).

[160] Linh Kästner, Vlad Catalin Frasineanu, and Jens Lambrecht. "A 3D-Deep-Learning-based Augmented Reality Calibration Method for Robotic Environments using Depth Sensor Data." In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2020–31. DOI: 10.1109/ICRA40945.2020.9197155 (cit. on p. 65).

[161] Jan Kallwies, Bianca Forkel, and Hans-Joachim Wuensche. "Determining and Improving the Localization Accuracy of AprilTag Detection." In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2020–31. DOI: 10.1109/ICRA40945.2020.9197427 (cit. on pp. 39, 52, 65).

[162] R. E. Kalman. "A New Approach to Linear Filtering and Prediction Problems." In: *J. Basic Eng.* 82.1 (Mar. 1960), pp. 35–45. ISSN: 0021-9223. DOI: 10.1115/1.3662552 (cit. on p. 19).

[163] Maria Karam and M.C. Schraefel. "A taxonomy of gestures in human computer interactions." In: (2005) (cit. on p. 11).

[164] James Kennedy, Paul Baxter, and Tony Belpaeme. "Children comply with a robot's indirect requests." In: *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. Bielefeld Germany: ACM, Mar. 2014, 198–199. DOI: 10.1145/2559636.2559820 (cit. on p. 132).

[165] Vanessa L. Kennedy, Carol A. Flavell, and Kenji Doma. "Intra-rater reliability of transversus abdominis measurement by a novice examiner: Comparison of "freehand" to "probe force device" method of real-time ultrasound imaging." In: *Ultrasound* 27.3 (Feb. 2019), pp. 156–166. ISSN: 1742-271X. DOI: 10.1177/1742271X19831720 (cit. on p. 4).

[166] Ali Khamene and Frank Sauer. "A Novel Phantom-Less Spatial and Temporal Ultrasound Calibration Method." In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*. Berlin, Germany: Springer, 2005, pp. 65–72. ISBN: 978-3-540-32095-1. DOI: 10.1007/11566489_9 (cit. on p. 17).

[167] Hyung-il Kim, Boram Yoon, Seo Young Oh, and Woontack Woo. "Visualizing Hand Force with Wearable Muscle Sensing for Enhanced Mixed Reality Remote Collaboration." In: *Transactions on Visualization and Computer Graphics (TVCG)* 29.11 (Oct. 2023), pp. 4611–4621. DOI: 10.1109/TVCG.2023.3320210 (cit. on p. 77).

[168] M. Kim, K. Oh, J. Choi, J. Jung, and Y. Kim. "User-Centered HRI: HRI Research Methodology for Designers." In: *Mixed Reality and Human-Robot Interaction*. Dordrecht, The Netherlands: Springer, 2011, pp. 13–33. ISBN: 978-94-007-0582-1. DOI: 10.1007/978-94-007-0582-1_2 (cit. on p. 21).

[169] Won Kim, F. Tendick, S. Ellis, and L. Stark. "A comparison of position and rate control for telemanipulations with consideration of manipulator system dynamics." In: *IEEE Journal on Robotics and Automation* 3.5 (Oct. 1987), pp. 426–436. DOI: 10.1109/JRA.1987.1087117 (cit. on pp. 10, 111, 113, 125).

[170] Yonjae Kim, Peter C. W. Kim, Rebecca Selle, Azad Shademan, and Axel Krieger. "Experimental evaluation of contact-less hand tracking systems for tele-operation of surgical tasks." In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 2014–07. ISBN: 978-1-4799-3685-4. DOI: 10.1109/ICRA.2014.6907364 (cit. on p. 112).

[171] Masaya Kitagawa, Daniell Dokko, Allison M. Okamura, and David D. Yuh. "Effect of sensory substitution on suture-manipulation forces for robotic surgical systems." In: *J. Thorac. Cardiovasc. Surg.* 129.1 (Jan. 2005), pp. 151–158. ISSN: 0022-5223. DOI: 10.1016/j.jtcvs.2004.05.029 (cit. on pp. 76–78, 93).

[172] Risto Kojcev, Ashkan Khakzar, Bernhard Fuerst, Oliver Zettinig, Carole Fahkry, Robert DeJong, Jeremy Richmon, Russell Taylor, Edoardo Sinibaldi, and Nassir Navab. "On the reproducibility of expert-operated and robotic ultrasound acquisitions." In: *Int. J. CARS* 12.6 (June 2017), pp. 1003–1011. ISSN: 1861-6429. DOI: 10.1007/s11548-017-1561-1 (cit. on p. 20).

[173] Vera M Kolb, ed. *Handbook of astrobiology*. en. Series in Astrobiology. London, England: CRC Press, Jan. 2019. ISBN: 9781439829448 (cit. on p. 23).

[174] M. H. Korayem, M. A. Madihi, and V. Vahidifar. "Controlling surgical robot arm using leap motion controller with Kalman filter." In: *Measurement* 178 (June 2021), p. 109372. ISSN: 0263-2241. DOI: 10.1016/j.measurement.2021.109372 (cit. on pp. 111, 112, 114).

[175] Jan Krieglstein, Gesche Held, Balázs A. Bálint, Frank Nägele, and Werner Kraus. "Skill-based Robot Programming in Mixed Reality with Ad-hoc Validation Using a Force-enabled Digital Twin." In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2023–02. DOI: 10.1109/ICRA48891.2023.10161095 (cit. on p. 1).

[176] Robert Kristof and Cristian Moldovan. "Human–Machine Interfaces for Industrial Robot Control: A Comprehensive Review of the State of the Art." In: *25th International Symposium on Measurements and Control in Robotics*. Cham, Switzerland: Springer, Apr. 2024, pp. 201–213. ISBN: 978-3-031-51085-4. DOI: 10.1007/978-3-031-51085-4_18 (cit. on p. 10).

[177] Ernst Kruijff, J. Edward Swan, and Steven Feiner. "Perceptual issues in augmented reality revisited." In: *2010 IEEE International Symposium on Mixed and Augmented Reality*. New York, USA: IEEE, 2010, pp. 13–16. DOI: 10.1109/ISMAR.2010.5643530 (cit. on p. 145).

[178] Dennis Krupke, Frank Steinicke, Paul Lubos, Yannick Jonetzko, Michael Görner, and Jianwei Zhang. "Comparison of Multimodal Heading and Pointing Gestures for Co-Located Mixed Reality Human-Robot Interaction." In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 01–05. DOI: 10.1109/IROS.2018.8594043 (cit. on pp. 38–41, 62).

[179] Rita Latikka, Nina Savela, Aki Koivula, and Atte Oksanen. "Attitudes toward robots as equipment and coworkers and the impact of robot autonomy level." In: *International Journal of Social Robotics* 13.7 (2021), pp. 1747–1759. DOI: 0.1007/s12369-020-00743-9 (cit. on p. 21).

[180] Cha Lee, Scott Bonebrake, Doug A Bowman, and Tobias Hollerer. "The role of latency in the validity of AR simulation." In: *2010 IEEE Virtual Reality Conference (VR)*. New York, USA: IEEE, Mar. 2010, pp. 11–18. DOI: 10.1109/vr.2010.5444820 (cit. on p. 131).

[181] Hvuniin Lee and Woontack Woo. "Investigating Display Position of a Head-Fixed Augmented Reality Notification for Dual-task." In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2022, pp. 12–16. DOI: 10.1109/VRW55335.2022.00143 (cit. on pp. 79, 95).

[182] Hyunjin Lee and Woontack Woo. "Exploring Augmented Reality Notification Placement while Communicating with Virtual Avatar." In: *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 2022, pp. 17–21. DOI: 10.1109/ISMAR-Adjunct57072.2022.00143 (cit. on pp. 79, 95).

[183] Joon Hyub Lee, Yongkwan Kim, Sang-Gyun An, and Seok-Hyung Bae. "Robot Telekinesis: Application of a Unimanual and Bimanual Object Manipulation Technique to Robot Control." In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2020–31. DOI: 10.1109/ICRA40945.2020.9197517 (cit. on pp. 1, 2, 28, 71–73, 111, 112).

[184] Michael D. Lee and Eric-Jan Wagenmakers. *Bayesian Cognitive Modeling*. Cambridge, UK: Cambridge University Press, 2014. DOI: 10.1017/cbo9781139087759 (cit. on p. 27).

[185] Timothy E. Lee, Jonathan Tremblay, Thang To, Jia Cheng, Terry Mosier, and Oliver Kroemer. "Camera-to-Robot Pose Estimation from a Single Image." In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2020–31. DOI: 10.1109/ICRA40945.2020.9196596 (cit. on pp. 38, 42, 62, 63, 65).

[186] Franziska Legler, Jonas Trezl, Dorothea Langer, Max Bernhagen, Andre Dettmann, and Angelika C Bullinger. "Emotional Experience in Human–Robot Collaboration: Suitability of Virtual Reality Scenarios to Study Interactions beyond Safety Restrictions." In: *Robotics* 12.6 (2023), p. 168 (cit. on p. 131).

[187] Dingwei Li, Jixiang Yang, Huan Zhao, and Han Ding. "Contact force plan and control of robotic grinding towards ensuring contour accuracy of curved surfaces." In: *Int. J. Mech. Sci.* 227 (Aug. 2022), p. 107449. ISSN: 0020-7403. DOI: 10.1016/j.ijmecsci.2022.107449 (cit. on p. 4).

[188] Dongrui Li, Zhigang Cheng, Gang Chen, Fangyi Liu, Wenbo Wu, Jie Yu, Ying Gu, Fengyong Liu, Chao Ren, and Ping Liang. "A multimodality imaging-compatible insertion robot with a respiratory motion calibration module designed for ablation of liver tumors: a preclinical study." In: *Int. J. Hyperthermia* (Nov. 2018). ISSN: 0265-6736. DOI: 10.1080/02656736.2018.1456680 (cit. on p. 21).

[189] Jamy Li. "The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents." In: *International Journal of Human-Computer Studies* 77 (2015), pp. 23–37. DOI: 10.1016/j.ijhcs.2015.01.001 (cit. on p. 132).

[190] Keyu Li, Yangxin Xu, and Max Q.-H. Meng. "An Overview of Systems and Techniques for Autonomous Robotic Ultrasound Acquisitions." In: *IEEE Transactions on Medical Robotics and Bionics* 3.2 (Apr. 2021), pp. 510–524. ISSN: 2576-3202. DOI: 10.1109/TMRB.2021.3072190 (cit. on p. 162).

[191] Rui Li, Marc van Almkerk, Sanne van Waveren, Elizabeth Carter, and Iolanda Leite. "Comparing human-robot proxemics between virtual reality and the real world." In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Daegu, South Korea: IEEE, Mar. 2019, 431–439. DOI: 10.1109/hri.2019.8673116 (cit. on pp. 132, 152, 153).

[192] Yang Li, Jin Huang, Feng Tian, Hong-An Wang, and Guo-Zhong Dai. "Gesture interaction in virtual reality." In: *Virtual Reality & Intelligent Hardware* 1.1 (Feb. 2019), pp. 84–112. ISSN: 2096-5796. DOI: 10.3724/SP.J.2096-5796.2018.0006 (cit. on pp. 12, 71).

[193] Yihan Li, Yong Hu, Zidan Wang, and Xukun Shen. "Evaluating the Object-Centered User Interface in Head-Worn Mixed Reality Environment." In: *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2022, pp. 17–21. DOI: 10.1109/ISMAR55827.2022.00057 (cit. on p. 79).

[194] Josip Tomo Licardo, Mihael Domjan, and Tihomir Orehovački. "Intelligent Robotics—A Systematic Review of Emerging Technologies and Trends." In: *Electronics* 13.3 (Jan. 2024), p. 542. ISSN: 2079-9292. DOI: 10.3390/electronics13030542 (cit. on p. 71).

[195] Tsung-Chi Lin, Achyuthan Unni Krishnan, and Zhi Li. "Physical Fatigue Analysis of Assistive Robot Teleoperation via Whole-body Motion Mapping." In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 03–08. DOI: 10.1109/IROS40897.2019.8968544 (cit. on p. 118).

[196] Yanping Lin, Shihang Chen, Wangjie Xu, Xiaoxiao Zhu, and Qixin Cao. "Robotic system for accurate percutaneous puncture guided by 3D–2D ultrasound." In: *Int. J. CARS* 18.2 (2023), pp. 217–225. DOI: 10.1007/s11548-022-02766-1 (cit. on pp. 21, 80, 83).

[197] Mingyu Liu, Mathieu Nancel, and Daniel Vogel. "Gunslinger: Subtle Arms-down Mid-air Interaction." In: *UIST '15: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. New York, NY, USA: Association for Computing Machinery, Nov. 2015, pp. 63–71. ISBN: 978-1-45033779-3. DOI: 10.1145/2807442.2807489 (cit. on pp. 112, 114).

[198] Oliver Liu, Daniel Rakita, Bilge Mutlu, and Michael Gleicher. "Understanding human-robot interaction in virtual reality." In: *2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN)*. New York, USA: IEEE, 2017, pp. 751–757. DOI: 10.1109/roman.2017.8172387 (cit. on p. 132).

[199] Julio Alegre Luna, Anthony Vasquez Rivera, Alejandra Loayza Mendoza, Andres Montoya, et al. "Development of a touchless control system for a clinical robot with multimodal user interface." In: *International Journal of Advanced Computer Science and Applications* 14.9 (2023). DOI: 10.14569/IJACSA.2023.01409111 (cit. on pp. 28, 72).

[200] Akshit Lunia, Ashley Stevens, Cavender Holt, Rhyan Morgan, Joshua Norris, Shailendran Poyyamozhi, and Yehua Zhong. *Modeling, Motion Planning, and Control of Manipulators and Mobile Robots*. URL: https://opentextbooks.clemson.edu/wangrobotics (cit. on p. 8).

[201] I. Scott MacKenzie and R. William Soukoreff. "Text Entry for Mobile Computing: Models and Methods,Theory and Practice." In: *Human–Computer Interaction* (Sept. 2002) (cit. on p. 23).

[202] Simone Macciò, Alessandro Carfì, and Fulvio Mastrogiovanni. "Mixed Reality as Communication Medium for Human-Robot Collaboration." In: *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 23–27. DOI: 10.1109/ICRA46639.2022.9812233 (cit. on p. 1).

[203] D. Magee, Y. Zhu, R. Ratnalingam, P. Gardner, and D. Kessel. "An augmented reality simulator for ultrasound guided needle placement training." In: *Med. Biol. Eng. Comput.* 45.10 (Oct. 2007), pp. 957–967. ISSN: 1741-0444. DOI: 10.1007/s11517-007-0231-9 (cit. on p. 17).

[204] Ville Mäkelä, Rivu Radiah, Saleh Alsherif, Mohamed Khamis, Chong Xiao, Lisa Borchert, Albrecht Schmidt, and Florian Alt. "Virtual field studies." In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, Apr. 2020, 1–15. DOI: 10.1145/3313831.3376796 (cit. on p. 132).

[205] Zhanat Makhataeva and Huseyin Atakan Varol. "Augmented Reality for Robotics: A Review." In: *Robotics* 9.2 (Apr. 2020), p. 21. ISSN: 2218-6581. DOI: 10.3390/robotics9020021 (cit. on pp. 29, 171).

[206] Ali Ahmad Malik, Tariq Masood, and Arne Bilberg. "Virtual reality in manufacturing: immersive and collaborative artificial-reality in design of human-robot workspace." en. In: *Int. J. Comput. Integr. Manuf.* 33.1 (Jan. 2020), pp. 22–37. DOI: 10.1080/0951192x.2019.1690685 (cit. on p. 131).

[207] Alexander Marquardt, Christina Trepkowski, Tom David Eibich, Jens Maiero, and Ernst Kruijff. "Non-Visual Cues for View Management in Narrow Field of View Augmented Reality Displays." In: *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2019, pp. 14–18. DOI: 10.1109/ISMAR.2019.000-3 (cit. on p. 96).

[208] Alejandro Martin-Gomez, Haowei Li, Tianyu Song, Sheng Yang, Guangzhi Wang, Hui Ding, Nassir Navab, Zhe Zhao, and Mehran Armand. "STTAR: Surgical Tool Tracking Using Off-the-Shelf Augmented Reality Head-Mounted Displays." In: *Transactions on Visualization and Computer Graphics (TVCG)* 30.7 (Jan. 2023), pp. 3578–3593. DOI: 10.1109/TVCG.2023.3238309 (cit. on p. 53).

[209] Alejandro Martin-Gomez, Alexander Winkler, Kevin Yu, Daniel Roth, Ulrich Eck, and Nassir Navab. "Augmented Mirrors." In: *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, pp. 09–13. DOI: 10.1109/ISMAR50242.2020.00045 (cit. on p. 50).

[210] Jeremy A. Marvel, Shelly Bagchi, Megan Zimmerman, and Brian Antonishek. "Towards Effective Interface Designs for Collaborative HRI in Manufacturing: Metrics and Measures." In: *J. Hum.-Robot. Interact.* 9.4 (May 2020), pp. 1–55. DOI: 10.1145/3385009 (cit. on p. 23).

[211] Jeremy A. Marvel and Karl Van Wyk. "Simplified framework for robot coordinate registration for manufacturing applications." In: *2016 IEEE International Symposium on Assembly and Manufacturing (ISAM)*. IEEE, pp. 21–22. DOI: 10.1109/ISAM.2016.7750718 (cit. on pp. 38–40, 42, 52, 62).

[212] M. J. Massimino. "Improved force perception through sensory substitution." In: *Control Eng. Pract.* 3.2 (Feb. 1995), pp. 215–222. ISSN: 0967-0661. DOI: 10.1016/0967-0661(94)00079-V (cit. on p. 75).

[213] Eloise Matheson, Riccardo Minto, Emanuele G. G. Zampieri, Maurizio Faccio, and Giulio Rosati. "Human–Robot Collaboration in Manufacturing Applications: A Review." In: *Robotics* 8.4 (Dec. 2019), p. 100. ISSN: 2218-6581. DOI: 10.3390/robotics8040100 (cit. on p. 27).

[214] Kim Mathiassen, Jørgen Enger Fjellin, Kyrre Glette, Per Kristian Hol, and Ole Jakob Elle. "An Ultrasound Robotic System Using the Commercial Robot UR5." In: *Front. Rob. AI* 3 (Feb. 2016), p. 168139. ISSN: 2296-9144. DOI: 10.3389/frobt.2016.00001 (cit. on p. 75).

[215] Florian Mathis, Kami Vaniea, and Mohamed Khamis. "RepliCueAuth: Validating the use of a lab-based virtual reality setup for evaluating authentication systems." In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama Japan: ACM, May 2021, 1–18. DOI: 10.1145/3411764.3445478 (cit. on pp. 131, 132).

[216] Daniel Mendes, Filipe Relvas, Alfredo Ferreira, and Joaquim Jorge. "The benefits of DOF separation in mid-air 3D object manipulation." In: *VRST '16: Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. New York, NY, USA: Association for Computing Machinery, Nov. 2016, pp. 261–268. ISBN: 978-1-45034491-3. DOI: 10.1145/2993369.2993396 (cit. on p. 29).

[217] Stephanie M. Merritt, Heather Heimbaugh, Jennifer LaChapell, and Deborah Lee. "I Trust It, but I Don't Know Why: Effects of Implicit Attitudes Toward Automation on Trust in an Automated System." In: *Hum. Factors* 55.3 (Nov. 2012), pp. 520–534. ISSN: 0018-7208. DOI: 10.1177/0018720812465081 (cit. on pp. 152, 160).

[218] Sébastien Mick, Daniel Cattaert, Florent Paclet, Pierre-Yves Oudeyer, and Aymar de Rugy. "Performance and Usability of Various Robotic Arm Control Modes from Human Force Signals." In: *Front. Neurorob.* 11 (Oct. 2017), p. 278858. ISSN: 1662-5218. DOI: 10.3389/fnbot.2017.00055 (cit. on p. 113).

[219] Paul Milgram and Fumio Kishino. "A taxonomy of mixed reality visual displays." In: *IEICE TRANSACTIONS on Information and Systems* 77.12 (1994), pp. 1321–1329. URL: https://globals.ieice.org/en_transactions/information/10.1587/e77-d_12_1321/_p (cit. on p. 13).

[220] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. "Augmented reality: a class of displays on the reality-virtuality continuum." In: *Proceedings Volume 2351, Telemanipulator and Telepresence Technologies*. Vol. 2351. SPIE, Dec. 1995, pp. 282–292. DOI: 10.1117/12.197321 (cit. on p. 14).

[221] Zhe Min and Max Q.-H. Meng. "Robust Generalized Point Set Registration using Inhomogeneous Hybrid Mixture Models via Expectation Maximization." In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 20–24. DOI: 10.1109/ICRA.2019.8794135 (cit. on p. 65).

[222] Sushmita Mitra and Tinku Acharya. "Gesture Recognition: A Survey." In: *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 37.3 (Apr. 2007), pp. 311–324. DOI: 10.1109/TSMCC.2007.893280 (cit. on p. 11).

[223] Yoshiaki Mizuchi and Tetsunari Inamura. "Cloud-based multimodal human-robot interaction simulator utilizing ros and unity frameworks." In: *2017 IEEE/SICE international symposium on system integration (SII)*. New York, USA: IEEE, 2017, pp. 948–955. DOI: 10.1109/sii.2017.8279345 (cit. on p. 131).

[224] Alexander Moortgat-Pick, Peter So, Michael J. Sack, Emma G. Cunningham, Benjamin P. Hughes, Anna Adamczyk, Andriy Sarabakha, Leila Takayama, and Sami Haddadin. "A-RIFT: Visual Substitution of Force Feedback for a Zero-Cost Interface in Telemanipulation." In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 23–27. DOI: 10.1109/IROS47612.2022.9981365 (cit. on pp. 77, 78, 93).

[225] Nicholas A. Nadeau, Ilian A. Bonev, and Ahmed Joubair. "Impedance Control Self-Calibration of a Collaborative Robot Using Kinematic Coupling." In: *Robotics* 8.2 (Apr. 2019), p. 33. ISSN: 2218-6581. DOI: 10.3390/robotics8020033 (cit. on pp. 38, 42).

[226] Andrew Yeh Ching Nee and Soh Khim Ong, eds. *Springer handbook of augmented reality*. en. 1st ed. Springer Handbooks. Cham, Switzerland: Springer Nature, Jan. 2023. ISBN: 978-3-030-67822-7 (cit. on p. 17).

[227] Federica Nenna, Valeria Orso, Davide Zanardi, and Luciano Gamberini. "The virtualization of human–robot interactions: a user-centric workload assessment." In: *Virtual Reality* 27.2 (2023), pp. 553–571. DOI: 10.1007/s10055-022-00667-x (cit. on pp. 131–133).

[228] Vinh Nguyen, Xiaofeng Liu, and Jeremy Marvel. "Augmented Reality Interface for Robot-Sensor Coordinate Registration." In: *J. Comput. Inf. Sci. Eng.* 24.3 (Oct. 2023), p. 034501. ISSN: 1530-9827. DOI: 10.1115/1.4063131 (cit. on pp. 38–40, 52, 62, 63).

[229] D. Ni, A. W. W. Yew, S. K. Ong, and A. Y. C. Nee. "Haptic and visual augmented reality interface for programming welding robots." In: *Adv. Manuf.* 5.3 (Sept. 2017), pp. 191–198. ISSN: 2195-3597. DOI: 10.1007/s40436-017-0184-7 (cit. on pp. 38, 39, 42).

[230] Saeed B Niku. *Introduction to robotics*. en. 3rd ed. Nashville, TN: John Wiley & Sons, Feb. 2020. ISBN: 978-1-11952762-6 (cit. on pp. 1, 7).

[231] Emanuel Nunez Sardinha, Nancy Zook, Virginia Ruiz Garate, David Western, and Marcela Munera. "Comparison of Three Interface Approaches for Gaze Control of Assistive Robots for Individuals with Tetraplegia." In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025 (cit. on p. 79).

[232] Cristina Nuzzi, Stefano Ghidini, Roberto Pagani, Simone Pasinetti, Gabriele Coffetti, and Giovanna Sansoni. "Hands-Free: a robot augmented reality teleoperation system." In: *2020 17th International Conference on Ubiquitous Robots (UR)*. IEEE, pp. 22–26. DOI: 10.1109/UR49135.2020.9144841 (cit. on pp. 38, 42).

[233] Tommy Öberg, Leif Sandsjö, and Roland Kadefors. "Subjective and objective evaluation of shoulder muscle fatigue." In: *Ergonomics* (Aug. 1994). DOI: 10.1080/00140139408964911 (cit. on p. 24).

[234] Zachary Ochitwa, Reza Fotouhi, Scott J. Adams, Adriana Paola Noguera Cundar, and Haron Obaid. "MSK-TIM: A Telerobotic Ultrasound System for Assessing the Musculoskeletal System." In: *Sensors* 24.7 (Apr. 2024), p. 2368. ISSN: 1424-8220. DOI: 10.3390/s24072368 (cit. on p. 4).

[235] A. M. Okamura. "Methods for haptic feedback in teleoperated robot-assisted surgery." In: *Industrial Robot: An International Journal* 31.6 (Dec. 2004), pp. 499–508. DOI: 10.1108/01439910410566362 (cit. on pp. 1, 75).

[236] Bukeikhan Omarali, Kaspar Althoefer, Fulvio Mastrogiovanni, Maurizio Valle, and Ildar Farkhatdinov. "Workspace Scaling and Rate Mode Control for Virtual Reality based Robot Teleoperation." In: *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, pp. 17–20. DOI: 10.1109/SMC52423.2021.9658796 (cit. on pp. 111, 113).

[237] Linda Onnasch and Eileen Roesler. "A Taxonomy to Structure and Analyze Human–Robot Interaction." In: *Int. J. Social Rob.* 13.4 (July 2021), pp. 833–849. ISSN: 1875-4805. DOI: 10.1007/s12369-020-00666-5 (cit. on p. 10).

[238] Luca Oppici, Kim Grütters, Felix Bechtolsheim, and Stefanie Speidel. "How does the modality of delivering force feedback influence the performance and learning of surgical suturing skills? We don't know, but we better find out! A review." In: *Surg. Endosc.* 37.4 (Apr. 2023), pp. 2439–2452. ISSN: 1432-2218. DOI: 10.1007/s00464-022-09740-7 (cit. on p. 78).

[239] Scott Ososky, Tracy Sanders, Florian Jentsch, Peter Hancock, and Jessie Y. C. Chen. "Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems." In: *Proceedings Volume 9084, Unmanned Systems Technology XVI*. Vol. 9084. SPIE, June 2014, pp. 112–123. DOI: 10.1117/12.2050622 (cit. on p. 152).

[240] Mikhail Ostanin, Stanislav Mikhel, Alexey Evlampiev, Valeria Skvortsova, and Alexandr Klimchik. "Human-robot interaction for robotic manipulator programming in Mixed Reality." In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2020–31. DOI: 10.1109/ICRA40945.2020.9196965 (cit. on p. 1).

[241] Mikhail Ostanin, Stanislav Mikhel, Alexey Evlampiev, Valeria Skvortsova, and Alexandr Klimchik. "Human-robot interaction for robotic manipulator programming in Mixed Reality." In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2020–31. DOI: 10.1109/ICRA40945.2020.9196965 (cit. on pp. 38–40, 62).

[242] Munir Oudah, Ali Al-Naji, and Javaan Chahl. "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques." In: *J. Imaging* 6.8 (July 2020), p. 73. ISSN: 2313-433X. DOI: 10.3390/jimaging6080073 (cit. on p. 12).

[243] Sharon Oviatt. "MULTIMODAL INTERFACES." In: *The Human-Computer Interaction Handbook*. Boca Raton, FL, USA: CRC Press, Sept. 2007, pp. 439–458. ISBN: 978-0-42916397-5. DOI: 10.1201/9781410615862-32 (cit. on p. 12).

[244] John O Oyekan, Windo Hutabarat, Ashutosh Tiwari, Raphael Grech, Min H Aung, Maria P Mariani, Laura López-Dávalos, Timothé Ricaud, Sumit Singh, and Charlène Dupuis. "The effectiveness of virtual environments in developing collaborative strategies between industrial robots and humans." In: *Robotics and Computer-Integrated Manufacturing* 55 (2019), pp. 41–54. DOI: 10.1016/j.rcim.2018.07.006 (cit. on p. 131).

[245] V. I. Pavlovic, R. Sharma, and T. S. Huang. "Visual interpretation of hand gestures for human-computer interaction: a review." In: *IEEE Trans. Pattern Anal. Mach. Intell.* 19.7 (Aug. 2002), pp. 677–695. DOI: 10.1109/34.598226 (cit. on p. 11).

[246] Huaishu Peng, Jimmy Briggs, Cheng-Yao Wang, Kevin Guo, Joseph Kider, Stefanie Mueller, Patrick Baudisch, and François Guimbretière. "RoMA: Interactive Fabrication with Augmented Reality and a Robotic 3D Printer." In: *ACM Conferences*. New York, NY, USA: Association for Computing Machinery, Apr. 2018, pp. 1–12. DOI: 10.1145/3173574.3174153 (cit. on pp. 38, 39).

[247] S. Camille Peres, Vickie Nguyen, Philip T. Kortum, Magdy Akladios, S. Bart Wood, and Andrew Muddimer. "Software ergonomics: relating subjective and objective measures." In: *ACM Conferences*. New York, NY, USA: Association for Computing Machinery, Apr. 2009, pp. 3949–3954. DOI: 10.1145/1520340.1520599 (cit. on p. 24).

[248] Christopher Peters, Chengjie Li, Fangkai Yang, Vanya Avramova, and Gabriel Skantze. "Investigating Social Distances between Humans, Virtual Humans and Virtual Robots in Mixed Reality." In: *ACM Conferences*. International Foundation for Autonomous Agents and Multiagent Systems, July 2018, pp. 2247–2249. DOI: 10.5555/3237383.3238137 (cit. on p. 131).

[249] Christopher Peters, Fangkai Yang, Himangshu Saikia, Chengjie Li, and Gabriel Skantze. "Towards the use of Mixed Reality for HRI Design via Virtual Robots." In: *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*. Cambridge, UK: ACM/IEEE, 2018, pp. 1–4 (cit. on p. 131).

[250] Lucas Plabst, Sebastian Oberdörfer, Francisco Raul Ortega, and Florian Niebling. "Push the Red Button: Comparing Notification Placement with Augmented and Non-Augmented Tasks in AR." In: *SUI '22: Proceedings of the 2022 ACM Symposium on Spatial User Interaction*. New

York, NY, USA: Association for Computing Machinery, Dec. 2022, pp. 1–11. ISBN: 978-1-45039948-7. DOI: 10.1145/3565970.3567701 (cit. on pp. 78, 79, 95).

[251] Jan Hendrik Plümer and Markus Tatzgern. "Towards a framework for validating XR prototyping for performance evaluations of simulated user experiences." In: *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. New York, USA: IEEE, Oct. 2023, pp. 810–819. DOI: 10.1109/ismar59233.2023.00096 (cit. on pp. 2, 131–136, 138, 140, 145, 148, 153, 154, 160, 163).

[252] Jan Hendrik Plümer, Kevin Yu, Ulrich Eck, Denis Kalkofen, Philipp Steininger, and Nassir Navab. "XR Prototyping of Mixed Reality Visualizations: Compensating Interaction Latency for a Medical Imaging Robot." In: *Proceedings of the 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. New York, USA: IEEE, 2024, pp. 21–25. DOI: 10.1109/ISMAR62088.2024.00014 (cit. on pp. 2, 131–135, 145, 146, 153, 160, 163).

[253] Shashi Poddar, Vipan Kumar, and Amod Kumar. "A Comprehensive Overview of Inertial Sensor Calibration Techniques." In: *J. Dyn. Syst. Meas. Contr.* 139.1 (Sept. 2016), p. 011006. ISSN: 0022-0434. DOI: 10.1115/1.4034419 (cit. on p. 17).

[254] Charles Pontonnier, Afshin Samani, Marwan Badawi, Pascal Madeleine, and Georges Dumont. "Assessing the ability of a VR-based assembly task simulation to evaluate physical risk factors." en. In: *IEEE Trans. Vis. Comput. Graph.* 20.5 (May 2014), pp. 664–674. DOI: 10.1109/tvcg.2013.252 (cit. on pp. 131, 132).

[255] Philip Pratt, Erik Mayer, Justin Vale, Daniel Cohen, Eddie Edwards, Ara Darzi, and Guang-Zhong Yang. "An effective visualisation and registration system for image-guided robotic partial nephrectomy." In: *J. Robotic Surg.* 6.1 (Mar. 2012), pp. 23–31. ISSN: 1863-2491. DOI: 10.1007/s11701-011-0334-z (cit. on p. 18).

[256] Bernhard Preim and Raimund Dachselt. *Interaktive Systeme*. de. 2nd ed. eXamen.Press. Berlin, Germany: Springer, Apr. 2015. ISBN: 978-3-642-45247-5 (cit. on pp. 11, 12, 71, 111).

[257] Alan M. Priester, Shyam Natarajan, and Martin O. Culjat. "Robotic ultrasound systems in medicine." In: *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 60.3 (Mar. 2013), pp. 507–523. DOI: 10.1109/TUFFC.2013.2593 (cit. on pp. 21, 82).

[258] David Puljiz, Katharina S. Riesterer, Björn Hein, and Torsten Kröger. "Referencing between a Head-Mounted Device and Robotic Manipulators." In: Apr. 2019. DOI: 10.48550/arXiv.1904.02480. eprint: 1904.02480 (cit. on pp. 38–41, 50, 62, 63, 65).

[259] Long Qian, Anton Deguet, and Peter Kazanzides. "ARssist: augmented reality on a head-mounted display for the first assistant in robotic surgery." In: *Healthcare Technol. Lett.* 5.5 (Oct. 2018), pp. 194–200. ISSN: 2053-3713. DOI: 10.1049/htl.2018.5065 (cit. on p. 53).

[260] Carla Maria Irene Quarato et al. "A Review on Biological Effects of Ultrasounds: Key Messages for Clinicians." In: *Diagnostics* 13.5 (Feb. 2023), p. 855. ISSN: 2075-4418. DOI: 10.3390/diagnostics13050855 (cit. on p. 20).

[261] Francis Quek, David McNeill, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. McCullough, and Rashid Ansari. "Multimodal human discourse: gesture and speech." In: *ACM Trans. Comput.-Hum. Interact.* 9.3 (Sept. 2002), pp. 171–193. ISSN: 1073-0516. DOI: 10.1145/568513.568514 (cit. on p. 12).

[262] Daniel S. Quintana and Donald R. Williams. "Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP." In: *BMC Psychiatry* 18.1 (June 2018). ISSN: 1471-244X. DOI: 10.1186/s12888-018-1761-4 (cit. on p. 27).

[263] Camilo Perez Quintero, Sarah Li, Matthew Kxj Pan, Wesley P. Chan, H. F. Machiel Van der Loos, and Elizabeth Croft. "Robot Programming Through Augmented Trajectories in Augmented Reality." In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 01–05. DOI: 10.1109/IROS.2018.8593700 (cit. on pp. 37–39, 41).

[264] Camilo Perez Quintero, Romeo Tatsambon, Mona Gridseth, and Martin Jägersand. "Visual pointing gestures for bi-directional human robot interaction in a pick-and-place task." In: *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 2015–04. DOI: 10.1109/ROMAN.2015.7333604 (cit. on p. 112).

[265] Jagdish Lal Raheja, Radhey Shyam, Umesh Kumar, and P. Bhanu Prasad. "Real-Time Robotic Hand Control Using Hand Gestures." In: *2010 Second International Conference on Machine Learning and Computing*. IEEE, pp. 09–11. DOI: 10.1109/ICMLC.2010.12 (cit. on p. 112).

[266] Md. Mijanur Rahman, Fatema Khatun, Ismat Jahan, Ramprosad Devnath, and Md. Al-Amin Bhuiyan. "Cobotics: The Evolving Roles and Prospects of Next-Generation Collaborative Robots in Industry 5.0." In: *J. Rob.* 2024.1 (Aug. 2024), p. 2918089. ISSN: 1687-9619. DOI: 10.1155/2024/2918089 (cit. on pp. 10, 71).

[267] Deepak Raina, Hardeep Singh, Subir Kumar Saha, Chetan Arora, Ayushi Agarwal, Chandrashekhara S. H., Krithika Rangarajan, and Suvayan Nandi. "Comprehensive Telerobotic Ultrasound System for Abdominal Imaging: Development and in-vivo Feasibility Study." In: *2021 International Symposium on Medical Robotics (ISMR)*. IEEE, pp. 17–19. DOI: 10.1109/ISMR48346.2021.9661578 (cit. on pp. 75, 76, 79, 80, 90, 152, 153, 159, 163).

[268] Ismo Rakkolainen, Ahmed Farooq, Jari Kangas, Jaakko Hakulinen, Jussi Rantala, Markku Turunen, and Roope Raisamo. "Technologies for Multimodal Interaction in Extended Reality—A Scoping Review." In: *Multimodal Technol. Interact.* 5.12 (Dec. 2021), p. 81. ISSN: 2414-4088. DOI: 10.3390/mti5120081 (cit. on p. 37).

[269] Philipp A. Rauschnabel, Reto Felix, Chris Hinsch, Hamza Shahab, and Florian Alt. "What is XR? Towards a Framework for Augmented and Virtual Reality." In: *Computers in Human Behavior* 133 (Aug. 2022), p. 107289. ISSN: 0747-5632. DOI: 10.1016/j.chb.2022.107289 (cit. on p. 13).

[270] Daniel J. Rea and Stela H. Seo. "Still Not Solved: A Call for Renewed Focus on User-Centered Teleoperation Interfaces." In: *Front. Rob. AI* 9 (Mar. 2022), p. 704225. ISSN: 2296-9144. DOI: 10.3389/frobt.2022.704225 (cit. on pp. 1, 21, 131).

[271] Carol E. Reiley, Takintope Akinbiyi, Darius Burschka, David C. Chang, Allison M. Okamura, and David D. Yuh. "Effects of visual force feedback on robot-assisted surgical task performance." In: *J. Thorac. Cardiovasc. Surg.* 135.1 (Jan. 2008), pp. 196–202. ISSN: 0022-5223. DOI: 10.1016/j.jtcvs.2007.08.043 (cit. on pp. 77, 78).

[272] Dominik Riedelbauch, Nico Höllerich, and Dominik Henrich. "Benchmarking Teamwork of Humans and Cobots—An Overview of Metrics, Strategies, and Tasks." In: *IEEE Access* 11 (Apr. 2023), pp. 43648–43674. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2023.3271602 (cit. on p. 135).

[273] Eric Rosen, David Whitney, Elizabeth Phillips, Gary Chien, James Tompkin, George Konidaris, and Stefanie Tellex. "Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays." In: *Int. J. Rob. Res.* 38.12-13 (Apr. 2019), pp. 1513–1526. ISSN: 0278-3649. DOI: 10.1177/0278364919842925 (cit. on pp. 38, 39, 41).

[274] Zvi Roth, Benjamin W. Mooring, and Bahram Ravani. "An overview of robot calibration." In: *IEEE Journal on Robotics and Automation* 3.5 (Jan. 2003), pp. 377–385. DOI: 10.1109/JRA.1987.1087124 (cit. on p. 17).

[275] Nina Rudigkeit, Marion Gebhard, and Axel Gräser. "Evaluation of control modes for head motion-based control with motion sensors." In: *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*. IEEE, 2015, pp. 07–09. ISBN: 978-1-4799-6477-2. DOI: 10.1109/MeMeA.2015.7145187 (cit. on pp. 111, 113, 125).

[276] Daniel Rueckert and Julia A. Schnabel. "Medical Image Registration." In: *Biomedical Image Processing*. Berlin, Germany: Springer, Oct. 2010, pp. 131–154. ISBN: 978-3-642-15816-2. DOI: 10.1007/978-3-642-15816-2_5 (cit. on p. 18).

[277] Rufat Rzayev, Sven Mayer, Christian Krauter, and Niels Henze. "Notification in VR: The Effect of Notification Placement, Task and Environment." In: *CHI PLAY '19: Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 199–211. ISBN: 978-1-45036688-5. DOI: 10.1145/3311350.3347190 (cit. on pp. 79, 95).

[278] Ofir Sadka, Jonathan Giron, Doron Friedman, Oren Zuckerman, and Hadas Erel. "Virtual-reality as a simulation tool for non-humanoid social robots." In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, Apr. 2020, 1–9. DOI: 10.1145/3334480.3382893 (cit. on p. 132).

[279] Albert Ali Salah, Javier Ruiz-del Solar, Çetin Meriçli, and Pierre-Yves Oudeyer. "Human Behavior Understanding for Robotics." In: *Human Behavior Understanding*. Berlin, Germany: Springer, 2012, pp. 1–16. ISBN: 978-3-642-34014-7. DOI: 10.1007/978-3-642-34014-7_1 (cit. on p. 21).

[280] Ane San Martín and Johan Kildal. "Audio-visual AR to Improve Awareness of Hazard Zones Around Robots." In: *ACM Conferences*. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 1–6. DOI: 10.1145/3290607.3312996 (cit. on p. 37).

[281] Tracy L. Sanders, Keith MacArthur, William Volante, Gabriella Hancock, Thomas MacGillivray, William Shugars, and P. A. Hancock. "Trust and Prior Experience in Human-Robot Interaction." In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61.1 (Sept. 2017), pp. 1809–1813. ISSN: 1071-1813. DOI: 10.1177/1541931213601934 (cit. on p. 163).

[282] Suprakas Saren, Abhishek Mukhopadhyay, Debasish Ghose, and Pradipta Biswas. "Comparing alternative modalities in the context of multimodal human–robot interaction." In: *J. Multimodal User Interfaces* 18.1 (Mar. 2024), pp. 69–85. ISSN: 1783-8738. DOI: 10.1007/s12193-023-00421-w (cit. on pp. 2, 28, 72, 73, 82, 91).

[283] Jeff Sauro and Joseph S. Dumas. "Comparison of three one-question, post-task usability questionnaires." In: *ACM Conferences*. New York, NY, USA: Association for Computing Machinery, Apr. 2009, pp. 1599–1608. DOI: 10.1145/1518701.1518946 (cit. on p. 25).

[284] Jeff Sauro and James R Lewis. *Quantifying the User Experience*. Cambridge, MA, USA: Elsevier, 2016, pp. 1–8. ISBN: 978-0-12-802308-2 (cit. on p. 24).

[285] Kristin E. Schaefer. "Measuring Trust in Human Robot Interactions: Development of the "Trust Perception Scale-HRI"." In: *Robust Intelligence and Trust in Autonomous Systems*. Boston, MA, USA: Springer, Boston, MA, Apr. 2016, pp. 191–218. ISBN: 978-1-4899-7668-0. DOI: 10.1007/978-1-4899-7668-0_10 (cit. on pp. 25, 152).

[286] Dieter Schmalstieg and Tobias Hollerer. *Augmented reality*. en. Boston, MA, USA: Addison-Wesley Educational, June 2016. ISBN: 978-0-32188357-5 (cit. on pp. 12–19).

[287] Xenia Schmalz, Jose Biurrun Manresa, and Lei Zhang. "What is a Bayes factor?" In: *Psychological methods* 28.3 (2023), p. 705. DOI: 10.1037/met0000421 (cit. on p. 26).

[288] Tanner Schmidt, Richard Newcombe, and Dieter Fox. "DART: dense articulated real-time tracking with consumer depth cameras." In: vol. 39. 3. Kluwer Academic Publishers, Oct. 2015, pp. 239–258. DOI: 10.1007/s10514-015-9462-z (cit. on p. 62).

[289] Dominik Schön, Thomas Kosch, Florian Müller, Martin Schmitz, Sebastian Günther, Lukas Bommhardt, and Max Mühlhäuser. "Tailor Twist: Assessing Rotational Mid-Air Interactions for Augmented Reality." In: *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, Apr. 2023, pp. 1–14. ISBN: 978-1-45039421-5. DOI: 10.1145/3544548.3581461 (cit. on pp. 112, 124).

[290] Samuel B. Schorr, Zhan Fan Quek, Ilana Nisky, William R. Provancher, and Allison M. Okamura. "Tactor-Induced Skin Stretch as a Sensory Substitution Method in Teleoperated Palpation." In: *IEEE Trans. Hum.-Mach. Syst.* 45.6 (Aug. 2015), pp. 714–726. DOI: 10.1109/THMS.2015.2463090 (cit. on pp. 77, 78).

[291] Danny Schott, Florian Heinrich, Lara Stallmeister, Julia Moritz, Bennet Hensen, and Christian Hansen. "Is this the vReal Life? Manipulating Visual Fidelity of Immersive Environments for Medical Task Simulation." In: *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, pp. 16–20. DOI: 10.1109/ISMAR59233.2023.00134 (cit. on pp. 131, 146, 161).

[292] Josefine Schreiter, Fabian Joeres, Christine March, Maciej Pech, and Christian Hansen. "Application Potential of Robot-Guided Ultrasound During CT-Guided Interventions." In: *Simplifying Medical Ultrasound*. Cham, Switzerland: Springer, Sept. 2021, pp. 116–125. ISBN: 978-3-030-87583-1. DOI: 10.1007/978-3-030-87583-1_12 (cit. on p. 136).

[293] Josefine Schreiter, Vladimir Semshchikov, Magnus Hanses, Norbert Elkmann, and Christian Hansen. "Towards a real-time control of robotic ultrasound using haptic force feedback." In: *Current Directions in Biomedical Engineering* 8.1 (July 2022), pp. 81–84. ISSN: 2364-5504. DOI: 10.1515/cdbme-2022-0021 (cit. on p. 91).

[294] Lovis Schwenderling, Anna Kleinau, Wilhelm Herbrich, Haswanth Kasireddy, Florian Heinrich, and Christian Hansen. "Activation modes for gesture-based interaction with a magic lens in AR anatomy visualisation." In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* (July 2023). ISSN: 2168-1163. DOI: 10.1080/21681163.2022.2157749 (cit. on p. 28).

[295] Lovis Schwenderling, Maximilian Schotte, Fabian Joeres, Florian Heinrich, Laura Isabel Hanke, Florentine Huettl, Tobias Huber, and Christian Hansen. "Teach Me Where to Look: Dual-task Attention Training in Augmented Reality." In: *ACM Conferences*. New York, NY, USA: Association for Computing Machinery, Apr. 2025, pp. 1–8. DOI: 10.1145/3706599.3720198 (cit. on p. 108).

[296] A. Scorza, S. Conforto, C. D'Anna, and S. A. Sciuto. "A Comparative Study on the Influence of Probe Placement on Quality Assurance Measurements in B-mode Ultrasound by Means of Ultrasound Phantoms." In: *Open Biomed. Eng. J.* 9 (July 2015), p. 164. DOI: 10.2174/1874120701509010164 (cit. on p. 20).

[297] Peter Karl Seitz, Beatrice Baumann, Wibke Johnen, Cord Lissek, Johanna Seidel, and Rolf Bendl. "Development of a robot-assisted ultrasound-guided radiation therapy (USgRT)." In: *Int. J. CARS* 15 (2020), pp. 491–501. DOI: 10.1007/s11548-019-02104-y (cit. on pp. 21, 76, 79, 80, 83).

[298] Mario Selvaggio, Marco Cognetti, Stefanos Nikolaidis, Serena Ivaldi, and Bruno Siciliano. "Autonomy in Physical Human-Robot Interaction: A Brief Survey." In: *IEEE Rob. Autom. Lett.* 6.4 (July 2021), pp. 7989–7996. DOI: 10.1109/LRA.2021.3100603 (cit. on p. 9).

[299] Stela H Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E Young. "Poor Thing! Would You Feel Sorry for a Simulated Robot?" In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. Portland Oregon USA: ACM, Mar. 2015, 125–132. DOI: 10.1145/2696454.2696471 (cit. on p. 132).

[300] Helen Sharp, Yvonne Rogers, and Jenny Preece. *Interaction Design*. en. 2nd ed. Chichester, England: John Wiley & Sons, Jan. 2007. ISBN: 978-0-47001866-8 (cit. on p. 24).

[301] F. Sherwani, Muhammad Mujtaba Asad, and B. S. K. K. Ibrahim. "Collaborative Robots and Industrial Revolution 4.0 (IR 4.0)." In: *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*. IEEE, pp. 26–27. DOI: 10.1109/ICETST49965.2020.9080724 (cit. on pp. 1, 10, 71).

[302] K. Kirk Shung. "Diagnostic Ultrasound: Past, Present, and Future." In: *Journal of Medical and Biological Engineering* 31.6 (2011), pp. 371–374. ISSN: 1609-0985 (cit. on p. 20).

[303] Bruno Siciliano and Oussama Khatib, eds. *Springer Handbook of Robotics*. en. 2008th ed. Springer Handbook of Robotics. Berlin, Germany: Springer, May 2008. ISBN: 978-3-319-32550-7 (cit. on p. 8).

[304] Bruno Siciliano, Lorenzo Sciavicco, Luigi Villani, and Giuseppe Oriolo. *Robotics*. en. 1st ed. Advanced Textbooks in Control and Signal Processing. London, England: Springer, Nov. 2008. ISBN: 978-1846286414 (cit. on pp. 7, 8).

[305] Shaishav Siddhpuria, Keiko Katsuragawa, James R. Wallace, and Edward Lank. "Exploring At-Your-Side Gestural Interaction for Ubiquitous Environments." In: *DIS '17: Proceedings of the 2017 Conference on Designing Interactive Systems*. New York, NY, USA: Association for Computing Machinery, June 2017, pp. 1111–1122. ISBN: 978-1-45034922-2. DOI: 10.1145/3064663.3064695 (cit. on pp. 112, 114).

[306] Anton Simorov, R. Stephen Otte, Courtni M. Kopietz, and Dmitry Oleynikov. "Review of surgical robotics user interface: what is the best way to control robotic surgery?" In: *Surg. Endosc.* 26.8 (Aug. 2012), pp. 2117–2125. ISSN: 1432-2218. DOI: 10.1007/s00464-012-2182-y (cit. on p. 71).

[307] Richard Skarbez, Missie Smith, and Mary C. Whitton. "Revisiting Milgram and Kishino's Reality-Virtuality Continuum." In: *Front. Virtual Real.* 2 (Mar. 2021), p. 647997. ISSN: 2673-4192. DOI: 10.3389/frvir.2021.647997 (cit. on p. 13).

[308] Tianyu Song, Ulrich Eck, and Nassir Navab. "Optimizing In-Contact Force Planning in Robotic Ultrasound with Augmented Reality Visualization Techniques." In: *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, pp. 21–25. DOI: 10.1109/ISMAR62088.2024.00070 (cit. on pp. 4, 75, 77, 78, 107, 162).

[309] Tianyu Song, Felix Pabst, Ulrich Eck, and Nassir Navab. "Enhancing Patient Acceptance of Robotic Ultrasound through Conversational Virtual Agent and Immersive Visualizations." In: *Transactions on Visualization and Computer Graphics (TVCG)* 31.5 (Mar. 2025), pp. 2901–2911. DOI: 10.1109/TVCG.2025.3549181 (cit. on p. 37).

[310] Christian Stetco, Stephan Mühlbacher-Karrer, Matteo Lucchi, Matthias Weyrer, Lisa-Marie Faller, and Hubert Zangl. "Gesture-based Contactless Control of Mobile Manipulators using Capacitive Sensing." In: *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, pp. 25–28. DOI: 10.1109/I2MTC43012.2020.9128751 (cit. on p. 113).

[311] Ryo Suzuki, Adnan Karim, Tian Xia, Hooman Hedayati, and Nicolai Marquardt. "Augmented Reality and Robotics: A Survey and Taxonomy for AR-enhanced Human-Robot Interaction and Robotic Interfaces." In: *ACM Conferences*. New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 1–33. DOI: 10.1145/3491102.3517719 (cit. on pp. 1, 13, 29, 37, 78, 94, 171).

[312] Toqeer Ali Syed, Muhammad Shoaib Siddiqui, Hurria Binte Abdullah, Salman Jan, Abdallah Namoun, Ali Alzahrani, Adnan Nadeem, and Ahmad B. Alkhodre. "In-Depth Review of Augmented Reality: Tracking Technologies, Development Tools, AR Displays, Collaborative AR, and Security Concerns." In: *Sensors* 23.1 (Dec. 2022), p. 146. ISSN: 1424-8220. DOI: 10.3390/s23010146 (cit. on p. 16).

[313] Leila Takayama, Wendy Ju, and Clifford Nass. "Beyond dirty, dangerous and dull: What everyday people think robots should do." In: *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots* (Mar. 2008), pp. 25–32. ISSN: 2167-2121. DOI: 10.1145/1349822.1349827 (cit. on p. 1).

[314] Ali Talasaz, Ana Luisa Trejos, and Rajni V. Patel. "The Role of Direct and Visual Force Feedback in Suturing Using a 7-DOF Dual-Arm Teleoperated System." In: *IEEE Trans. Haptics* 10.2 (Oct. 2016), pp. 276–287. DOI: 10.1109/TOH.2016.2616874 (cit. on pp. 4, 75, 77, 78, 152, 158).

[315] Robert J. Teather and I. Scott MacKenzie. "Position vs. Velocity Control for Tilt-Based Interaction." In: *Taylor & Francis* (Nov. 2020), pp. 51–58. DOI: 10.1201/9781003059325-7 (cit. on p. 113).

[316] Minh Q. Tram, Joseph M. Cloud, and William J. Beksi. "Intuitive Robot Integration via Virtual Reality Workspaces." In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2023–02. DOI: 10.1109/ICRA48891.2023.10160699 (cit. on pp. 111, 112, 131).

[317] Amedeo Troiano, Francesco Naddeo, Erik Sosso, Gianfranco Camarota, Roberto Merletti, and Luca Mesin. "Assessment of force and fatigue in isometric contractions of the upper trapezius muscle by surface EMG signal and perceived exertion scale." In: *Gait Posture* 28.2 (Aug. 2008), pp. 179–186. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2008.04.002 (cit. on p. 24).

[318] Tom Tullis and Bill Albert. *Measuring the User Experience*. Morgan Kaufmann, 2013. ISBN: 978-0-12-415781-1. DOI: 10.1016/C2011-0-00016-9 (cit. on p. 22).

[319] Marcus Tönnis. *Augmented Reality*. de. 2010th ed. Informatik Im Fokus. Berlin, Germany: Springer, Aug. 2010. ISBN: 978-3642141782 (cit. on pp. 14–17, 63).

[320] Sayako Ueda, Ryoichi Nakashima, and Takatsune Kumada. "Influence of levels of automation on the sense of agency during continuous action." In: *Sci. Rep.* 11.2436 (Jan. 2021), pp. 1–13. ISSN: 2045-2322. DOI: 10.1038/s41598-021-82036-3 (cit. on p. 153).

[321] Dorin Ungureanu et al. "HoloLens 2 Research Mode as a Tool for Computer Vision Research." In: *arXiv* (Aug. 2020). DOI: 10.48550/arXiv.2008.11239 (cit. on p. 29).

[322] Universal Robots A/S. *Universal Robots e-Series User Manual: Original Instruction*. Original Instruction. Universal Robots A/S. Denmark, 2020 (cit. on p. 38).

[323] Jan BF Van Erp et al. "Guidelines for the use of vibro-tactile displays in human computer interaction." In: *Proceedings of eurohaptics*. Vol. 2002. Citeseer. 2002, pp. 18–22 (cit. on p. 108).

[324] Karl Van Wyk, Joe Falco, and Geraldine Cheok. "Efficiently Improving and Quantifying Robot Accuracy In Situ." In: *IEEE transactions on automation science and engineering : a publication of the IEEE Robotics and Automation Society* 1 (2019), 10.48550/arXiv.1908.07273. DOI: 10.48550/arXiv.1908.07273 (cit. on p. 63).

[325] Karl Van Wyk and Jeremy A. Marvel. "Strategies for Improving and Evaluating Robot Registration Performance." In: *IEEE Trans. Autom. Sci. Eng.* 15.1 (July 2017), pp. 320–328. DOI: 10.1109/TASE.2017.2720478 (cit. on pp. 38–40, 42, 50, 52, 62–65).

[326] Louise Veling and Conor McGinn. "Qualitative Research in HRI: A Review and Taxonomy." In: *Int. J. Social Rob.* 13.7 (Nov. 2021), pp. 1689–1709. ISSN: 1875-4805. DOI: 10.1007/s12369-020-00723-z (cit. on p. 24).

[327] Federico Vicentini. "Collaborative robotics: a survey." In: *Journal of Mechanical Design* 143.4 (2021), p. 040802. DOI: 10.1115/1.4046238 (cit. on pp. 80, 81).

[328] Valeria Villani, Lorenzo Sabattini, Giuseppe Riggio, Cristian Secchi, Marco Minelli, and Cesare Fantuzzi. "A Natural Infrastructure-Less Human–Robot Interaction System." In: *IEEE Rob. Autom. Lett.* 2.3 (Mar. 2017), pp. 1640–1647. DOI: 10.1109/LRA.2017.2678541 (cit. on pp. 111, 113).

[329] Christian Vogel, Christoph Walter, and Norbert Elkmann. "Safeguarding and Supporting Future Human-robot Cooperative Manufacturing Processes by a Projection- and Camera-based Technology." In: *Procedia Manuf.* 11 (Jan. 2017), pp. 39–46. ISSN: 2351-9789. DOI: 10.1016/j.promfg.2017.07.127 (cit. on p. 37).

[330] Alexandra Voit, Sven Mayer, Valentin Schwind, and Niels Henze. "Online, VR, AR, lab, and in-situ." In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, May 2019, 1–12. DOI: 10.1145/3290605.3300737 (cit. on pp. 131, 132, 145).

[331] Juan Pablo Wachs, Mathias Kölsch, Helman Stern, and Yael Edan. "Vision-based hand-gesture applications." In: *Commun. ACM* 54.2 (Feb. 2011), pp. 60–71. ISSN: 0001-0782. DOI: 10.1145/1897816.1897838 (cit. on p. 1).

[332] Lihui Wang, Sichao Liu, Hongyi Liu, and Xi Vincent Wang. "Overview of Human-Robot Collaboration in Manufacturing." In: *Proceedings of 5th International Conference on the Industry 4.0 Model for Advanced Manufacturing*. Cham, Switzerland: Springer, May 2020, pp. 15–58. ISBN: 978-3-030-46212-3. DOI: 10.1007/978-3-030-46212-3_2 (cit. on pp. 11, 12).

[333] Tian Wang, Pai Zheng, Shufei Li, and Lihui Wang. "Multimodal Human–Robot Interaction for Human-Centric Smart Manufacturing: A Survey." In: *Adv. Intell. Syst.* 6.3 (Mar. 2024), p. 2300359. ISSN: 2640-4567. DOI: 10.1002/aisy.202300359 (cit. on p. 174).

[334] Xian Wang, Luyao Shen, and Lik-Hang Lee. "A Systematic Review of XR-Enabled Remote Human-Robot Interaction Systems." In: *ACM Comput. Surv.* 57.11 (June 2025), pp. 1–37. ISSN: 0360-0300. DOI: 10.1145/3730574 (cit. on pp. 12, 71).

[335] Xihao Wang, Hao Shen, Hui Yu, Jielong Guo, and Xian Wei. "Hand and Arm Gesture-based Human-Robot Interaction: A Review." In: *ACM Other conferences*. New York, NY, USA: Association for Computing Machinery, Sept. 2022, pp. 1–7. DOI: 10.1145/3564982.3564996 (cit. on p. 11).

[336] Ziheng Wang, Isabella Reed, and Ann Majewicz Fey. "Toward Intuitive Teleoperation in Surgery: Human-Centric Evaluation of Teleoperation Algorithms for Robotic Needle Steering." In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 21–25. DOI: 10.1109/ICRA.2018.8460729 (cit. on pp. 111, 113, 125).

[337] Maximilian Weiß, Katrin Angerbauer, Alexandra Voit, Magdalena Schwarzl, Michael Sedlmair, and Sven Mayer. "Revisited: Comparison of empirical methods to evaluate visualizations supporting crafting and assembly purposes." In: *Transactions on Visualization and Computer Graphics (TVCG)* 27.2 (2020), pp. 1204–1213. DOI: 10.1109/tvcg.2020.3030400 (cit. on pp. 131, 132).

[338] Vincent Weistroffer, Alexis Paljic, Lucile Callebert, and Philippe Fuchs. "A methodology to assess the acceptability of human-robot collaboration using virtual reality." In: *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*. Singapore: ACM, Oct. 2013, 39–48. DOI: 10.1145/2503713.2503726 (cit. on pp. 2, 131).

[339] Vincent Weistroffer, Alexis Paljic, Philippe Fuchs, Olivier Hugues, Jean-Paul Chodacki, Pascal Ligot, and Alexandre Morais. "Assessing the acceptability of human-robot co-presence on assembly lines: A comparison between actual situations and their virtual reality counterparts." In: *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. New York, USA: IEEE, 2014, pp. 377–384. DOI: 10.1109/roman.2014.6926282 (cit. on pp. 131–135, 152, 153).

[340] Greg Welch and Gary Bishop. *An Introduction to the Kalman Filter*. Tech. rep. USA, 1995 (cit. on pp. 19, 97).

[341] Jay B. West, J. Michael Fitzpatrick, Steven A. Toms, Calvin R. Maurer Jr., and Robert J. Maciunas. "Fiducial Point Placement and the Accuracy of Point-based, Rigid Body Registration." In: *Neurosurgery* 48.4 (Apr. 2001), p. 810. ISSN: 0148-396X (cit. on pp. 39, 40, 49, 50, 52, 63, 64).

[342] Ruud Wetzels, Dora Matzke, Michael D. Lee, Jeffrey N. Rouder, Geoffrey J. Iverson, and Eric-Jan Wagenmakers. "Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests." In: *Perspect. Psychol. Sci.* 6.3 (May 2011), pp. 291–298. ISSN: 1745-6916. DOI: 10.1177/1745691611406923 (cit. on pp. 25, 26).

[343] David Whitney, Eric Rosen, Elizabeth Phillips, George Konidaris, and Stefanie Tellex. "Comparing Robot Grasping Teleoperation Across Desktop and Virtual Reality with ROS Reality." In: *Robotics Research*. Cham, Switzerland: Springer, Nov. 2019, pp. 335–350. ISBN: 978-3-030-28619-4. DOI: 10.1007/978-3-030-28619-4_28 (cit. on pp. 72, 73, 131–134).

[344] David Whitney, Eric Rosen, Daniel Ullman, Elizabeth Phillips, and Stefanie Tellex. "ROS Reality: A Virtual Reality Framework Using Consumer-Grade Hardware for ROS-Enabled Robots." In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 01–05. DOI: 10.1109/IROS.2018.8593513 (cit. on pp. 132, 133).

[345] Christopher D Wickens, Sallie Gordon-Becker, and Yili D Liu. *An Introduction to Human Factors Engineering*. en. Upper Saddle River, NJ: Pearson, Dec. 1997. ISBN: 978-0131837362 (cit. on p. 23).

[346] Daniel Wigdor and Dennis Wixon. *Brave NUI World*. Burlington, USA: Morgan Kaufmann, 2011. ISBN: 978-0-12-382231-4. DOI: 10.1016/C2009-0-64091-5 (cit. on p. 11).

[347] Luc Wijnen, Séverin Lemaignan, and Paul Bremner. "Towards using virtual reality for replicating HRI studies." In: *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. Cambridge, UK: ACM, Mar. 2020, 514–516. DOI: 10.1145/3371382.3378374 (cit. on pp. 132, 133).

[348] Rand R. Wilcox. "Chapter 8 - Comparing Multiple Dependent Groups." In: (2022). Ed. by Rand R. Wilcox, pp. 467–539. DOI: 10.1016/B978-0-12-820098-8.00014-2 (cit. on pp. 46, 85, 100, 119, 140, 154).

[349] Tom Williams, Leanne Hirshfield, Nhan Tran, Trevor Grant, and Nicholas Woodward. "Using augmented reality to better study human-robot interaction." en. In: *Lecture Notes in Computer Science*. Lecture notes in computer science. Cham: Springer International Publishing, 2020, pp. 643–654. DOI: 10.1007/978-3-030-49695-1_43 (cit. on p. 131).

[350] Sebastian Wojcinski, Kathrin Brandhorst, Gelareh Sadigh, Peter Hillemanns, and Friedrich Degenhardt. "Acoustic Radiation Force Impulse Imaging with Virtual Touch Tissue Quantification: Measurements of Normal Breast Tissue and Dependence on the Degree of Pre-compression." In: *Ultrasound Med. Biol.* 39.12 (Dec. 2013), pp. 2226–2232. ISSN: 0301-5629. DOI: 10.1016/j.ultrasmedbio.2013.06.014 (cit. on p. 4).

[351] Murphy Wonsick and Taskin Padir. "A Systematic Review of Virtual Reality Interfaces for Controlling and Interacting with Robots." In: *Appl. Sci.* 10.24 (Dec. 2020), p. 9051. ISSN: 2076-3417. DOI: 10.3390/app10249051 (cit. on pp. 13, 37).

[352] Julia Woodward and Jaime Ruiz. "Analytic Review of Using Augmented Reality for Situational Awareness." In: *Transactions on Visualization and Computer Graphics (TVCG)* 29.4 (Jan. 2022), pp. 2166–2183. DOI: 10.1109/TVCG.2022.3141585 (cit. on p. 78).

[353] World Medical Association. "World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects." In: *JAMA* 310.20 (Nov. 2013), pp. 2191–2194. ISSN: 1538-3598. DOI: 10.1001/jama.2013.281053 (cit. on p. 33).

[354] Adam Grant Worrallo and Thomas Hartley. "Robust Optical Based Hand Interaction for Virtual Reality." In: *Transactions on Visualization and Computer Graphics (TVCG)* 28.12 (May 2021), pp. 4186–4197. DOI: 10.1109/TVCG.2021.3083411 (cit. on p. 29).

[355] Wan-Chen Wu, Cagatay Basdogan, and Mandayam A Srinivasan. "Visual, haptic, and bimodal perception of size and stiffness in virtual environments." In: *ASME International Mechanical Engineering Congress and Exposition*. Vol. 16349. American Society of Mechanical Engineers. 1999, pp. 19–26 (cit. on pp. 134, 148, 153).

[356] Yuqiang Wu, Pietro Balatti, Marta Lorenzini, Fei Zhao, Wansoo Kim, and Arash Ajoudani. "A Teleoperation Interface for Loco-Manipulation Control of Mobile Collaborative Robotic Assistant." In: *IEEE Rob. Autom. Lett.* 4.4 (July 2019), pp. 3593–3600. DOI: 10.1109/LRA.2019.2928757 (cit. on pp. 1, 28, 71–73).

[357] Xiangjie Yan, Chen Chen, and Xiang Li. "Adaptive Vision-Based Control of Redundant Robots with Null-Space Interaction for Human-Robot Collaboration." In: *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 23–27. DOI: 10.1109/ICRA46639.2022.9812218 (cit. on pp. 1, 112).

[358] Xiangjie Yan, Yongpeng Jiang, Guokun Wu, Chen Chen, Gao Huang, and Xiang Li. "Multi-Modal Interaction Control of Ultrasound Scanning Robots with Safe Human Guidance and Contact Recovery." In: *arXiv preprint arXiv:2302.05685* (2023). DOI: 10.48550/arXiv.2302.05685 (cit. on p. 21).

[359] Xiangjie Yan, Yongpeng Jiang, Guokun Wu, Chen Chen, Gao Huang, and Xiang Li. *Multi-Modal Interaction Control of Ultrasound Scanning Robots with Safe Human Guidance and Contact Recovery*. 2023. arXiv: 2302.05685 [cs.RO]. URL: https://arxiv.org/abs/2302.05685 (cit. on p. 136).

[360] Chenguang Yang, Jing Luo, and Ning Wang. *Human-In-the-loop Learning and Control for Robot Teleoperation*. Elsevier, Academic Press, 2023. ISBN: 978-0-323-95143-2. DOI: 10.1016/C2021-0-02620-1 (cit. on p. 1).

[361] Guang-Zhong Yang et al. *Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy*. Mar. 2017. DOI: 10.1126/scirobotics.aam8638 (cit. on pp. 10, 71).

[362] Ziv Yaniv. *Rigid Registration*. Jan. 2008. ISBN: 978-0-387-73856-7. DOI: 10.1007/978-0-387-73858-1_6 (cit. on pp. 18, 19).

[363] Ziv Yaniv. "Which pivot calibration?" In: *Medical imaging 2015: Image-guided procedures, robotic interventions, and modeling*. Vol. 9415. SPIE. 2015, pp. 542–550 (cit. on pp. 17, 19, 30, 44, 53).

[364] Shih-Ching Yeh, Eric Hsiao-Kuang Wu, Ying-Ru Lee, R. Vaitheeshwari, and Chen-Wei Chang. "User Experience of Virtual-Reality Interactive Interfaces: A Comparison between Hand Gesture Recognition and Joystick Control for XRSPACE MANOVA." In: *Appl. Sci.* 12.23 (Nov. 2022), p. 12230. ISSN: 2076-3417. DOI: 10.3390/app122312230 (cit. on p. 71).

[365] Sierra N. Young and Joshua M. Peschel. "Review of Human–Machine Interfaces for Small Unmanned Systems With Robotic Manipulators." In: *IEEE Trans. Hum.-Mach. Syst.* 50.2 (Feb. 2020), pp. 131–143. DOI: 10.1109/THMS.2020.2969380 (cit. on p. 10).

[366] Hui Yu, Shengzhi Du, Anish Kurien, Barend Jacobus van Wyk, and Qingxue Liu. "The Sense of Agency in Human–Machine Interaction Systems." In: *Appl. Sci.* 14.16 (Aug. 2024), p. 7327. ISSN: 2076-3417. DOI: 10.3390/app14167327 (cit. on p. 153).

[367] Ahmad Anas Yusof, Takuya Kawamura, and Hironao Yamada. "Evaluation of Construction Robot Telegrasping Force Perception Using Visual, Auditory and Force Feedback Integration." In: *Journal of Robotics and Mechatronics* 24.6 (Dec. 2012), pp. 949–957. DOI: 10.20965/jrm.2012.p0949 (cit. on pp. 77, 78, 94).

[368] Oliver Zettinig, Benjamin Frisch, Salvatore Virga, Marco Esposito, Anna Rienmüller, Bernhard Meyer, Christoph Hennersperger, Yu-Mi Ryang, and Nassir Navab. "3D ultrasound registration-based visual servoing for neurosurgical navigation." In: *Int. J. CARS* 12.9 (Sept. 2017), pp. 1607–1619. ISSN: 1861-6429. DOI: 10.1007/s11548-017-1536-2 (cit. on pp. 21, 80, 83).

[369] S. Zhai and P. Milgram. "Human performance evaluation of manipulation schemes in virtual environments." In: *Proceedings of IEEE Virtual Reality Annual International Symposium*. IEEE, pp. 18–22. ISBN: 978-0-7803-1363. DOI: 10.1109/VRAIS.1993.380784 (cit. on p. 12).

[370] Guanyi Zhao, Chao Zeng, Weiyong Si, and Chenguang Yang. "A human-robot collaboration method for uncertain surface scanning." In: *CAAI Trans. Intell. Technol.* n/a.n/a (May 2023). ISSN: 2468-2322. DOI: 10.1049/cit2.12227 (cit. on pp. 75, 76, 79, 80, 90, 152, 159, 162, 163).

[371] Wei Zhao, Juliang Xiao, Sijiang Liu, Saixiong Dou, and Haitao Liu. "Robotic direct grinding for unknown workpiece contour based on adaptive constant force control and human–robot collaboration." In: *Industrial Robot* 50.3 (Apr. 2023), pp. 376–384. ISSN: 0143-991X. DOI: 10.1108/IR-01-2022-0021 (cit. on p. 4).

[372] Xiaofeng Zhao, Esma Ersoy, and Dianna L. Ng. "Comparison of low-cost phantoms for ultrasound-guided fine-needle aspiration biopsy training." In: *Journal of the American Society of Cytopathology* 12.4 (July 2023), pp. 275–283. ISSN: 2213-2945. DOI: 10.1016/j.jasc.2023.03.005 (cit. on p. 32).

[373] Tian Zhou, Maria E. Cabrera, Juan P. Wachs, Thomas Low, and Chandru Sundaram. "A comparative study for telerobotic surgery using free hand gestures." In: *J. Hum.-Robot. Interact.* 5.2 (Sept. 2016), pp. 1–28. DOI: 10.5898/JHRI.5.2.Zhou (cit. on pp. 1, 2, 28, 72, 73, 82, 91, 111).

[374] Jing Zhu, Yutao Chen, Ming Xu, Erbao Dong, Hao Zhang, and Xuming Tang. "Graphical Force and Haptic Feedback Teleoperation System for Live Power Lines Maintaining Robot." In: *2019 IEEE International Conference on Mechatronics and Automation (ICMA)*. IEEE, 2019, pp. 04–07. DOI: 10.1109/ICMA.2019.8816455 (cit. on pp. 75, 77).

# EHRENERKLÄRUNG

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; verwendete fremde und eigene Quellen sind als solche kenntlich gemacht. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Ich habe insbesondere nicht wissentlich:

- Ergebnisse erfunden oder widersprüchliche Ergebnisse verschwiegen,
- statistische Verfahren absichtlich missbraucht, um Daten in ungerechtfertigter Weise zu interpretieren,
- fremde Ergebnisse oder Veröffentlichungen plagiiert,
- fremde Forschungsergebnisse verzerrt wiedergegeben.

Mir ist bekannt, dass Verstöße gegen das Urheberrecht Unterlassungs- und Schadensersatzansprüche des Urhebers sowie eine strafrechtliche Ahndung durch die Strafverfolgungsbehörden begründen kann. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht.

*Magdeburg, den 14.08.2025*

Tonia Mielke

## APPLICATION OF NATURAL LANGUAGE PROCESSING

Generative AI systems (specifically ChatGPT-4) and the writing assistance tool Grammarly were intentionally used to support the linguistic refinement of this dissertation. Their use was limited to improving grammar, phrasing, and clarity of expression. No AI-generated content was used in the development of ideas, scientific arguments, or original contributions. All AI-assisted edits were critically reviewed and verified by the author to ensure full compliance with the principles of good scientific practice.

## COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/