# Human-Robot Interfaces to Enable Effective and Effortless Control for Remote Manipulation of Tele-nursing Robot

by

Tsung-Chi Lin

A Dissertation submitted to the Faculty of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Robotics Engineering

Committee in charge:

Prof. Jane Li, . . . . . Robotics Engineering, WPI (Advisor)

Prof. Jing Xiao, . . . . . Robotics Engineering, WPI (Department Head)

Prof. Cagdas Onal, . . . . . Robotics Engineering, WPI

Prof. He Huang, . . . . . Joint Department of Biomedical Engineering, UNC/NC

Prof. Berk Calli, . . . . . Robotics Engineering, WPI

April 2023

Worcester, Massachusetts

# Human-Robot Interfaces to Enable Effective and Effortless Control for Remote Manipulation of Tele-nursing Robot

Tsung-Chi Lin

## ABSTRACT

Tele-nursing robots offer a safe approach to patient care in quarantine areas during pandemics, reducing infection risks for healthcare workers and patients. They may also address nursing shortages and improve healthcare viability by providing remote care in various settings. However, remote manipulation is a challenging task due to limited feedback, latency, lack of sensory feedback, limited dexterity, and environmental factors. These factors make it difficult to control objects remotely with accuracy and precision, requiring advanced technology and skilled operators to achieve successful outcomes. In this dissertation, we investigate human-robot interfaces for remote manipulation of tele-nursing robots, which aims to enable effective and effortless control.

This dissertation addresses challenges related to remote manipulation for tele-nursing robots, focusing on developing workload-adaptive teleoperation interfaces for efficient and intuitive control, designing human-preferred remote manipulation assistance based on natural human perception-action coordination for improved user adaptivity, and additionally investigating the impact of teleoperation assistance reliability on workload and user preference.

We advance the design of human-robot collaboration for remote manipulation, enabling humans and robots to contribute their strengths. We further develop objective engagement and workload estimation methods using human motion and gaze tracking. The developed human-robot interfaces are effective across various remote manipulation systems. The dissertation concludes by summarizing key takeaways and identifying potential directions for future research.

*To my family. Especially to my beloved grandparents, whose unwavering love and support throughout my life have inspired me to pursue my dreams and strive for excellence.*

# Acknowledgments

First and foremost, I am deeply grateful to my advisor, Jane, for her invaluable guidance and continuous support throughout my PhD, and for the countless hours she has dedicated to helping me achieve my goals. I would also like to thank my dissertation committee, Professors Jing Xiao, Cagdas Onal, He Huang, and Berk Calli, for their constructive feedback and valuable insights, which have greatly improved the quality of this dissertation. And my PhD qualifier committee members, Professors Loris Fichera and Haichong Zhang, for helping me develop a stronger research agenda in the early stage of my PhD. All the user studies with nursing participants in this dissertation would not have been possible without collaborating with Worcester State University Nursing. I would like to especially thank the former Nursing head, Dr. Paula Bylaska-Davies, for providing unlimited support. I am also grateful to the registered nurses and nursing students who participated in the user studies, particularly during the COVID-19 pandemic.

I would like to thank my colleagues at HiRo lab, Alexandra, Kenechukwu, Zhaoyuan, Zhuoyun, Lorena and Nikita, for their camaraderie, support, and helpful discussions, which have enriched my graduate school experience and helped me develop new ideas. For the work in this dissertation, I enjoyed working with Achyuthan, for your enthusiasm and hard work on the projects we undertook. I would also like to thank all of the undergraduate and master's students that I worked with during my PhD, for your patience as I learned how to best advise you.

I am immensely grateful to my family in Taiwan for their unwavering support and encouragement throughout my PhD. And my soulmate, ChiChi, thank you for your unconditional love and support that meant the world to me. I could not have accomplished this without you. The therapeutic presence of my three lovely cats has also been irreplaceable. Lastly, I can't wait to embark on a new chapter with my baby girl!

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

The ability of robots to handle dull, dirty, and dangerous tasks is undoubtedly one of their greatest advantages. However, robots still lack the cognitive abilities of humans and may encounter unexpected situations (e.g., malfunctions, unforeseen obstacles, and interactions with humans) that they are not programmed to handle. Therefore, human-robot interaction is necessary to monitor and control the robot's actions, provide additional information or instructions, and intervene when needed to ensure that the robot operates safely and effectively. Some of the most significant challenges in human-robot interaction are ensuring safety, trust, communication, adaptability, and user interface. Safety is crucial, as robots can pose physical risks to humans if they malfunction or are used improperly [5]. Humans must trust robots to behave predictably and reliably, and this trust must be established through effective communication and transparency of the robot's actions [6]. Robots need to communicate with humans in a way that is clear and easy to understand [7], which requires developing natural language processing and other communication tools. Adaptability is essential, as robots must be flexible and able to adjust to changing environments and situations [8]. Effective user interface design is critical in enabling humans to control and monitor robots' actions [9]. With the increasing utilization of robots in various industries, it is crucial to prioritize the seamless integration of human-robot collaboration into existing workflows, which allows for efficient and collaborative work between humans and robots, ultimately leading to optimal outcomes.

***Motivation of This Work —*** While the challenges of human-robot interaction are complex, robots are increasingly being developed to address specific needs in various fields. Tele-nursing robots

have emerged as a promising solution for handling dangerous tasks in quarantine areas [10], particularly during times of pandemics such as COVID-19 [11] and Ebola [12]. These robots offer a safe approach to patient care, allowing healthcare providers to remotely monitor and communicate with patients without the need for direct physical contact. This can significantly reduce the risk of infection for both healthcare workers and patients as well as help prevent the spread of contagious diseases. Moreover, tele-nursing robots offer a potential solution to address the shortage of nursing workers and improve long-term healthcare viability for an aging society by providing remote care in hospitals, homes, and nursing facilities. Currently, most of the contemporary nursing robots (e.g., Intouch RT-Vita [10, 13], Ava [14], BeamPro 2 [15], Vecna VGo [16]) are limited to providing just mobility and telepresence. A few advanced robot prototypes (see Figure 1.1) can perform nursing tasks that require manipulation capability and mobility. However, even equipped with state-of-the-art autonomy, these robots cannot perform nursing tasks at operational speed, or handle high-complexity tasks. They are also not able to perform reliably without human direct control or intervention [17]. While teleoperation interfaces become a natural and practical solution to this problem, teleoperation interfaces may also become the threshold for human-robot teaming. Unlike surgical robots that are specialized for structured operations, nursing robots are designed for assisting a wide range of tasks (see the representative tasks in Figure 1.1) that require the coordination of the control of the robot arm, hand, and base, as well as the active sensors and telepresence. Prior research has demonstrated that the performance of human-robot teaming via teleoperation is limited by the usability of control interfaces rather than the robot's physical capabilities. The high workload and learning effort associated with robot teleoperation interfaces may prevent nursing workers from using the nursing robots on a daily basis and impose barriers to the nursing profession. The recent advances in human-robot interfaces provide a wide range of approaches for the teleoperation and assistance of complex motion coordination. However, it is still unclear which interface and assistance design could be most suitable for the remote control of tele-nursing robots. This dissertation aims to fill the gap in remote manipulation by proposing new design philosophies

for effective and effortless control methods.


Figure 1.1: The prototypes of advanced tele-nursing robots (i.e., TRINA 1.0 [1], TRINA 2.0 [2], and Moxi robot [3]) and representative nursing tasks.

***Limitation of Related Work —*** The state-of-the-art interfaces and teleoperation assistance in terms of freeform control via motion mapping (e.g., motion capture systems [18], portable motion capture devices [19, 20], and exoskeletons [21]) with action support [22] and shared autonomy [23] result in an interface design that is dexterous, precise, and reliable in remote manipulation. However, this type of interface has not been evaluated *in the context of nursing robots with the end-users*. Although efficient and intuitive, teleoperation via motion mapping may cause more *physical fatigue* than conventional desktop robot controllers which has not been thoroughly studied. From the perspective of active telepresence camera control, previous research has developed interfaces for intuitive camera control, such as head/gaze tracking [24], and robot autonomy for dynamic viewpoint selection [25]. However, these designs are mostly hand-engineered and *lack a deep understanding of human behavior and preference for perception-action coupling*. Furthermore, they are designed for single-camera systems and cannot handle active telepresence via multiple cam-

eras. Moreover, related work has proposed various approaches to assist human teleoperators with remote perception and motion control. These approaches include enhancing the interface's sensory capabilities [26], using alternative sensory feedback [27], and delegating difficult tasks to reliable autonomous systems [28]. However, existing research mostly compares similar approaches to validate their effectiveness, and *no study has compared different types of approaches to inform how to choose or integrate them when multiple options are available*. Lastly, existing research has analyzed the causes and effects of failures in human-robot interactions and proposed methods to mitigate negative impacts (see [29]). However, *it is still unclear how the frequency and types of errors affect human performance, workload, perception, and preference for the level of autonomy in robotic assistance*.

***Design Objectives —*** To enable effective tele-nursing, it is crucial that the robot used in the process provides manipulation capabilities in addition to navigation and telemedicine. In general, a tele-nursing robot serves a variety of functions, including communication, mobility, measurement, general manipulation, and tool use. However, manipulation presents a significant bottleneck as non-technology experts are required to operate highly complex robots in stressful and time-sensitive situations for extended periods. Additionally, the level of strength and dexterity required for manipulation tasks can vary significantly, from basic cleaning to precise measurements. The dissertation addresses these concerns by focusing on three key design objectives for human-robot interfaces. The first objective is *efficiency*, which involves completing tasks quickly and effectively, especially in time-sensitive situations. The second objective is *intuitiveness*, which aims to reduce the burden of learning and increase ease of use. The third objective is *ergonomics*, which focuses on designing the interface to minimize physical and cognitive workload, allowing for continuous operation over extended periods. We are also investigating operator augmentation approaches that improve direct control with sensory assistance and action support. We hypothesize that two key factors can impact the natural and casual use of assistance. The first factor is *human preference*, as people often

have unique preferences for how they want to receive assistance. By taking human preferences into account, the robot can be personalized to meet the needs of each user, increasing the chances of successful user adaptability. The second factor is *assistance reliability*, which pertains to the dependability and consistency of the assistive technology. This includes the accuracy and precision of the technology in performing its intended function. If the assistance technology is unreliable, users may be less likely to use it due to a loss of trust and acceptance.

## 1.1 Research Objectives and Contributions

With the objective of enhancing the design of human-robot interfaces for remote manipulation, I investigate the following research questions:

***R1*** *What would be the most appropriate interface design for nursing robots, tasks, and workers?*

*Problem Statement* — Tele-nursing robots hold great promise for providing solutions for quarantine and remote patient care. While more complex robotic systems are being developed for human-robot teaming in future healthcare workplaces, it remains unclear which teleoperation interfaces are most suitable for teleoperating with a wide range of nursing tasks. Given that end-users, such as nursing workers, are typically not familiar with robot operation, it is crucial that the teleoperation interface be efficient, intuitive, and ergonomic.

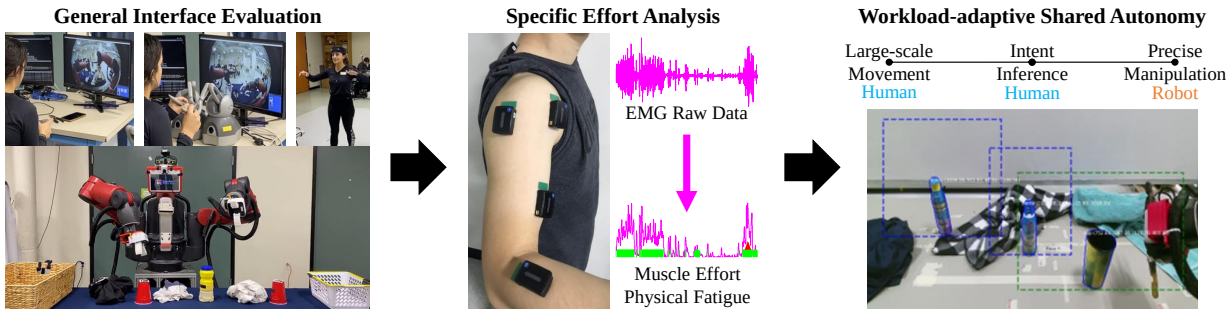



Figure 1.2: A proposed framework for the evaluation and evolution of the human-robot interface.

*Proposed Methods* — We propose a framework to evaluate the control interfaces to iteratively develop an intuitive, efficient, and ergonomic teleoperation interface (see Figure 1.2). The framework is a hierarchical procedure that incorporates general to specific assessment and its role in design evolution. To demonstrate the efficacy of the proposed framework, we conducted three user studies for the evaluation and evolution of the tele-nursing robot interfaces. Specifically, we compared three representative interface designs frequently used for the direct teleoperation of low-autonomy robots (i.e., the robot autonomy is limited to basic collision avoidance and inverse kinematics), including gamepad, stylus/joystick, and whole-body motion mapping via a motion capture system. The evaluations were performed over a set of "pseudo" nursing tasks that require the motion coordination skills frequently performed in a wide range of nursing tasks, including: arm-hand coordination (object grasping), bimanual coordination (for handling large, deformable and heavy objects), loco-manipulation (for navigating in a cluttered workspace), and camera selection and control to support these tele-actions.

*Key Findings and Novel Contributions* — Our general evaluation found that: whole-body motion mapping interfaces have the best task performance and learning efforts among freeform teleoperation interfaces for nursing robots. However, it may cause non-negligible physical fatigue and prevent users from teleoperating with the robots for a long time. Our integrated participant interview and survey also identified and ranked the fatigue-causing factors, including maintaining steady postures for wrist camera controls and adjusting arm posture for stable object grasping. Our robot autonomy design and specialized evaluation focused on reducing the physical workload of the interface and improving the ergonomics of the interface. For the specialized evaluation, we proposed a novel Electromyography (EMG)-based muscle effort index, to provide a more detailed, objective, and accurate physical workload assessment. The outcomes of the specialized evaluation validated the efficacy of our robot autonomy design, as well as our proposed framework for interface evaluation and evolution. We believe the proposed framework, including most of the

evaluation metrics, is applicable to human-robot teaming interfaces beyond teleoperation.

***R2*** *How to design remote manipulation assistance based on human natural perception-action coordination?*

*Problem Statement* — Among the many aspects of motion control, the coordination between perception and action is most critical to tele-nursing task performance. Knowledge about perception-action coupling has been leveraged in human-robot interaction to a limited extent and has already yielded effective models and approaches for predicting human intent [30], optimizing camera motions and viewpoints [31], interactive perception [32], and sensory augmentation of human-robot interfaces for motor skill training and rehabilitation [33]. While remote robots limit human perception and motion capabilities, they also provide opportunities for the human motor system to explore. Novel perception-action coupling skills do not exist in the repertoire of human motor control, yet are critical for robot teleoperation.

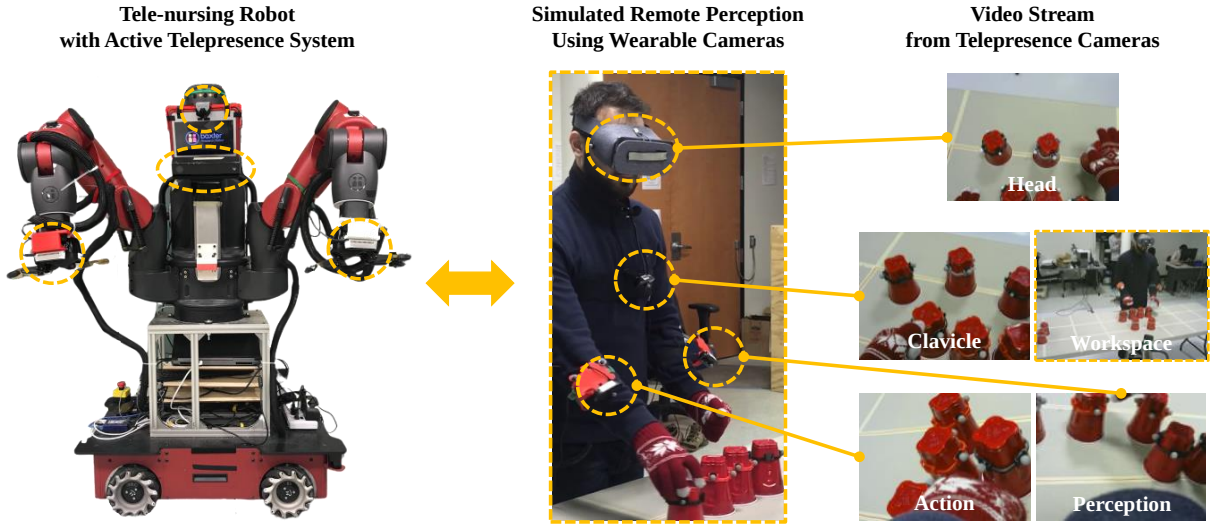



Figure 1.3: The novel experimental paradigm to study the perception-action coordination in the usage of active telepresence cameras.

*Novel Experimental Paradigm* — The research efforts aim to bridge this gap, by proposing a novel experimental paradigm (see Figure 1.3) that can simulate human natural behavior and preference

in the usage of active telepresence. We further conduct comprehensive user studies with this experimental paradigm to 1) discover the perception-action coupling of a coupled human-robot system, and 2) reveal its implications to the design of robot teleoperation interface and assistive autonomy.

*Key Findings* — The findings from our human experiment further imply the suitable design for intuitive camera control and autonomous camera selection. Moreover, people actively resort to every possible haptic feedback, to compensate for the lost depth information of the visual feedback. Inspired by this, from a high-level perspective, we propose a philosophy for tele-robotic interface that we may restore the lost haptic perception by adding vibrotactile feedback, we may replace haptic with augmented reality visual display, or we may delegate the task components that heavily rely upon haptic feedback to reliable robot autonomy.

***R3*** *How to achieve augmentation adaptation to distinct feedback mechanisms, types of assistance, and autonomy levels?*

*Problem Statement* — The use of augmentation techniques can enhance human-robot collaboration in manipulating tasks. However, achieving effective augmentation adaptation to different feedback mechanisms, types of assistance, and autonomy levels remains a significant challenge. The existing interface designs [34] and feedback mechanisms [35, 36] may not be suitable for all tasks, resulting in a decreased level of control and performance. Therefore, there is a need to investigate how to achieve effective augmentation adaptation to distinct feedback mechanisms, types of assistance, and autonomy levels to improve human-robot collaboration and increase user preference.

*Proposed Methods* — Shown in Figure 1.4, we first implement sensory feedback in terms of haptic and augmented reality (AR) visual cues to represent four types of information critical to the precision and performance of a tele-manipulation task, namely: (1) target location; (2) constraint alert; (3) grasping affordance; and (4) grasp confirmation. We further propose systematic approaches for perception and action assistance. For perception augmentation, we redesign the layout of AR visual

cues to convey the visual information difficult to perceive in a 2D camera for the pick-and-place task. We also provide a picture-in-picture (PIP) complementary camera viewpoint from a significantly different perspective, in which missing visual information such as loss of depth perception. The perception augmentation in form of the PIP can be presented always or dynamically given the robot and task states. And for action augmentation, in addition to the automated precise manipulation provided by shared autonomy control, humans can now use hand pose tracking and a trackpad on the handheld controller to control the robot's motions for both freeform and constrained motion. Another approach is the scaling of the operator-to-robot interface mapping is dynamically adjusted to support both large-scale movements and fine adjustments. As AR visual cues for remote manipulation with varying autonomy levels are also a crucial concern. We propose a systematic set of AR visual cues to assist in remote robot manipulation, ranging from direct to supervisory control. These AR cues provide guidance for the human operator to control the robot's motion towards a target, navigate around obstacles in a 3D workspace, and indicate when the robot's autonomy is activated and its planned motion or action.

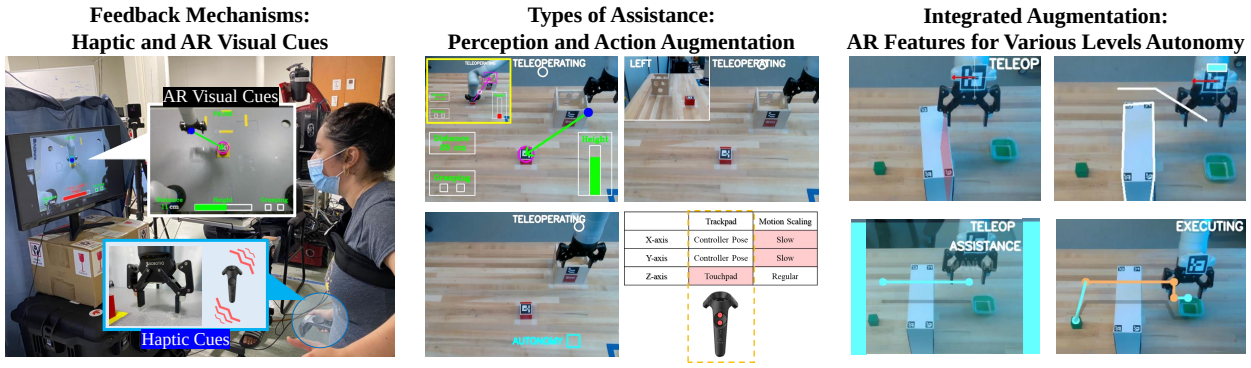



Figure 1.4: The proposed feedback mechanisms, different types of assistance, and integration of augmentation.

*Key Findings and Novel Contributions* — Based on the user study, we found that: (1) transparency, performance, and workload were improved with the haptic and AR visual cues over the baseline without sensory feedback. Also, the preference for sensory integration tended to avoid feedback

overlap when involving the different types of secondary tasks, (2) the effectiveness of and preference for the perception and action augmentation depend on the task performance objective, the user's need for assistance, and the types of users, and (3) participants' preferred AR visual cues changed as the level of autonomy increased from direct to supervisory control. Initially, cues guided human motion for robot control, but as autonomy increased, participants preferred global guidance. AR cues duplicating the robot's autonomy were less effective and not preferred. Experienced users' recommendations influenced participants' cue preferences regardless of initial selections based on video instructions. The novel contributions are threefold: (1) a generalizable design of haptic and AR visual cues crucial for precise and efficient remote manipulation, (2) an integrated comparison of various perception and action augmentations to facilitate optimal human-robot collaboration in freeform tele-manipulation, and (3) a novel method for objective estimation of physical and cognitive workload based on human motion and eye-tracking devices.

***R4*** *What are the implications of reliability on the usage of autonomy and how can they impact human preferences?*

*Problem Statement* — Robot autonomy can provide assistance for perception and action, such as object recognition, human intent inference, and motion planning and control, in various levels and types. These assistive capabilities are expected to improve collaboration between humans and robots in the remote control. However, the reliability of robot autonomy may vary due to uncertainty in perception and action, as well as the complexity of the task. It is currently unclear how to adjust the level and type of robot assistance to accommodate these variations in reliability.

*Proposed Methods* — We explore the impact of robot autonomy reliability on human performance, workload, and preference for robot assistance in remote manipulation. Our study proposes two human-robot collaboration (see Figure 1.5) paradigms: shared control, where humans control gross manipulation and robot autonomy controls precise actions, and supervisory control, where robot autonomy controls both gross and precise actions, but humans detect and correct manipulation

errors. The errors may happen to the autonomous actions triggered by the operator, at the action level (e.g., picking up the wrong object) or the motion level (e.g., missing to grasp an object).

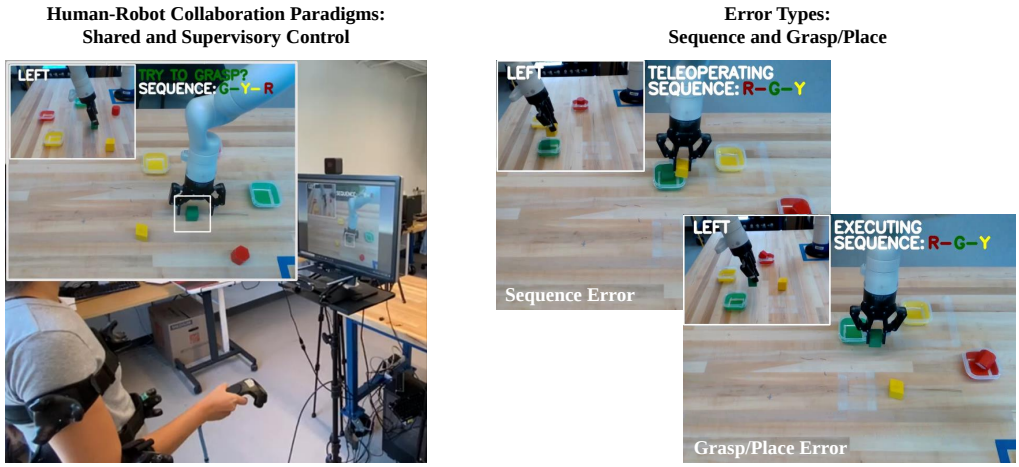



Figure 1.5: Shared and supervisory paradigms for assisted remote robot manipulation and introduced error types.

*Key Findings and Novel Contributions* — We conduct two user studies: one to compare the two HRC paradigms and characterize their effectiveness when the assistive autonomy is reliable; and the other to investigate how the type and frequency of the errors affect the tasks and human operators in the two HRC paradigms when assistive autonomy is not reliable. Our results show that: (1) the interface with a higher level of reliable autonomy yields significantly better performance, lower workload, and higher user preference but lower engagement, and (2) the frequency and type of the error have significant impacts on the task performance and human workload but only partially affects the operator's preference and usage of autonomy.

## 1.2 Outline of the Dissertation

**Chapter 1 —** This chapter introduces the specific application that the dissertation focuses on and outlines the key research objectives that guide the study.

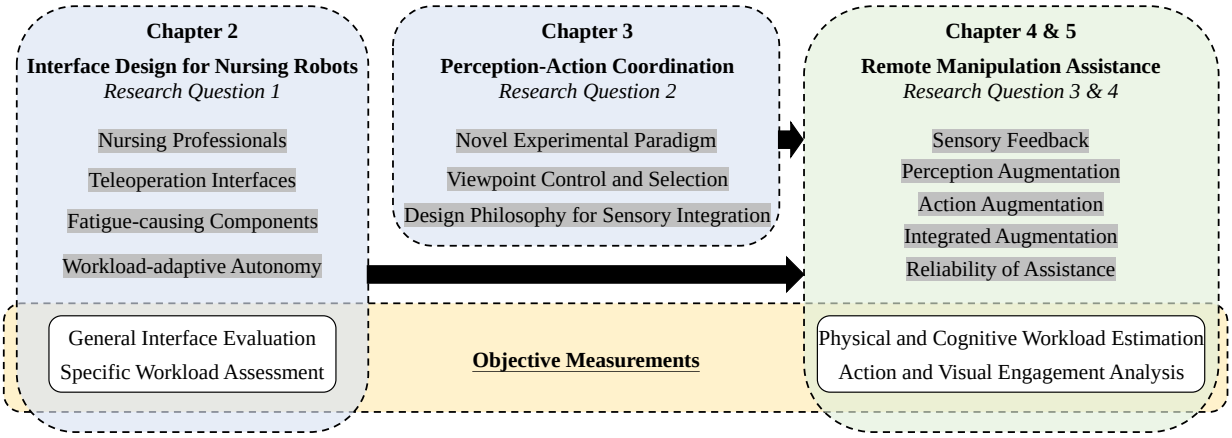



Figure 1.6: Dissertation overview.

**Chapter 2 —** This chapter presents the hierarchical framework that evaluates human-robot interfaces in a general-to-specific manner, addressing limitations rather than augmenting autonomous functions. We apply this to the design evolution of tele-nursing robots, focusing on freeform teleoperation skills and the needs of nurses as the primary users. Findings in this chapter were published in this journal article and these robotics refereed full conference papers:

[J1] **T. C. Lin**, A. U. Krishnan, and Z. Li, "Intuitive, Efficient and Ergonomic Tele-Nursing Robot Interfaces: Design Evaluation and Evolution", *ACM Transactions on Human-Robot Interaction (THRI), 2022*.

[C1] **T. C. Lin**, A. U. Krishnan, and Z. Li, "Physical Fatigue Analysis of Assistive Robot Teleoperation via Whole-body Motion Mapping", *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019*.

[C2] **T. C. Lin**, A. U. Krishnan, and Z. Li, "Shared Autonomous Interface for Reducing Physical Effort in Robot Teleoperation via Human Motion Mapping", *International Conference on Robotics and Automation (ICRA), 2020*.

**Chapter 3 —** This chapter aims to deepen the understanding of human adaptation by proposing

a novel experimental paradigm that simulates human natural behavior and preferences in active telepresence usage. We conducted comprehensive user studies with this paradigm to discover the perception-action coupling of a coupled human-robot system and its implications for designing robot teleoperation interfaces and assistive autonomy. Findings in this chapter were published in this journal article and robotics refereed full conference paper:

[J2] **T. C. Lin**, A. U. Krishnan, and Z. Li, "Perception-Motion Coupling in Active Telepresence: Human Behavior and Teleoperation Interface Design", *ACM Transactions on Human-Robot Interaction (THRI), 2022*.

[C3] **T. C. Lin**, A. U. Krishnan, and Z. Li, "How People Use Active Telepresence Cameras in Tele-manipulation", *International Conference on Robotics and Automation (ICRA), 2021*.

**Chapter 4 —** This chapter explores user preferences of remote manipulation assistance include: 1) multi-sensory integration using haptic and augmented reality visual cues, 2) perception augmentation with additional viewpoints for different perspectives on visual feedback, 3) action augmentation with assistive autonomy and constrained/scaled motion, and 4) a combination of perception and action augmentation with varying levels of robot autonomy and suitable AR features. Findings in this chapter were published in or submitted to this journal article and these robotics refereed full conference papers:

[J3] **T. C. Lin**, A. U. Krishnan, and Z. Li, "Perception and Action Augmentation for Teleoperation Assistance in Freeform Tele-manipulation", *Submitted to ACM Transactions on Human-Robot Interaction (THRI), 2022*.

[C4] **T. C. Lin**, A. U. Krishnan, and Z. Li, "Comparison of Haptic and Augmented Reality Visual Cues for Assisting Tele-manipulation", *International Conference on Robotics and Automation (ICRA), 2022*.

[C5] A. U. Krishnan, **T. C. Lin**, and Z. Li, ”Improve the Control Precision of a free-form Telemanipulation Interface”, *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2022*.

[C6] A. U. Krishnan, **T. C. Lin**, and Z. Li, ”Human Preferred Augmented Reality Visual Cues for Remote Robot Manipulation Assistance: from Direct to Supervisory Control”, *Submitted to IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2023*.

**Chapter 5 —** This chapter looks at how unreliable autonomy affects shared and supervisory control of unstructured remote manipulation tasks, and how it impacts operator performance, workload, and preference. Findings in this chapter were submitted to this journal article:

[J4] **T. C. Lin**, A. U. Krishnan, and Z. Li, ”The Impacts of Unreliable Autonomy in Human-Robot Collaboration on Shared and Supervisory Control for Remote Manipulation”, *Submitted to IEEE Robotics and Automation Letters (RA-L), 2023*.

**Chapter 6 —** Lastly, this dissertation concludes by summarizing key takeaways, highlighting main contributions, and identifying potential directions for future research.

# Chapter 2

# Desirable Control Interface for Tele-nursing Robot

## 2.1 Motivation

Prior research has demonstrated that the performance of human-robot teaming via teleoperation is limited by the usability of teleoperation interfaces rather than the robot's physical capabilities. The high workload and learning effort associated with robot teleoperation interfaces also prevent nursing workers from using the nursing robots on a daily basis, and impose barriers to the nursing profession. The recent advances in human-robot interfaces provide a wide range of interfaces for the teleoperation of complex motion coordination. However, it is still unclear which interface design could be the most suitable for nursing robots, nursing tasks, and workers.

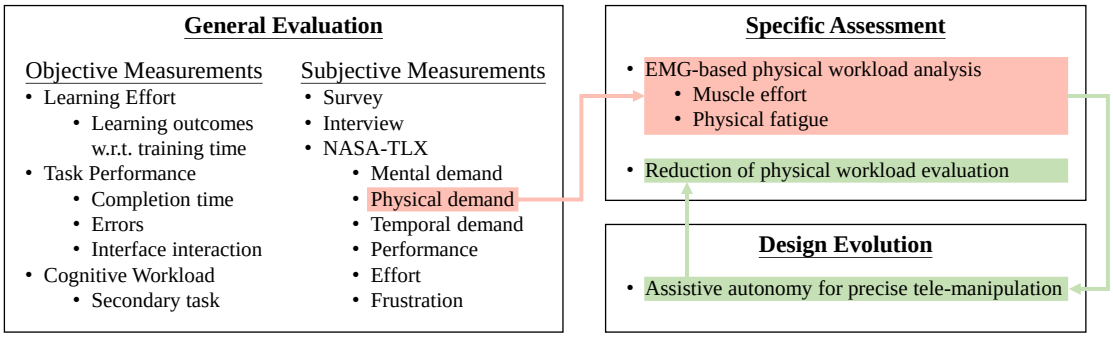



Figure 2.1: A proposed hierarchical framework for the evaluation and evolution of the human-robot interface.

We hypothesize that the desirable teleoperation interfaces for nursing robots should be efficient (high task performance), ergonomic (low cognitive and physical workload), and intuitive (low learning efforts). In this chapter, we presented the proposed hierarchical framework for the evaluation and evolution of a human-robot interface. The concept and the procedures are presented in Figure 2.1. We first conduct a general user study evaluation, to characterize the interfaces based on their performance, workload, and learning effort. With the results from this general evaluation, we further conducted an integrated interview and survey with participants to identify the causing factors and the extent of their influence. Robot autonomy is further designed to address the major interface limitations discovered in the general evaluation user study. A specialized evaluation was used to assess the efficacy of the designed robot autonomy.

## 2.2 Literature Review

**Teleoperation Interface and Assistance —** The capabilities of the nursing robots are not fundamentally limited by the hardware, but by the usability of interfaces. Thus far, research efforts on tele-medicine interfaces primarily focus on tele-surgical robots [37]. Since most of the surgical robots are focused on specific procedures (e.g., [38]), the interface design as well as the evaluation methods and metrics tend to be platform- and task-specific. The state-of-the-art interfaces and teleoperation assistance methods for complex robot platforms (e.g., mobile manipulators and humanoid robots) could be considered for tele-nursing robots of higher motion capabilities [39]. For the online control of motion coordination, these interfaces either map the *human motions* to the robots (using customized cockpits [40], commercial virtual reality and gaming controllers [41, 42, 43, 44], soft/hard exoskeletons [21, 45] and data gloves [46, 47], marker-less or marker-based motion capturing device [18, 19, 20, 45, 48, 49, 50, 51, 52, 53, 54, 55]), or map *human motor commands* using myoelectric devices [43, 56, 57] and brain-computer interfaces [58, 59, 60] (see the representative

interfaces in Table 2.1). Particularly, motion mapping interfaces such as motion capture systems (e.g., Vicon [18]), portable motion capture devices (e.g., Microsoft Kinect [19, 50, 52, 61], Xsens MVN [20, 51, 53, 62]) and exoskeletons [21, 45] are most natural to control humanoid robots to perform hand-arm coordination, bimanual manipulation, locomotion, and whole-body coordination. Within the spectrum of robot autonomy level, ranging from fully manual control to fully autonomous control (see the review in [63]), *action support* and *shared control* are often used to assist the freeform teleoperation using motion mapping interfaces. *Action support* like tremor filtering [64], obstacle avoidance [65] and precise orientation assistance [22, 66], usually assists the execution of a selected action. Shared control is mostly used to assist the operator in actions towards a goal or generating motion along certain trajectories [23, 28, 67, 68, 69]. Overall, the freeform control via motion mapping augmented by robot autonomy results in an interface design that is dexterous, precise, and reliable in motion control, which is desirable for a tele-nursing robot. However, it is hard to specify the design choices of the interface and robot autonomy without appropriate evaluation.

Table 2.1: Representative interfaces for online control of humanoid robot motion coordination.

| Input interface | Controlled motion |
|---|---|
| **Human motions** | |
| Customized cockpits | Whole-body [40] |
| VR controller | Multiple hand configuration [43] |
| | Whole-body [70] |
| Exosuit | Balancing [21] |
| | Manipulation and positioning [71] |
| RGB-D camera | Whole-body [19] |
| | Bi-manual manipulation [72] |
| Marker-based | Whole-body [18] |
| IMU-based | Whole-body [20] |
| **Human motor commands** | |
| Myo | Arm pose [57] |
| BCI | Pick-and-place [73] |
| | Navigation [60] |

**Evaluation Metrics and Methods —** Prior research efforts have provided many generic frameworks for the evaluation of human-robot teaming (HRT) performance and the usability of interfaces [74, 75, 76, 77, 78, 79, 80, 81]. The evaluation of human-robot interfaces is usually coupled with the assessment of task effectiveness (e.g., time- and error-based [82]) and human performance (e.g., workload [82], situational awareness [83]), and therefore share the same set of metrics [84, 85, 86, 87, 88, 89]. Some recent research efforts have also proposed metrics for the evaluation at system level [90, 91], or suggest evaluating dynamic aspects of the human-robot teaming (e.g., interactivity [92], teaming fluency [93], transparency [94], interface learning efforts [95]). For a particular robotic system, modality of the interface, work context and primary user group, the generic framework need to be augmented with the domain- and application-specific metrics [5, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 110]. Generally speaking, the work on human-robot interface evaluation falls into two categories, which may be: (1) a *general evaluation* which uses a framework of metrics to characterize the interface, or (2) a *specific evaluation* to assess the efficacy of some design choices using the metrics that emphasize the interface improvement. The novel contribution of this chapter is to propose an evaluation framework that integrates the general and specific evaluations with the interface and robot autonomy design, to close the loop of interface evaluation and evolution.

The choice of evaluation metrics depends on the available evaluation methods and data collection approaches. Bethel *et al* reviewed the methods of HRI human studies [111], while Abou reviewed the related work for the approaches for HRI performance assessment [80]. In the nursing research community, there are reviews on the usage of nursing robots [112] and on the evaluation of tele-nursing for its satisfaction, self-care practices, and cost savings [113]. The recently developed nursing robot platforms are usually presented with lab experiment evaluations [1, 114, 115, 116, 117, 118]. Although both the field study and lab experiments can collect data for quantitative evaluation, the subjective evaluation primarily replies to interview and survey feedback, while the

objective evaluation usually relies upon the measurements of the robot, task, and human states. The standard practice of evaluating HRT interfaces is to use general-purpose surveys (e.g., NASA-TLX and System Usability Scale [119, 120, 121]), or use a customized questionnaire design only applicable to a specific experiment (e.g., [122]). Participant interview, which has proven to be very useful for connecting human-robot teaming performance to the specific aspects of the interface design characteristics [123, 124], hasn't been well utilized for human-robot interface evaluation. Recently, the evaluation of HRT interfaces has more objective and quantitative metrics adopted for assessing human performance. For instance, the assessment of physical workload can be estimated based upon objective (neuro-) physiological measurements, including Electromyography (EMG) and Electroencephalograms (EEG) [125, 126]. Eye tracking has also been used for objective and accurate assessment of mental workload [127, 128, 129], attention [130, 131] and situational awareness [132, 133, 134]. Thus far, most nursing robots are (mobile) telepresence robots controlled using GUI interfaces, for which the cognitive workload has more influence on the usability of the interface. As the future tele-nursing robots will demand more complex motion control, interfaces that map human motions to robots will become more desirable for the efficient operation of the tasks. The physical workload of using such interfaces on a daily basis will no longer be negligible. Prior research has investigated the physical fatigue associated with conventional tele-robotic interfaces. For instance, the physical fatigue during tele-robotic surgery causes muscle tremors and may result in dangerous situations in critical surgical steps [135]. Besides, fatigue level also negatively affects the Quality of Teleoperation (QoT), which indicates a teleoperator's confidence in commands and decisions [136]. Beyond the teleoperation of medical robots, increased fatigue results in reduced performance during the teleoperation of Unmanned Ground Vehicles (UGV) [137]. The accurate assessment of muscle efforts and physical fatigue, therefore, becomes critical for evaluating interface usability. Recent robot interfaces have incorporated the physical fatigue assessment using EMG sensing [8, 18, 138] and biomechanical human modeling (e.g., OpenSim [139, 140]). Besides, the learning efforts for the interfaces, measured as the difference in task performance and

workload, are also important for the nursing robots to be accepted by nursing workers. To address these needs, our proposed interface evaluation framework will also incorporate novel methods and metrics for the quantitative and objective assessment of the physical workload, and the interface learning efforts of nursing workers.

## 2.3 Telerobotic System

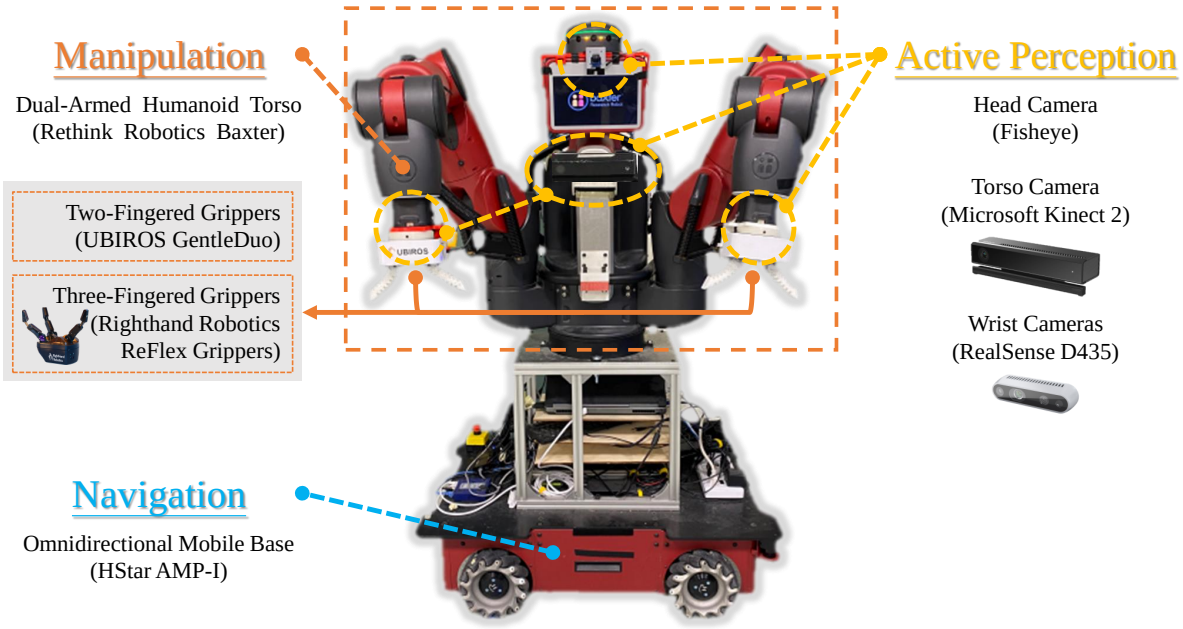



Figure 2.2: Tele-robotic Intelligent Nursing Assistant (TRINA) system.

### 2.3.1 Robot Platform

Shown in Figure 2.2, the Tele-Robotic Intelligent Nursing Assistant (TRINA) consists of a dual-armed humanoid torso (Rethink Robotics Baxter), and an omnidirectional mobile base (HStar AMP-I). For grasping objects, two Righthand Robotics Reflex grippers were used in the experiments evaluating physical fatigue indices and benefits of automating grasping while two two-

fingered soft grippers (UBIROS GentleDuo) were used for evaluating teleoperation interfaces. The visual sensor suite consists of a ELP-USBFHD01M 180° fisheye camera mounted on the robot head, two Intel RealSense D435 cameras on the wrists, and a Microsoft Kinect 2 on the middle of the Torso. The RGB-D cameras on the Kinect are used for object detection for teleoperation assistance.

### 2.3.2 Representative Teleoperation Interfaces

The control input for the teleoperation interfaces is selected based on whether the interface provides point control (i.e., discrete control of robot states) or free control (i.e., simultaneous control of multiple degrees of freedom). Figure 2.3 highlights the selected input devices based on the spectrum of representative teleoperation interfaces, ranging from single-axis control [141] to multi-axis control [142, 143] with three-axis control [62, 144] lying in between.

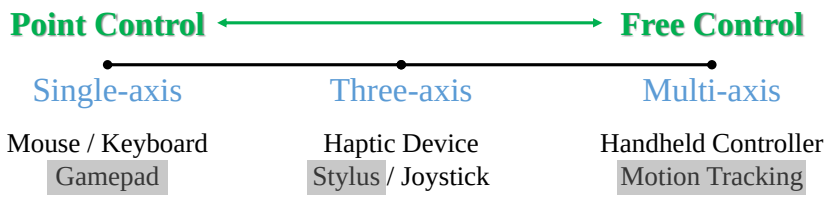



Figure 2.3: Spectrum of representative teleoperation interfaces.

**Gamepad —** The TRINA robot is controlled using a gamepad (Logitech F710) as shown in Figure 2.4. The gamepad control interface consists of 2 modes: arm and base mode. The arm mode controls the motion of the robot arm while the base mode controls the motion of the robot base. There are dedicated buttons to cycle between the two modes, to return the arms to a pre-defined starting position and hold the current gripper pose. The left and right trigger buttons are used to switch between the arm being currently controlled and controlling the gripper, respectively. The

gripper of the currently active arm will be controlled when operated using the trigger. The joysticks of the gamepad are to control the motion of the base and arms depending on the mode the operator is in currently. In the arm mode, the left joystick moves the arms up and down while the right joystick moves the arm forward, backward, and sideways. In the base mode, the left joystick moves the base forward, backward, and sideways while the left gamepad rotates the base clockwise and counterclockwise. Specifically, the end-effector position and base motion are controlled using velocity control based on the location of the joystick on the gamepad:

$$V_x = X_{left}/5 \tag{2.1}$$

$$V_y = Y_{left}/5 \tag{2.2}$$

$$V_z = Y_{right}/5 \tag{2.3}$$

where $V_x$, $V_y$ and $V_z$ are the velocities in the x,y, and z directions, $X_{left}$, $Y_{left}$ and $Y_{right}$ are the x-coordinate and y-coordinate position of the left joystick and y-coordinate position of the right joystick respectively. When controlling the base $V_x$, $V_y$ and $V_z$ correspond to the motion of the base in the x and y directions and rotation of the base respectively.


Figure 2.4: Gamepad controller configuration for teleoperation interface.

As it is not possible to represent the cartesian positions of the gripper through the gamepad, the

grippers are controlled through velocity control. The direction in which the gamepad is held and the extent to which it is moved from the center determines the velocity with which the gripper moves. This principle applies to moving the arm up, down, and sideways. For the gamepad interface, the gripper is designed such that the face of the gripper when open always points downwards. The interface was designed in this manner because this orientation of the gripper enables the robot to pick up even objects that are very small or laundry that can lie flat against the table. However, this interface is limited in its joint control abilities. The addition of this extra mode to the interface might make the interface more complicated.

To illustrate this interface, we explain how the robot is controlled to pick up an object off a table. The user first switches to the base mode to position the robot near the table and the object using the two joysticks. The user then switches to the arm mode and uses the trigger switch to cycle between the left and the right arm. The user uses the two joysticks to position the arm near the object. The right trigger (RT) button is used to close the gripper and the Hold (A) button is used to lock the gripper's position. The user can then use the joysticks to lift the arm off the table.

**Stylus Device —** The TRINA robot is controlled using the stylus device (Geomagic Touch) as shown in Figure 2.5. Due to the limited amount of user inputs that can be extracted from the device, the interface is split into three modes namely the base, arm, and gripper modes. There are two buttons on both styluses that can cycle between the modes. Once a mode is selected, the dedicated engage button can be used to activate the mode of teleoperation.

The motion of the styluses can control the respective arms in arm mode. To make the interface as intuitive as possible, the orientation of the robot arm from the robot's "elbow" to the gripper is mapped to the orientation of the stylus of the device. To control the robot end-effector positions Cartesian position control was used. The position and orientation of the left and right styluses were scaled to the robot frame and the robot limbs were moved to the desired location by solving the Inverse Kinematics of the Baxter robot arms. In the base mode, the left stylus can move the base


Figure 2.5: Stylus device (Geomagic Touch) configuration for teleoperation interface.

linearly while the right stylus is used to rotate the base. This means that the base moves forwards, backward, and sideways depending on the way the left stylus is moved. The right stylus is used to rotate the base clockwise and counterclockwise. The base control uses velocity control similar to the one described in the Gamepad interface where the location of the right and left styluses control the robot base rotation and translation respectively.

To illustrate this interface, we again use the example of grasping an object off a table. The user in the base mode uses the stylus motion to move the base closer to the table and the object. The user then must switch to the arm mode to control the arms using the styluses. Once in a suitable pose, the user can close the gripper in the gripper mode. The gripper can be toggled between fully open and fully closed. Once the object is fully grasped, the user can switch back to arm mode to lift the arm.

**Motion Mapping —** The Vicon Nexus motion capture system is used to develop a motion mapping interface to control the TRINA robot (Figure 2.6). The human motion was captured at 100 Hz by 10 infrared cameras and streamed at 50 Hz for robot control. The subject's physical attributes do not affect the end-effector positions of the robot as only the position and orientation of the wrist and the swivel angle of the teleoperator are mapped to the robot during teleoperation. The swivel

angle is defined as the rotation of the elbow of the operator with respect to the axis connecting the centers of the shoulder and wrist joints [4]. This angle is then used as an indicator of the operator's arm posture.

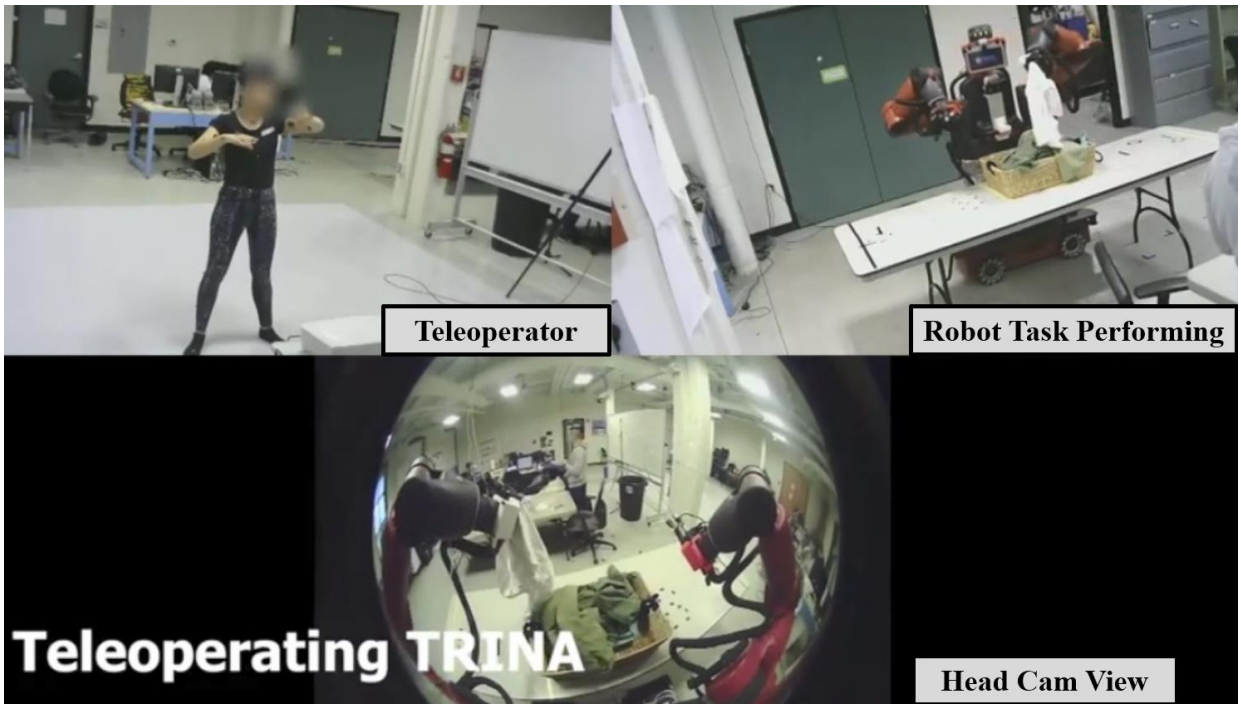



Figure 2.6: Overview of robot teleoperation via motion mapping interface.

The motion mapping interface developed using the VICON motion capture set-up also controlled the robot motion using the Cartesian position similar to the method described in the Stylus device. The location of the operator's hands in the human skeleton captured by the motion capture system is scaled to the robot frame. The Inverse Kinematics for the Baxter robot arms is solved to control the motion of the robot end-effectors. The robot base is controlled using the desired velocity. The direction and magnitude of offset of the operator's feet from the origin defined in the operator's workspace is the direction and magnitude of the velocity with which the base is required to move.

Table 2.2 defines the controls for the motion mapping interface. Robot teleoperation can be

engaged and disengaged by squatting. The operator can control the robot's arms by moving their arms in the desired manner and the robot will replicate these movements. The robot base can be moved by the operator stretching their leg out in the desired direction of motion. For example, the operator can move the robot forward, backward, left, and right by stretching their leg out forward, backward, left, and right respectively. The robot base can also be moved diagonally depending upon how the operator moves their leg. The opening and closing of the robot grippers can be achieved by the operator opening and closing their fingers.

Table 2.2: Motion mapping teleoperation interface. The arm posture is measured by the swivel angle, i.e., the rotation of the elbow position with respect to the axis connecting the shoulder and wrist positions [4].

| **Teleoperation Input** | **Robot Function** |
|---|---|
| **Robot's Upper Body** | |
| Hand position and orientation | End-effector position and orientation |
| Arm posture and orientation | Manipulator arm posture |
| Rotate upper body | Rotate mobile base orientation |
| Hand open/close | Gripper opens/closes |
| Right shank flexion | Activate teleoperation assistance |
| **Robot's Lower Body** | |
| Squat | Engage/Disengage teleoperation |
| Leg steps forward/backward | Mobile base moves front/back |
| Left (right) leg steps left (right) | Mobile base moves left (right) |
| Lift right leg | Switch the camera view |

We shall use the example of picking up an object from a table to explain this interface. The operator moves the robot toward the table and the object by stretching their legs in the necessary direction. Once at a convenient position, the operator can make the robot reach out to the object by reaching out in the corresponding direction in their workspace. The robot replicates this motion and once at the right position, the gripper can be closed when the operator closes their fingers.

**Graphical User Interface —** A Graphical User Interface (GUI) is used to provide the teleoperator with the video stream from the fisheye camera (e.g. in the bottom of Figure 2.6 ) and the two

wrist cameras. The video stream and the control GUI are presented to the user on a monitor at the teleoperation workstation. This video stream provides the teleoperator with a real-time view of the workspace. The GUI also tells which mode the operator is in while using the gamepad and stylus interface or which arm is being controlled when in the gamepad mode. The current robot state is also provided as a 3D model in a simulated environment.

### 2.3.3 Teleoperation Assistance

We proposed a workload-adaptive shared autonomy control to enable humans and robots to complement each other's strength in collaborative control. Human operators can use their hand motions and pose to guide the robot's gross manipulation to approach a target and move across the cluttered workspace so that they can freely determine the task procedures and action sequence. Meanwhile, the robot autonomously performs precise manipulation actions based on human goals inferred by user input. The flowchart in Figure 2.7 describes the design of the autonomous grasping function for teleoperation assistance. The Microsoft Kinect attached to the robot was used for capturing the workspace. Mask-RCNN [145, 146] is used to detect objects and generate bounding boxes of $(2 \times height) \times (3 \times thickness) \times (5 \times width)$ $(cm^3)$ around the center of an object. This enhanced region around the object is where teleoperation assistance is available for the user and is termed the Teleoperation Assistance Zone (TAZ). The TAZ was designed in this manner based on the inputs from an initial pilot study while designing this interface.

The bounding box generated by the Mask-RCNN model on the Kinect RGB-stream around the detected objects was projected to the depth stream from the Kinect. The center of this bounding box was determined to be the center of the object. The depth value of the center of the object thus obtained can be used to find the remaining two coordinates of the object using the pinhole camera equations [147]. The coordinates of the object formed in the Kinect frame are transformed into

the coordinate system of the Baxter. This provides us with the location of the detected object in the three-dimensional workspace. If multiple objects are present in the workspace (e.g. cluttered environment), it will return the coordinate of the closest object as default.

The teleoperator is indicated if the end-effector is in this region by audio and visual notifications. Points on the mid-points of the left and right vertical sides of the original bounding box around the object are identified as target grasping points. Based on where the robot arm presents in the TAZ, a reaching-to-grasp motion is planned for the corresponding nearest target point by solving the inverse kinematics for this location. The assistance does not control the gripper and this action is left to the discretion of the operator who goes will complete the grasping action if he/she believes the gripper is in an appropriate position for grasping.

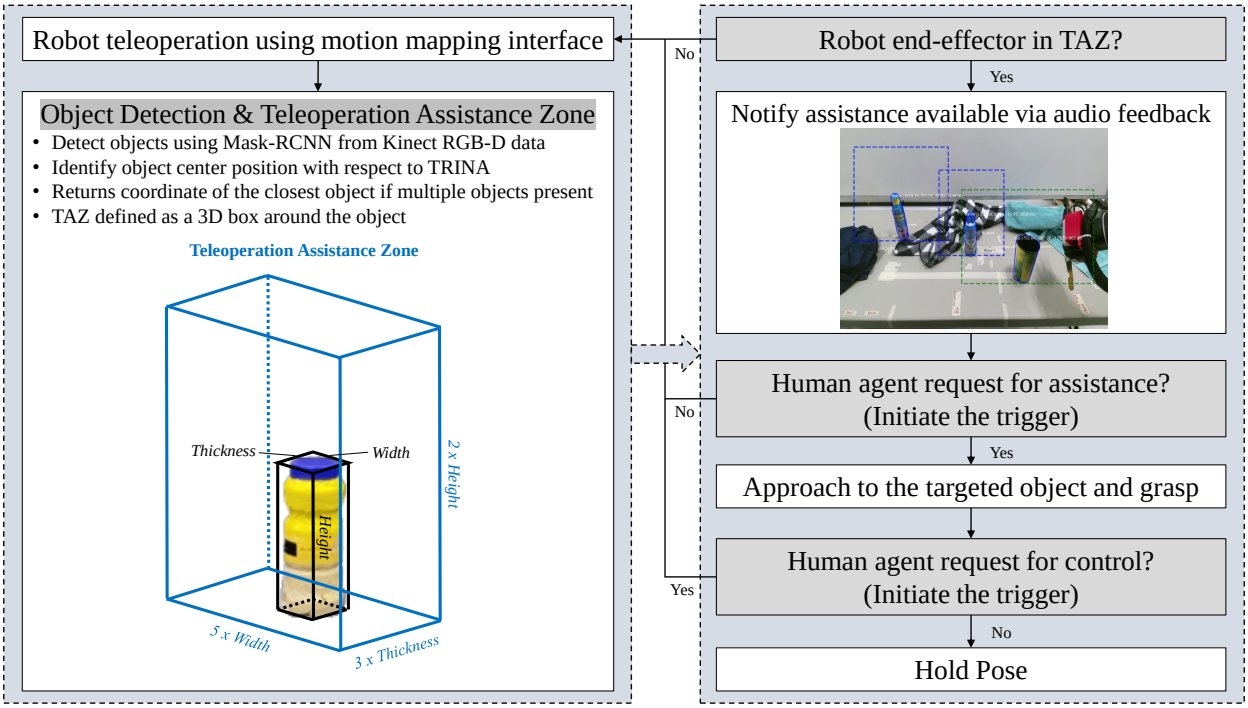



Figure 2.7: Workload-adaptive shared autonomy feature for teleoperation assistance.

## 2.4 User Study I: Comparison of Teleoperation Interfaces

To appraise the usability of the general to specific teleoperation interface evaluation framework, we conducted a series of user studies to validate each component step-by-step. In the first user study, we present the usage of the general pre-defined metrics to evaluate three representative designs of contemporary robot teleoperation interfaces. We tried to find a feasible teleoperation interface that is intuitive and easy to learn for the primary users (nursing workers) who usually do not have an engineering background.

### 2.4.1 Experimental Setup

First, the three teleoperation interfaces are compared to find the best-suited interface for the nursing workers. This experiment requires the subject to control the robot platform using the gamepad, stylus, and motion mapping interfaces (See Figure 2.8). The teleoperator receives a real-time video stream from the fish-eye camera of the robot workspace. The user performs different tasks that involve them manipulating different objects placed on a table in two different experiment stages, the training and performance phase. The task performance time, number and type of errors, and the number of mode switches were recorded. The responses to arithmetic questions asked during the performance phase are also recorded. After all the tasks, the user answers a survey that captures the user's preferences for the different interfaces.

### 2.4.2 Participants and Tasks

Our user study involves (N=8) nursing students (eight female, 19-21 years old) who represent the future users of the teleoperation interface. We also recruited three registered nurses to get their feedback and attitude toward the use of tele-nursing assistive robots on a daily basis. All the par-

ticipants have experience working in healthcare and are familiar with the hospital environment. They also do not have any robotics or engineering expertise and have almost zero gaming experience with the gamepad controller. The experimental protocol was reviewed and approved by the Worcester Polytechnic Institute Institutional Review Board.

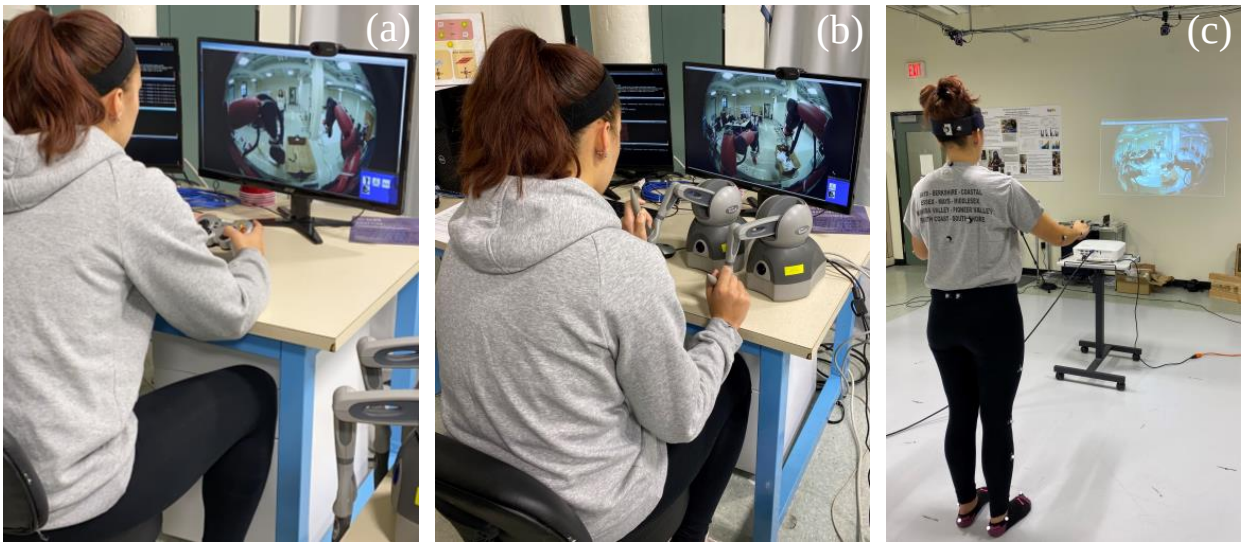



Figure 2.8: Nursing robot teleoperation via: (a) Gamepad (b) Stylus-Style Joysticks and (c) Motion Mapping Interface.

The participants in this user study performed two tasks (see Figure 2.9). The *testing task* (left) is to collect an individual object on the counter workspace and is designed to examine the user's learning effort and the outcome of practice time for each interface. This task is performed both before and after the practice session to evaluate the learning outcome. The *evaluation task* (right) is to clean and organize several objects scattered in the workspace and is designed to evaluate the interface's usability. Our prior study shows that most of the tele-nursing tasks require the users to perform: 1) free control of reaching-to-grasp rigid and deformable objects, 2) free control to move the mobile base to facilitate manipulation, and 3) point control to facilitate engaging/disengaging the interface or mode switching. Based on this finding, we set up a "pseudo-task" which incorporates these necessary teleoperation skills in the context of workspace cleaning and organization by a nurse. Specifically, the user will teleoperate with the nursing robot to collect several rigid and

deformable objects randomly placed on a counter workspace and sort them into two separate bins. This task integrates the free control of precise and gross manipulation in a cluttered environment, locomotion, and point control of robot states, all of which are necessary primitive robot control skills for tele-nursing tasks.

Figure 2.9: The tasks of the user study include collecting a single object (left), and cleaning and organizing a counter workspace (right).

### 2.4.3 Experimental Procedure

The user study consists of a *Training Phase* and a *Performance Phase*. The order of interfaces was randomized for each participant. For each user interface, the experimenter introduced and demonstrated the interface functions. In the *Training Phase*, the subject performs the testing task described in the previous section twice. The two trials of the testing task are separated by a practice session of 15 minutes. The participants can stop the practice session anytime they feel comfortable with the interface and are confident in performing the second trial of the training task. The time consumed for practice and the change in performance between the two trials of the testing task is used in evaluating the learning effort associated with the interface. The participants move on to the *Performance Phase* after completing the training phase. In the *Performance Phase* the

users will perform the evaluation task, i.e., collecting 3 grocery items and 3 pieces of clothing, and sorting them into two baskets. The participants were required to answer two-digit arithmetic problems continuously for the entirety of the task (as in [148]). This secondary task helps identify the mental effort required for decision-making during robot teleoperation and provides an objective measurement of the cognitive workload of the teleoperators.

### 2.4.4 General Evaluation Metrics

In order to investigate the most feasible teleoperation interfaces that can be utilized by non-robotics-related populations (e.g. healthcare workers and nurses), we performed the evaluation of three representative designs of contemporary control interfaces (handheld gamepad, stylus-based devices, and human motion mapping) by a system of objective and subjective metrics to appraise the learning effort, task performance and operation workload.

**Objective Measurements —** For **learning effort** evaluation, we consider the learning outcomes (indicated by the task completion time, numbers of mode switches, and errors before and after practice) and practice time used during the training phase for each teleoperation interface. This measurement helps us to quantify how fast the users can build confidence to use the control interface and how much impact training has on task efficiency and accuracy. On the other hand, the investigator could just simply adopt the method of asking the participants to perform multiple trials of the training task and using only the task completion time to evaluate the interface. This is however not sufficient as an index as it will ignore other critical factors that might be essential in influencing the learning curve.

For **task performance** evaluation, we measure the time required to complete the task, the number of interactions with the interface (mode switches in our case) during teleoperation, as well as the number and type of errors. The type of errors that influence the performance include errors that:

(1) reduce efficiency, (2) decrease accuracy and (3) diminish safety. For example, in the pick-and-place sub-task in our evaluation task, the types of errors we include are: (1) dropping or knocking over objects, (2) inappropriate grasps, and (3) collisions with the table.

For **operation workload** evaluation, we used a secondary task where the participants had to solve simple arithmetic questions [148] while teleoperating the robot. Users were allowed to skip questions if they deemed it too difficult. A mentally demanding task would make the user perceive a problem to be more difficult than they usually would have, and they would answer fewer questions or have more errors or skip more questions. The longer response time indicates that the user requires a higher cognitive workload for robot teleoperation, and therefore has less capacity for the other aspects of the tasks (e.g. professional decision-making, patient interaction, and information inquiry).

**Subjective Measurements —** We utilized NASA-TLX as the workload assessment tool that helps record the user's self-evaluation of Mental, Physical and Temporal Demands, Performance, Effort, and Frustrations [149]. We further calculated the overall NASA-TLX score to identify the subjective workload by weighting each effort demand. The weighting coefficients were generated by choosing from a series of pairs of rating scale factors that were deemed to be important based on the official instructions. In addition to the NASA-TLX evaluation form, the comprehensive custom questionnaire is an integral part that captures the users' feedback and attitude toward the newly implemented methods and interfaces. Unlike the traditional customized questionnaire in human-robot interaction, we performed a post-study interview to identify the causing factors and features that will improve the interface usability.

### 2.4.5 Feasible Interface to Control Tele-nursing Robot

**Learning Effort and Outcome —** Figure 2.10(a) shows the comparison of interfaces in terms of learning effort. The dotted line indicates the mean of the maximum and minimum value from all participants for both learning effort and outcome. For N=8 nursing students, the motion mapping interface (Mocap) has a better learning outcome and lower learning effort, compared to the other interfaces. Each ellipse plots the mean and standard deviation of the testing task completion time with respect to the mean and standard deviation of the user's practice time. The red, green, and blue colors are for the motion mapping interface (Mocap), stylus, and gamepad, respectively. The learning outcomes can be found by comparing the ellipses of the same color. On average, the nursing students spent less time (219$\pm$39 sec) learning the Mocap interface than the gamepad (792$\pm$57 sec) and stylus device (870$\pm$20 sec) interfaces. ANOVA analysis showed that: 1) The learning effort for Mocap was significantly lower than that for the gamepad ($F(2,21)=71.137$, $p<0.001$) and stylus interfaces ($F(2,21)=71.137$, $p<0.001$); 2) The Mocap interface also had the least completion time (61$\pm$6.7 sec) after practice for the testing task, followed by the gamepad (90$\pm$8.2 sec), and then the stylus (228$\pm$57 sec); 3) In the evaluation task, the Mocap interface also has a significantly faster completion time than the gamepad ($F(1,14)=6.979$, $p<0.05$) and stylus interfaces ($F(1,14)=8.296$, $p<0.05$). We also noticed that the completion time for the testing task before the practice was significantly slower than after practice, when using the gamepad ($F(1,14)=5.624$, $p<0.05$) and stylus device ($F(1,14)=5.442$, $p<0.05$). The significant effects of practice for these two interfaces were confirmed by the participants' reports in the post-study interview. The effect of practice is also more significant for the gamepad than for the stylus interface, which may be because the gamepad is a widely used gaming interface for the public. However, the participants also report that it is difficult to remember the many different functions associated with the gamepad buttons. On the other hand, the Mocap interface has a low learning effort and the trivial effect of practice indicates that this interface is the most intuitive one for nursing robot teleoperation.

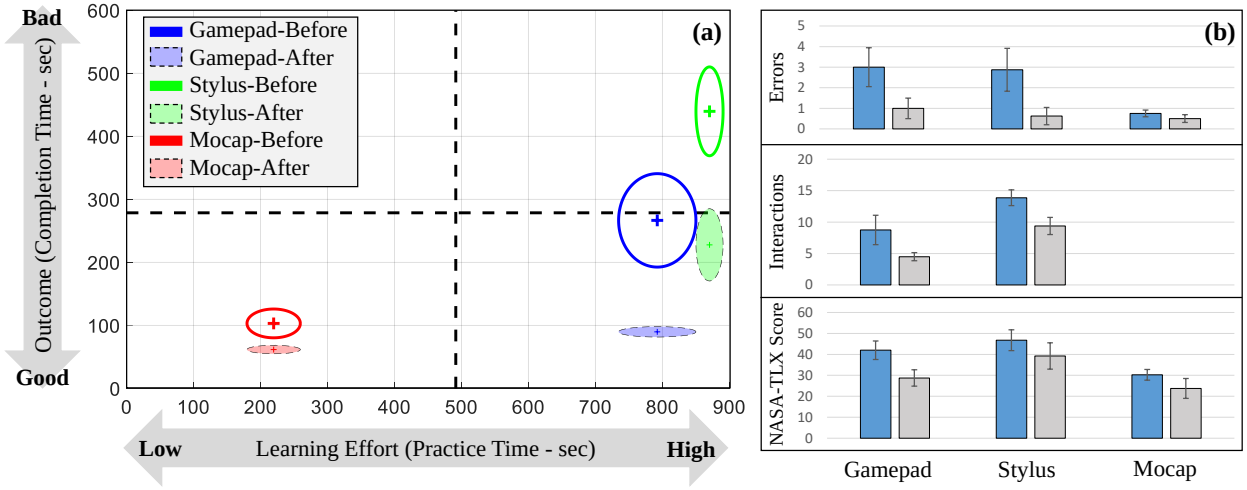



Figure 2.10: (a) Practice time vs Completion time for nursing students. (b) Comparison of the number of errors, interactions, and subjective workload (NASA-TLX) across interfaces for nursing students.

We further use the weighted NASA-TLX scores to measure the subjective workload before and after practice with the interface. The weighting coefficients were selected as follows: mental demand=5, physical demand=4, temporal demand=0, performance=1, effort=3, frustration=2. In Figure 2.10(b), the Interactions field refers to the number of times a mode has to be switched while using the interface. A switch from controlling the base of the robot to the right arm of the robot is considered an interaction when evaluating the gamepad and stylus interfaces. However, for the Mocap interface since the operator can control all aspects of the robot functionality at the same time, this part of the graph is empty as no interactions/mode switching is required. The comparison of all the interfaces for all the nursing students shows no significant difference in the number of errors and interactions and subjective workload during the testing task performed before and after practice. However, the motion mapping interface allowed users to control the robot arm and mobile base simultaneously eliminating the complexity of mode switches that resulted in a lower total subjective workload (23$\pm$4.7) than the gamepad (29$\pm$3.8) and stylus (40$\pm$6.2).

**Task Performance —** Figure 2.11(a) compares the performance in the evaluation tasks among all the interfaces, using the following objective metrics: 1) the completion time of the evaluation task,

and 2) the response time for each math question (i.e., the secondary task). The dotted line indicated the mean of the maximum and minimum values from all participants for both cognitive workload and performance. For nursing students, the task completion time using the Mocap interface was less (404±50 sec) than the gamepad (745±176 sec) and stylus (1367±165 sec). ANOVA analysis shows the significant differences in completion time between Mocap and stylus ($F(2,21)=11.667$, $p<0.001$) and between the gamepad and stylus device ($F(2,21)=11.667$, $p=0.015$). The results also indicate that the subjects took lesser time to solve the arithmetic questions while using the Mocap interface (12.8±1.6) than the gamepad (16±1.5) and stylus device (20.3±2.9).

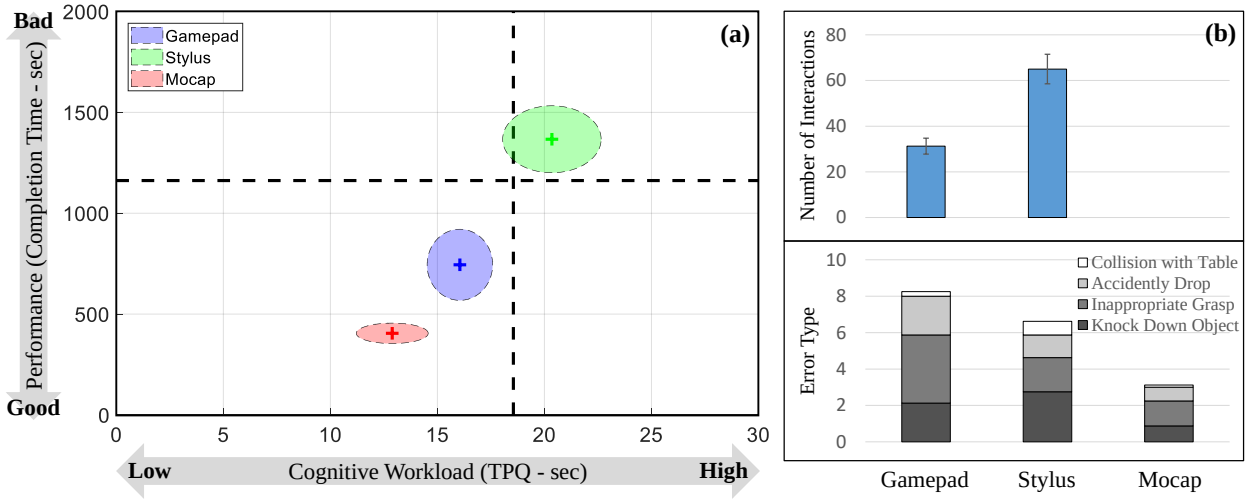
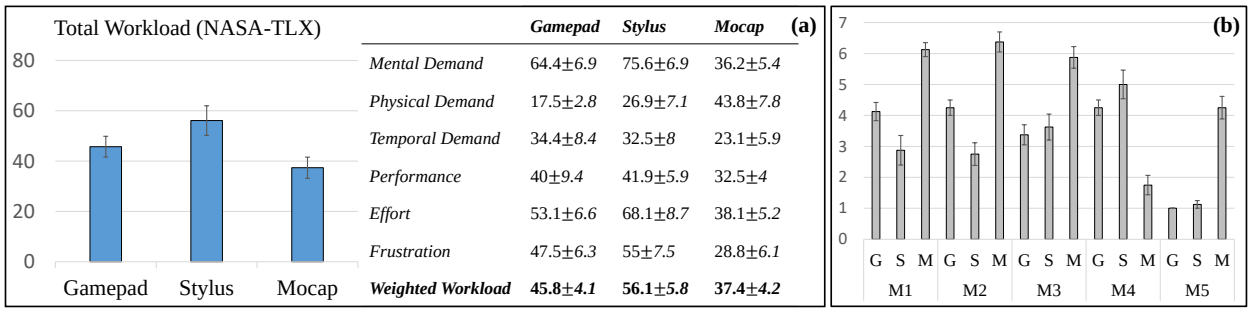



Figure 2.11: (a) Completion time vs Cognitive workload based on the time per question answered for nursing students. (b) The number of interactions and type of errors for nursing students.

Figure 2.11(b) further compares the interfaces by the number of errors and mode switches for the evaluating task. The nursing students have fewer total operation errors using Mocap interfaces (3.1±0.6), compared to using the gamepad (8.6±1.7) and stylus (6.6±1.7). The ANOVA analysis indicated significant differences between the Mocap interface and gamepad ($F(2,21)=3.513$, $p<0.05$). The breaking-down of error types shows that the Mocap and gamepad interfaces tend to cause more inappropriate grasps. Our post-study interviews show that this is due to the lack of depth perception in the visual feedback. Additionally, the stylus interface requires significantly

more mode switches during operation than the gamepad because the user has only one button to cycle between the hand, arm, and base control. The Mocap interface does not need any mode switching as all robot components can be controlled simultaneously via whole-body motion mapping.

**Subjective Operation Workload —** As seen in Figure 2.12(a), the total workload while teleoperating the robot using the Mocap interface was lower (37.4±4.2) than the gamepad (45.8±4.1) and stylus device (56.1±5.8) among nursing students. The ANOVA analysis on the user-reported feedback regarding mental demand shows that the Mocap interface demands significantly lower mental effort than the gamepad ($F(2,21)=9.828$, $p<0.05$) and stylus ($F(2,21)=9.828$, $p<0.001$).




| | *Gamepad* | *Stylus* | *Mocap* |
|---|---|---|---|
| *Mental Demand* | 64.4±6.9 | 75.6±6.9 | 36.2±5.4 |
| *Physical Demand* | 17.5±2.8 | 26.9±7.1 | 43.8±7.8 |
| *Temporal Demand* | 34.4±8.4 | 32.5±8 | 23.1±5.9 |
| *Performance* | 40±9.4 | 41.9±5.9 | 32.5±4 |
| *Effort* | 53.1±6.6 | 68.1±8.7 | 38.1±5.2 |
| *Frustration* | 47.5±6.3 | 55±7.5 | 28.8±6.1 |
| ***Weighted Workload*** | **45.8±4.1** | **56.1±5.8** | **37.4±4.2** |

Figure 2.12: (a) Subjective workload (NASA-TLX) for nursing students. (b) Users' preference rating for the gamepad (G), stylus device (S), and motion mapping interface (M) based on controllability (M1), efficiency (M2), accuracy (M3), mental demand (M4) and physical demand (M5).

**Users' Preference —** The survey feedback from the customized questionnaire is shown in Figure 2.12(b). All nursing students chose the human motion mapping interface as the easiest one to learn and preferred using it as the robot teleoperation interface for future use. Moreover, they reported that the motion mapping interface had better controllability, efficiency, accuracy, and lower mental demand but required greater physical effort.

**Registered Nurses' Performance and Feedback —** Table 2.3 shows the practice time, learning effort (testing task completion time, number of error and mode switches), and subjective workload for three registered nurses across all interfaces. Due to the small population, we are not able to

conclude anything significant. However, the practice time and testing task completion time after practice indicates that: 1) registered nurses require more effort to learn the interface, particularly for the gamepad and stylus indicated by the greater usage of the practice time; 2) for the gamepad interface, practicing the interface does not have as much an impact as it had for the nursing students; 3) the motion mapping interface is still the easiest to learn among all the interfaces for elder nursing workers as identified by the lower practice time, testing task completion time and errors during teleoperation; (4) the overall weighted workload after practice tends to be lower while using motion mapping interface.

Table 2.3: Learning effort and outcome of registered nurses.

| | | **Learning Effort** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *Completion Time (s)* | | *Errors* | | *Mode Switch* | | *Overall NASA-TLX* | |
| *Nurses* | *Practice Time (s)* | *Before Practice* | *After Practice* | *Before Practice* | *After Practice* | *Before Practice* | *After Practice* | *Before Practice* | *After Practice* |
| | | | | | Gamepad | | | | |
| 1 | 900 | 157 | 177 | 2 | 4 | 8 | 14 | 88 | 55 |
| 2 | 900 | 372 | 201 | 8 | 4 | 14 | 10 | 43 | 31 |
| 3 | 900 | 231 | 414 | 2 | 11 | 1 | 12 | 42 | 47 |
| | | | | | Stylus Devices | | | | |
| 1 | 900 | 915 | 213 | 10 | 2 | 33 | 6 | 100 | 80 |
| 2 | 900 | 259 | 212 | 1 | 1 | 9 | 12 | 38 | 34 |
| 3 | 900 | 356 | 153 | 0 | 0 | 12 | 7 | 37 | 22 |
| | | | | | Motion Mapping | | | | |
| 1 | 420 | 54 | 64 | 0 | 0 | 0 | 0 | 12 | 21 |
| 2 | 292 | 76 | 35 | 0 | 0 | 0 | 0 | 44 | 29 |
| 3 | 472 | 63 | 44 | 1 | 1 | 0 | 0 | 28 | 26 |

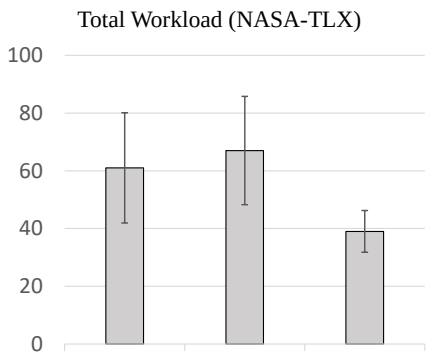
For registered nurses, Figure 2.13(a) shows the performance in the evaluation tasks among all the interfaces using the same metrics used for the nursing students. The motion mapping interface outperforms the gamepad and stylus devices in terms of faster task completion time, fewer errors, and mode switches. The secondary task results also demonstrate that the nurses can solve the arithmetic questions faster while using the motion mapping interface than the gamepad and stylus devices. The total workload while teleoperating the robot using the motion mapping interface was lower than the gamepad and stylus device (Figure 2.13(b)) and the motion mapping interface demands lower mental effort than the gamepad and stylus device interfaces based on the user-reported feedback. All registered nurses chose the motion mapping interface as the most intuitive and easiest to learn. They also preferred using it for nursing assistive robot teleoperation on a daily

basis to help them with routine tasks. Furthermore, they also reported that the motion mapping interface had better controllability, efficiency, accuracy, and lower cognitive workload but was concerned about the heavier physical demand, especially for extended usage.

**(a)**

| Task Performance | | | | |
|---|---|---|---|---|
| *Nurses* | *Completion Time (s)* | *Errors* | *Mode Switch* | *Secondary Task (s)* |
| | | Gamepad | | |
| 1 | 1150 | 7 | 36 | 46 |
| 2 | 579 | 5 | 20 | 17.5 |
| 3 | 1360 | 10 | 38 | 30.9 |
| | | Stylus Devices | | |
| 1 | 1903 | 2 | 55 | 50.1 |
| 2 | 1624 | 7 | 56 | 25.4 |
| 3 | 1474 | 7 | 64 | 32 |
| | | Motion Mapping | | |
| 1 | 400 | 1 | 0 | 14.8 |
| 2 | 325 | 3 | 0 | 15.5 |
| 3 | 377 | 4 | 0 | 17.1 |

**(b)**




| | ***Registered Nurses*** | | |
|---|---|---|---|
| | *Gamepad* | *Stylus* | *Mocap* |
| *Mental Demand* | 80±*13.2* | 75±*18* | 40±*17.3* |
| *Physical Demand* | 40±*30* | 55±*22.9* | 50±*5* |
| *Temporal Demand* | 60±*20* | 50±*25* | 35±*13.2* |
| *Performance* | 55±*17.3* | 70±*15* | 30±*10* |
| *Effort* | 60±*21.7* | 80±*20* | 35±*13.2* |
| *Frustration* | 60±*26.4* | 50±*25* | 25±*8.6* |
| ***Weighted Workload*** | **61±*19*** | **67±*18.7*** | **39±*7.2*** |

Figure 2.13: (a) Task performance of registered nurses. (b) Subjective workload (NASA-TLX) for registered nurses.

To this end, the **whole-body motion mapping interface** resulted in the best task performance and learning outcomes as well as the lowest cognitive workload which indicates it is a feasible interface. However, because of the higher physical demand of the motion mapping interface, the participants reported experiencing a *non-trivial physical workload*, particularly when using the interface for an extended duration.

## 2.5 User Study II: Analysis of Physical Effort

The results from User Study I show that teleoperating the robot via the human motion mapping interface outperforms the handheld gamepad and stylus-based devices in terms of better task performance and lower learning effort as well as cognitive workload. However, non-trivial physical fatigue may prevent such interfaces from being widely used for robot teleoperation, particularly for daily usage made of long work hours. In the second user study, we propose the objective assessment of muscle effort and physical fatigue while teleoperating the robot using surface EMG to investigate the fatigue-causing components.

### 2.5.1 Experimental Setup

The participant was asked to teleoperate with the robot by standing in the center of the motion capture workspace (10 Vero cameras coupled with the Nexus platform from VICON). The real-time visual feedback was projected in front of the participant with the default view being the feed from the fisheye camera. The participant can also cycle through two cameras placed on the wrists of the robot arms to provide depth perception.

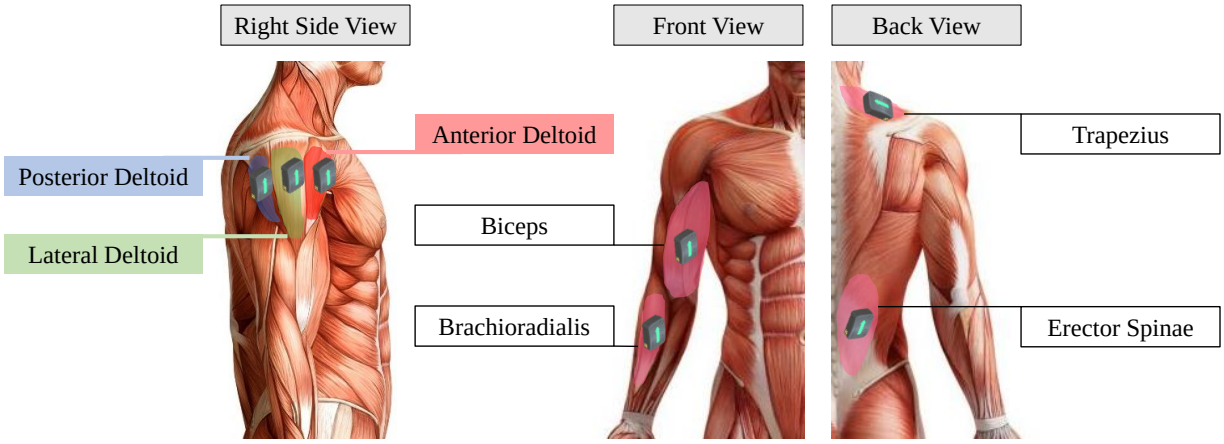



Figure 2.14: Surface EMG placement.

During the experiment, the Vicon motion capture system records human motion at 100 Hz and streams human motion for robot control at 50 Hz. As shown in Figure 2.14, wireless sEMG sensors (Trigno$^{TM}$ from Delsys Inc.) are used to record the EMG signals at 1,000 Hz of 14 individual muscles (Anterior, Lateral and Posterior Deltoids, Biceps, Brachioradialis, Trapezius, and Erector Spinae Muscles of the left and right sides of the body). These 14 muscle groups are most involved in controlling the motion of the upper body.

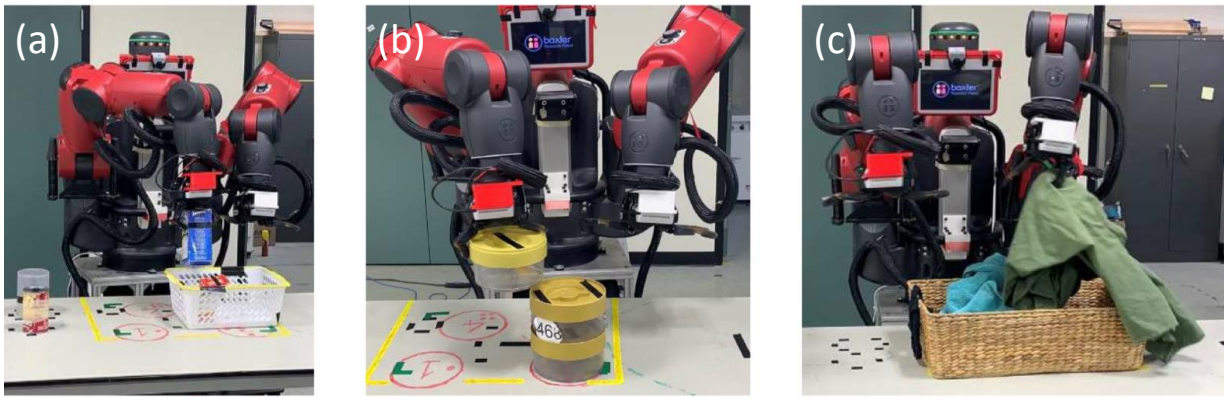



Figure 2.15: Robot teleoperation tasks: (a) collecting, (b) stacking and (c) laundry.

### 2.5.2 Participants and Preparation

Our experiment involved 6 male ($25 \pm 3$ years old) and 2 female participants (28–29 years old). Seven participants had an engineering-related background and one female participant had no experience in engineering or robot control. All the participants had a normal skeletal muscle system in the upper extremities and normal trunk function. The experimental protocol was approved by the Worcester Polytechnic Institute Institutional Review Board.

After the EMG sensor attachment, all the participants were asked to perform the set of subject-specific maximum voluntary contraction tests to record the maximum force generated by each targeted muscle. The MVC test involved a series of single-joint motions to isolate the contraction

from each muscle. The experimenter tries to resist the subject's single joint motion during the MVC test by applying a resisting force. The MVC movements include: (1) shoulder front raise, (2) shoulder lateral raise, (3) shoulder reverse fly, (4) shoulder shrug, (5) biceps curl, (6) wrist extension, and (7) lower back extension. These MVC signals served as the baseline to normalize the EMG signal recorded during the task performance.

### 2.5.3 Tasks and Procedure

As shown in Figure 2.15, the participants are instructed to perform three robot teleoperation tasks in the experiment, namely: (1) Collecting: collect six scattered grocery items on a large table into a container; (2) Stacking: stack food containers in the instructed order; (3) Laundry: collect towels and blankets (3 pieces of laundry) into a laundry basket and take them out in a pre-defined sequence. Each participant performs each task three times. For each iteration of each task, the items were replaced in the same position to ensure that the tasks were executed in largely the same manner. The first repetition of each task was used to analyze the muscle usage during the robot teleoperation.

At the start of the user study, each participant was briefed about the capabilities of the robot platform, the way to use the motion mapping interface, and the objectives of the experiment. Each participant was attached with surface EMG sensors on the muscle groups and reflective markers for tracking human motion. Then, we performed a MVC test for each participant to normalize the EMG signal.

Before the experiment, participants could become familiar with the TRINA system through a training session. Participants first performed a quick practice session that lets them use the functions listed in Table 2.2. The training tasks were similar to the performance tasks but with fewer items. They could practice in this training session until they felt confident and comfortable using the

motion mapping interface independently to teleoperate with the robot to accomplish the tasks.

The order of the tasks was randomized, and participants took a minute's break after finishing each task iteration. They moved on to the next task only if they felt completely rested and they felt no fatigue in any area of their body. After accomplishing the experiment, each participant answered a custom questionnaire. The task completion time for each task was also monitored.

### 2.5.4 Objective Assessment of Physical Workload

By the general evaluation metrics, robot teleoperation via motion mapping has been demonstrated to be an intuitive, efficient, and low learning curve approach for controlling the motion coordination of humanoid robots. However, the trade-off with using human motion mapping as a teleoperation interface is the non-trivial physical fatigue associated with this interface and this information is also captured in the NASA-TLX self-evaluation and post-study questionnaire. Muscle fatigue has been defined as any exercise-induced reduction in the maximum capacity to generate force or power output [150]. Assessment of physical fatigue can be based on the measurements of force, power, and torque. Besides, heart rate can also be used to detect muscle contraction and infer overall physical fatigue level [151]. Among all the approaches, sEMG measurement has proven to be more effective as it measures muscle activity in a non-invasive and real-time environment and provides the ability to monitor the physical fatigue of a particular muscle [152]. Thus, in addition to the subjective measurements, we developed indices that assess the physical workload objectively using wireless surface EMG sensors (Trigno$^{TM}$ from Delsys Inc. at 1,000 Hz sample rate). Our analysis of the EMG signals aims to: (1) investigate individual muscle effort and physical fatigue development during teleoperation using motion mapping; (2) compare the physical fatigue induced by different tasks and thus identify movements that cause fatigue, giving us the directions to facilitate the fatigue-adaptive interface design. Figure 2.16 illustrates our data preparation process for the

muscle effort and physical fatigue estimation.

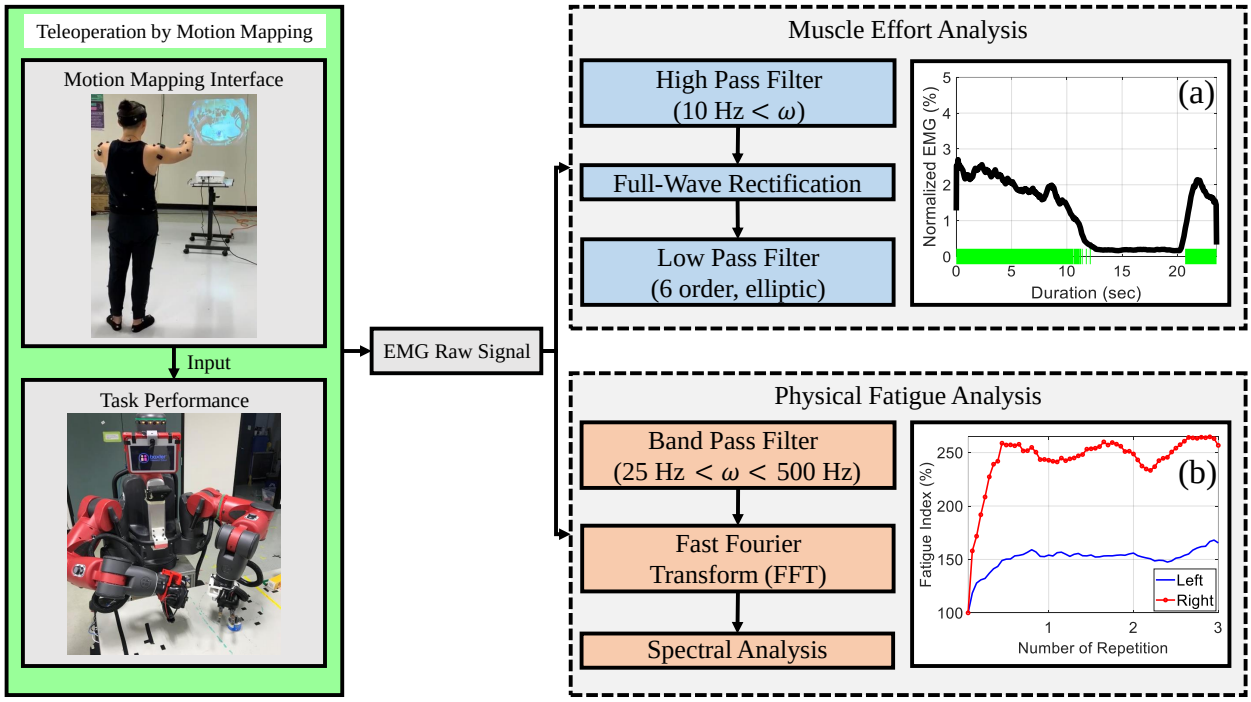



Figure 2.16: Process of the muscle effort and physical fatigue analysis.

**Muscle Effort Analysis —** The recorded EMG signals are within the 40 Hz-700 Hz range in the spectrum domain. The raw EMG data were pre-processed using a high pass filter (cutoff frequency 10 Hz), to remove the soft tissue artifact and offset the frequency baseline. The processed signal further went through a full-wave rectification and then a sixth-order elliptical low pass filter (cutoff frequency 50 Hz), to remove noise and transients and develop a linear envelope of the EMG signal. We use computer-based methods to determine the onset and offset of muscle contraction of the processed EMG signal. The tunable parameters include the threshold value (standard deviation of the baseline signal) and the number of samples (sliding windows in the units of a millisecond) for which the mean must surpass the defined threshold. We choose the combination of three times the standard deviation of the muscle static contraction obtained from the first 200 frames of the EMG signal in the maximum voluntary contraction (MVC) test and 25 milliseconds as the signal sliding window size. This window size has shown similar results to results from the visually derived

data [153]. The MVC test also serves as the tool to normalize the EMG signal with respect to the maximum force generated by each muscle [154]. Figure 2.16(a) shows the individual muscle contraction levels (represented as the black line which is the normalized processed EMG signal from the MVC test) and contraction duration (indicated by the green bars).

**Fatigue Analysis —** We use the band pass filter (25 to 500 Hz) to filter the recorded EMG signals and apply the conventional fast Fourier transformation to convert the signal from time domain to power spectrum domain to calculate the spectral density. Previous studies have shown that the mean and median frequencies of the surface EMG signals decrease as the muscle contraction duration increases, and therefore can be used to measure the fatigue in isometric contraction [155]. To address the inadequate sensitivity of the transitional fatigue indices during dynamic contractions, increased fatigue can be measured by the highly sensitive Dimitrov spectral fatigue indices ($FI_{nsmk}$) [156]. These indices are the features extracted from the spectral moments computed from the EMG power-spectral density (PSD) function. The spectral moment ($M_k$) can be calculated using equation (2.4):

$$M_k = \int_{f_{\min}}^{f_{\max}} f^k PS(f) df \tag{2.4}$$

where $M_k$ indicates the spectral moment, $f$ is the frequency, $f_{max}$ and $f_{min}$ represent the bandwidth of the signal and $PS(f)$ is the EMG power-frequency spectrum as a function of frequency and k is the chosen order. The Dimitrov spectral fatigue indices are represented by the ratio of the spectral moments of order (-1) and order k which is in the range of 2-5:

$$FI_{nsmk} = \frac{M_{-1}}{M_k} = \frac{\int_{f_{\min}}^{f_{\max}} f^{-1} PS(f) df}{\int_{f_{\min}}^{f_{\max}} f^k PS(f) df} \quad , k = 2 \quad to \quad 5 \tag{2.5}$$

The fifth-order $FI_{nsm5}$ data was selected for generating the objective physical fatigue index

since the variation across the repetitions tended to be wider when the order $k$ of the normalizing spectral moment was higher. The relative changes in the fatigue index is calculated against the first repetition within the trial. This means the fatigue index will always start at 100 % and the values increase with increasing fatigue.

$$Objective\ Fatigue\ Index = \frac{FI^{n}_{nsm5}}{FI^{1}_{nsm5}} \times 100\% \quad , n = repetition\ number\ in\ the\ trial \tag{2.6}$$

Figure 2.16(b) shows the example output of the fatigue index for the muscle on the left and right sides of the body (blue and red lines in the graph on the bottom right).

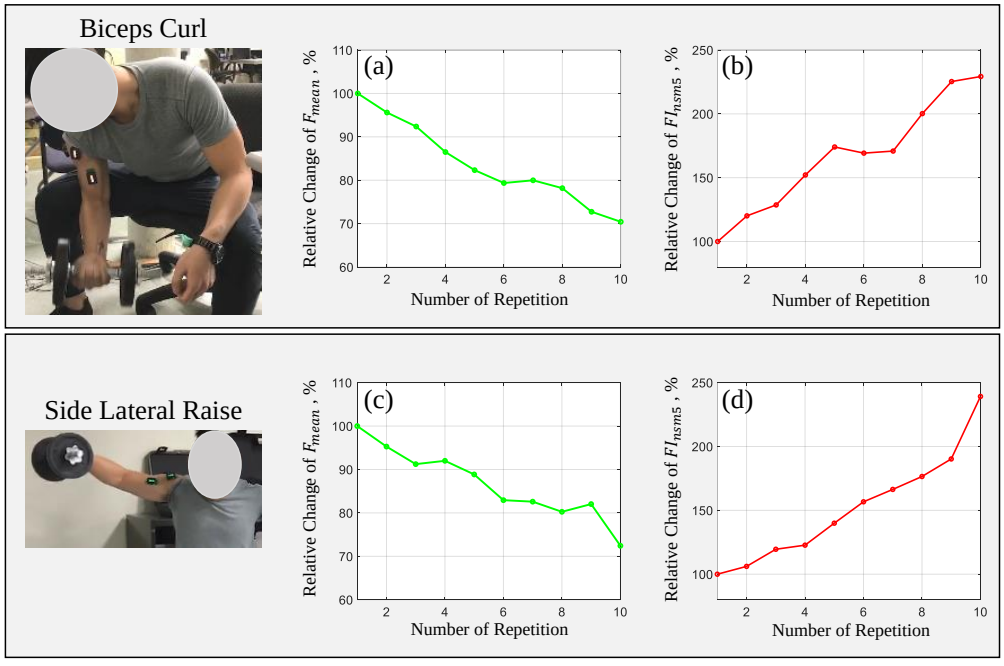
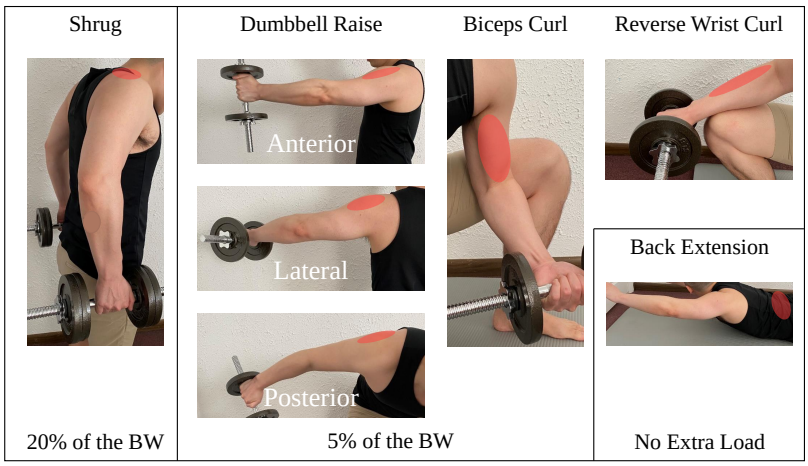



Figure 2.17: Objective physical fatigue indices validation.

**Pilot Testing —** For **verification of our methods with literature,** we evaluated the effectiveness of the objective indices for measuring fatigue using a validation test. Participants ($N = 5$) were instructed to perform single joint movements like Biceps curls and side lateral raises using

their dominant hand to lift a dumbbell (9.5 kg) for 10 repetitions (left in Figure 2.17), which was expected to fatigue their Bicep and lateral Deltoid in the signals detected by the surface EMG sensors attached. As the fatigue increases, the static fatigue index (for isometric contraction using mean frequencies, $F_{mean}$) decreases (Figure 2.17(a) and 2.17(c) shows the results from a representative participant) while the dynamic fatigue index (for the dynamic contraction using Dimitrov spectral indices, $FI_{nsm5}$) increases (Figure 2.17(b) and 2.17(d) shows the results from a representative participant) that those indices share the same trend with the results reported in previous literature [155, 156]. We choose the dynamic fatigue index based on spectral indices as it is more sensitive and was reported to be more suitable for measuring the fatigue caused due to motion.




| | Male | Female |
|---|---|---|
| Deltoid (Anterior) | 150.68 | 123.64 |
| Deltoid (Lateral) | 171.30 | 137.95 |
| Deltoid (Posterior) | 199.85 | 192.33 |
| Biceps | 179.88 | 201.53 |
| Trapezius | 200.65 | 181.94 |
| Lower Back | 268.24 | 248.82 |
| Forearm | 171.04 | 150.79 |

Figure 2.18: A series of isolation exercises (left) and the muscle-specific fatigue threshold (right). Bodyweight is denoted as BW in the figure.

For **identification of fatigue threshold,** one important issue in fatigue assessment is to determine an appropriate fatigue threshold which may vary largely across muscle groups. Prior research chose to use a certain percentage of MVC value as the fatigue threshold, which was expected to be suitable for their tasks. Our robot teleoperation involves tasks that utilize mostly the upper body. Thus, we propose an experimental approach to determine the fatigue threshold for each targeted muscle. The participants (two male and one female) were instructed to perform a series of isolation exercises (pictured to the left in Figure 2.18) using their dominant and non-dominant hands. The

participants had to lift a dumbbell (20 percent of the body weight for the trapezius; 5 percent of the body weight for the anterior, lateral, posterior deltoid, biceps, and forearm; no extra load for the lower back) for 3 sets with 12 repetitions each. They can rest between each repetition for one minute. After the weight-lifting experiment, we asked the participants to point out the specific session and repetition at which they struggled to continue the task. On average, the participants pointed to the $9^{th}$ repetition in the third set as the fatigue threshold. The ratio (in terms of percentage) between the spectral indices of the 9th repetition and the 1st repetition will be used as the threshold for identifying the onset of physical fatigue. The muscle-specific fatigue threshold is defined as the mean of the index value identified for the dominant and non-dominant hand (Figure 2.18 in the right).

### 2.5.5 Fatigue-Causing Components in Tele-manipulation

**Muscle Effort Analysis —** Figure 2.19 highlights how long the different muscle groups were contracted as a percentage of task completion time while performing different trials of the Collecting, Stacking, and Laundry tasks averaged across all the participants. The lateral and anterior deltoid muscles, bicep muscles, trapezius muscles, and forearm muscles showed considerable activity during the execution of the teleoperation tasks.

As seen in Figure 2.20, the subjects were separated into two groups. P1-P3 were users who were familiar with the teleoperation interface while P4-P8 were users who were relative novices to teleoperation. The results show that familiarity with the teleoperation interface reduces task completion time. The effect of familiarity with teleoperation on task performance requires further investigation with a greater number of subjects, but the preliminary results show that it has a positive effect on task completion times.

**Physical Fatigue Analysis —** As shown in Figure 2.21 for novice and expert representative par-

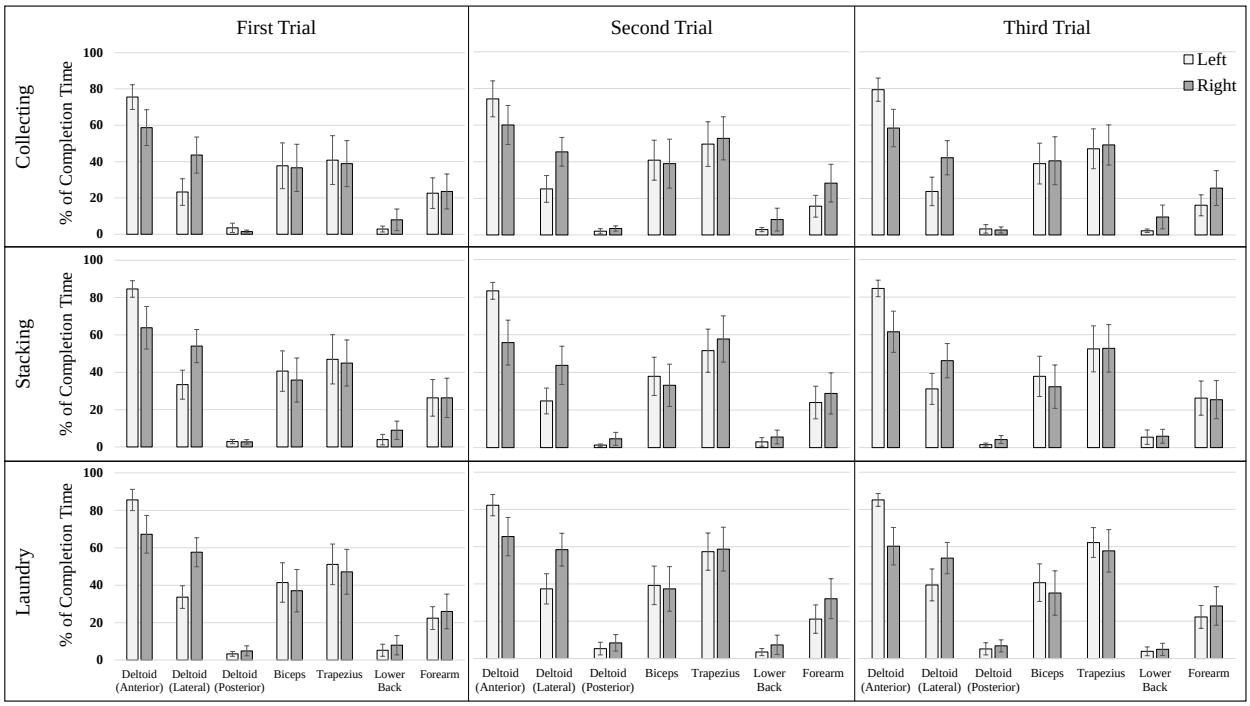



Figure 2.19: The muscle effort across all muscle groups averaged across all the participants for the three trials and three tasks. Muscle effort is identified as the percentage of task completion time that the muscle is contracted.

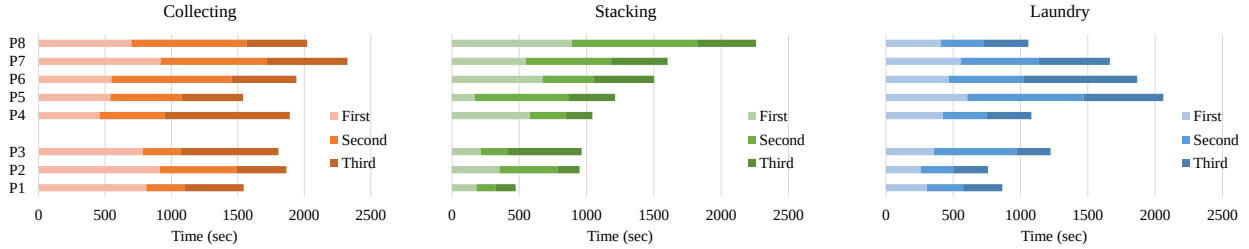



Figure 2.20: The task completion time across user groups and tasks.

ticipants, the area in red in the figure indicates the time when the fatigue index is above the fatigue threshold for the particular muscle identified using the technique described in pilot testing. The Anterior and Lateral Deltoids were found to have been muscle groups susceptible to physical fatigue in addition to the Biceps and Trapezius. The novice and expert groups were created based on the subject's familiarity with teleoperation. This user study shows that physical fatigue developed in the users is lesser if the familiarity with teleoperation is more.

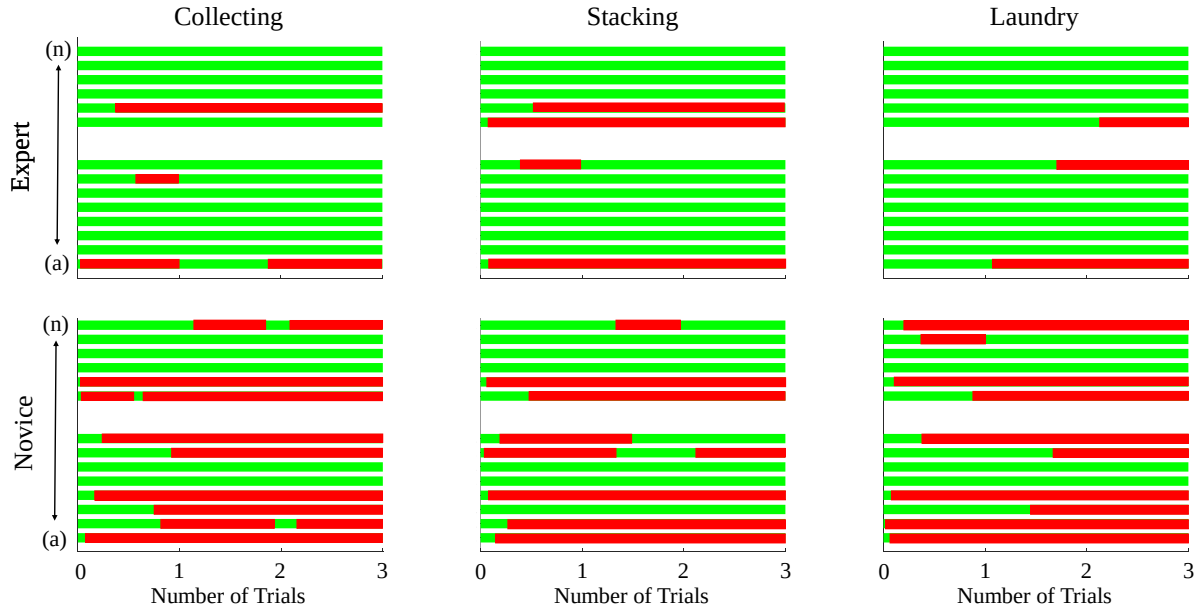



Figure 2.21: Representative participant result of physical fatigue across tasks and muscle groups of the (a) left anterior deltoid, (b) right anterior deltoid, (c) left lateral deltoid, (d) right lateral deltoid, (e) left posterior deltoid, (f) right posterior deltoid, (g) left biceps, (h) right biceps, (i) left trapezius, (j) right trapezius, (k) left lower back, (l) right lower back, (m) left forearm and (n) right forearm.

In Figure 2.22, it is clear that the novice users incur greater fatigue across the Anterior and Lateral Deltoids, Biceps, Trapezius, and Forearms while the expert users show considerably less fatigue across these same muscle groups. The users might become more efficient with their motions with increased familiarity. Similar to identifying how task completion time is related to interface familiarity, further testing is required to identify how physical fatigue of the muscles reduces with interface familiarity.

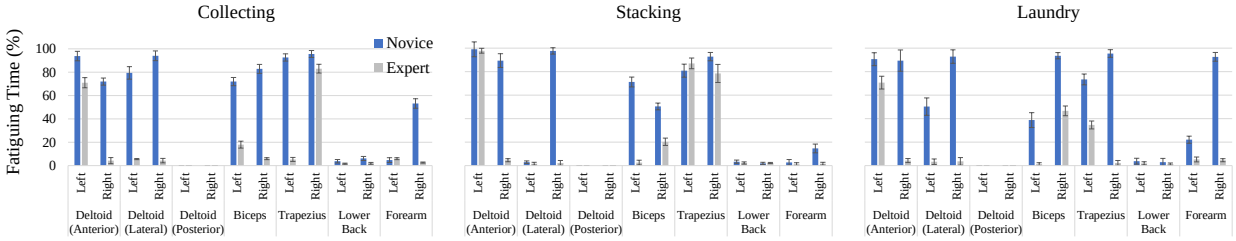



Figure 2.22: The duration of muscle fatigue across all three trials as a percentage of the total task performance time.

**Survey Results —** We surveyed the level of physical demand for each task, and the impact of several possible fatigue-causing factors. The reported physical demand of the three tasks is ordered as: Stacking > Collecting > Laundry, corresponding to the amount of precise manipulation and active perception each task requires. The teleoperation actions that cause the most fatigue (above a 3 rating out of 5 in the survey) include: (a) holding a steady pose of the wrist camera for observation, (b) aligning objects, (c) raising the arm up for a long time during teleoperation, (d) grasping small objects, and (e) adjusting camera view for the best perspective. The less fatigue-causing actions (below 3 in rating out of 5) include: (f) picking objects from the top, (g) grasping large objects, (h) picking up objects from the side, (i) placing objects, (j) carrying grasped objects, and (k) lifting the leg to change camera. The results confirm the fatigue-causing task characteristics and teleoperation actions, implied in the muscle effort and fatigue analysis.

## 2.6 User Study III: Evaluation of Shared Autonomy

The results from User Study II identify the actions that cause the most physical fatigue, namely steady arm postures for wrist camera control and precise manipulation for grasping objects. This physical fatigue as previously mentioned will deter future adoption of teleoperation techniques and needs to be addressed. In this study, an interface was developed that automated the robotic grasping of objects. This will eliminate the need for depth perception through wrist camera control and reduce the teleoperator's effort for manipulation.

### 2.6.1 Experimental Setup

The experimental setup in this user study is similar to the second user study, where the participant teleoperated the robot by standing in the motion capture workspace, with the real-time fisheye

camera feed of the workspace projected in front of them (see the left half of Figure 2.16). In addition to visual feedback, we also provided the camera feed from the Kinect which provides the detected objects and the location of the teleoperation assistance zone for the users. Audio cues are also played when the robot end-effector is within the TAZ.

Wireless sEMG sensors are used to monitor muscle activity during the experiments to evaluate the physical effort. We focused on 10 individual muscles, namely the Anterior and Lateral fibers of the Deltoid, the Biceps, the Brachioradialis (Forearm) and the Trapezius of the left and right sides of the body (see Figure 2.14). Comparing muscle activity with and without teleoperation helps us understand the impact of teleoperation assistance on the teleoperation experience.

### 2.6.2 Participants and Preparations

Our experiment included 6 male (ranging from 22 to 25 years old) and 2 female participants (29 and 31 years old). The Six male participants had engineering-related backgrounds and both female participants had no experience in engineering or robot control. All the participants had a normal skeletal muscle system in the upper extremities and normal trunk function. The experimental protocol was approved by the Worcester Polytechnic Institute Institutional Review Board. Out of the 8 participants in this experiment, 1 male and 1 female participant had also participated in User Study II. Prior to the start of each experiment, all participants also went through a training stage where they were allowed to become familiar with the interface and thus the past experiences of some participants did not inherently make them better than the participants without prior experiences.

After the EMG sensor attachment, all the participants were asked to perform the subject-specific maximum voluntary contraction test to record the maximum force generated by each targeted muscle. The MVC test was conducted similarly to the exercise done in the second user study.

### 2.6.3 Tasks

As shown in Figure 2.23, the participants performed the following tasks: (a) reaching to grasp an individual object, and (b) grasping multiple objects (bottles and cups) in a cluttered workspace. User Study II has indicated that precise manipulation is one of the most fatigue-causing factors in teleoperation. We choose these tasks because precise orientation control in reaching-to-grasp is challenging for the operators during teleoperation and requires careful design of teleoperation interface assistance (e.g., [22]).

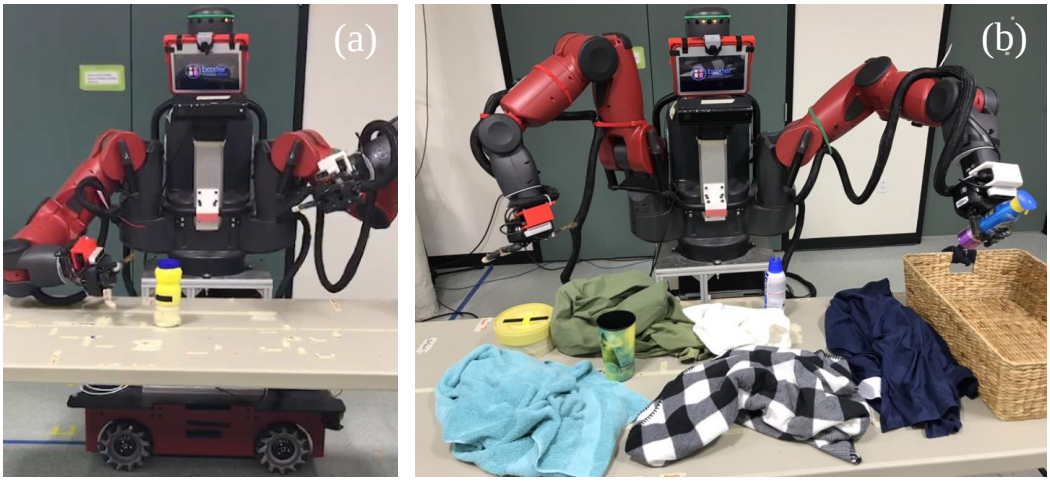



Figure 2.23: Teleoperation tasks: (a) reaching-to-grasp an individual object; (b) collecting multiple objects in a cluttered counter workspace.

### 2.6.4 Experimental Procedure

**Training —** Each participant undergoes a training session to get familiar with the teleoperation interface, the autonomous grasping function, and the robot. The training task is to pick up a bottle on the counter and place it in a basket. The participants can practice in this training session until they feel confident and comfortable using the teleoperation interface and assistive function.

**Section 1 —** In this session, a participant was instructed to reach and grab a bottle placed on the

counter (Figure 2.23(a)). The participants were asked to grab the object for five repetitions each, under the following conditions: (a) using their dominant and non-dominant arms; (b) with and without the teleoperation assistance (Total number of trials = 5 repetitions $\times$ 2 arms $\times$ 2 modes). The order of arms and the availability of assistance was randomized. All the repetitions of the object grasping task had the same initial robot arm configuration, and initial and final location of the object. The participants were required to pick up and place the object in a stable manner. During each trial, we record the time for completing the task, the number of times the object was knocked down, and the EMG signal of the muscle groups for physical workload analysis. The participants also answered survey questions about their teleoperation experience, in the NASA Task Load Index (NASA-TLX) format.

**Section 2 —** In this session, the user has to pick up three cylindrical objects in a cluttered workspace and place it in a basket (see Figure 2.23(b)). This task was to simulate a real-world scenario in which a nursing robot needs to clean and organize a workspace with medical supplies, patient room debris, and laundry (based on the tasks identified in [1]). The participant could choose between picking up the object manually or using teleoperation assistance. If the object was dropped, they are allowed to pick it up unless the object falls off the counter. We counted the number of times that the user uses teleoperation assistance. We also scored the participant's task performance in the following way: (1) +10 points for picking up each object and placing it in the basket; (2) -20 points for knocking an object down or dropping an object when moving it to the basket. The scoring system helps compare quantitatively the performance of the participants who use assistance and don't use assistance by comparing the scores they were able to achieve.

### 2.6.5 Data Analysis and Results

Our analysis of the sEMG data aims to evaluate physical workload (muscle effort and fatigue) during teleoperation using motion mapping with and without the assistance feature. Figure 2.16 illustrates our data analysis process where we have determined individual muscle contraction duration and individual physical fatigue level.

We compared the physical efforts, task completion time, and number of errors in Session 1 (object grasping task), to objectively and quantitatively assess the teleoperators' physical workload reduction when using teleoperation assistance. We further use the results from the NASA-TLX survey and customized questionnaires in Sessions 1 and 2 to assess their perception of workload, preference for teleoperation assistance, and their change in attitude toward teleoperated robot technologies.

**Performance and Efforts —** We analyzed the recorded data to evaluate the teleoperator's efficiency, accuracy, and effort to perform the object grasping task in Experiment Session 1. For *Efficiency* (T) and *Accuracy* (A), we averaged task completion time and the number of errors across all five repetitions in the four discrete conditions (with and without assistance, and for both the dominant and non-dominant hands). The *Effort* (E) is measured by the average contraction duration for all the muscle groups. For each participant, these three indices were then normalized to range between 0 and 1, with respect to the difference between maximum and minimum values across all the conditions. Figure 2.24 compares the performance of Radar Charts across participants. Overall, the teleoperation assistance improves the task *Efficiency* and *Accuracy* for all the participants and for teleoperation using both the non-dominant and dominant arms. The reduction of **Effort** is more prominent and consistent for the non-dominant arm across the teleoperators.

Our ANOVA analysis further reveals the improvement in task Efficiency and Accuracy when using teleoperation assistance for the object grasping task. This can be seen by the recorded task

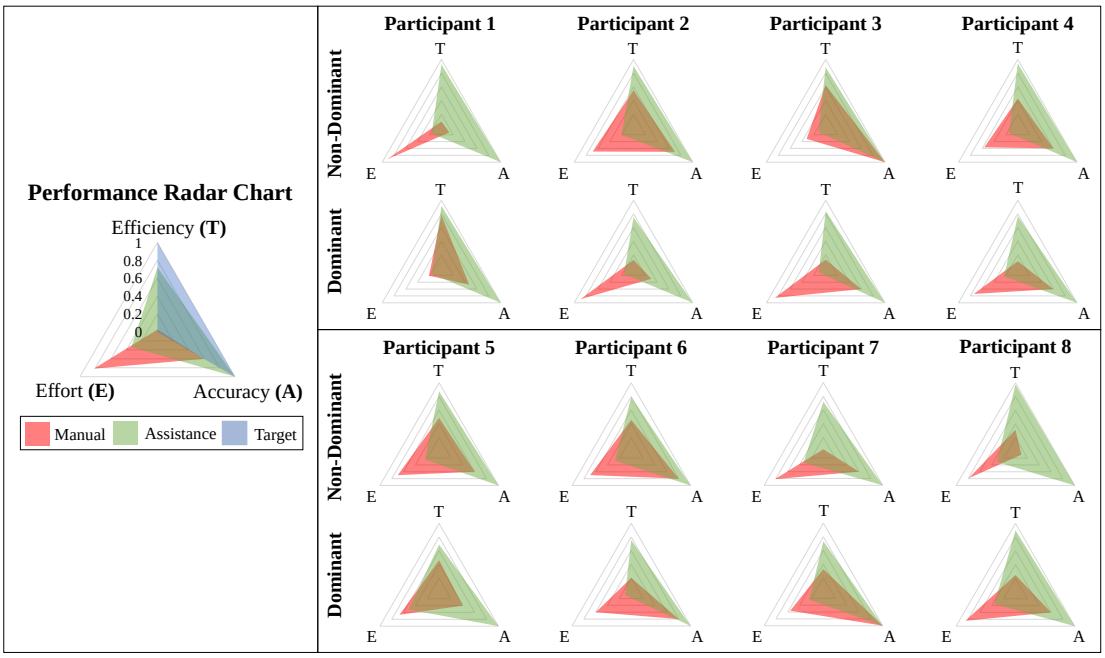



Figure 2.24: Performance evaluation procedure and summary for object grasping across all subjects.

completion times (non-dominant arm: $F(1,12)= 33.87$, $P< 0.01$; dominant arm: $F(1,12)= 52.35$, $P< 0.01$), number of errors (non-dominant arm: $F(1,12)= 6.02$, $P< 0.05$; dominant arm: $F(1,12)= 9.85$, $P< 0.01$) and duration of muscle contraction (non-dominant arm: $F(1,12)= 5.93$, $P< 0.05$; dominant arm: $F(1,12)= 7.93$, $P< 0.05$). Overall, grasping without teleoperation assistance took 13.3 seconds longer for the non-dominant arm and 11.9 seconds longer for the dominant arm on average. This is mostly because the teleoperation assistance reduced the risk of knocking down the object during grasping and the effort for precise manipulation.

We further compared the muscle efforts and physical fatigue between teleoperation with and without assistance across all muscle groups for each participant. The different levels of muscle effort were calculated using the Kullback-Leibler (KL) divergence measurement based on muscle contraction duration and all the results were normalized by the maximum value. As shown in Figure 2.25, most of the muscles had a significant reduction in physical effort (marked as green) with a higher level of relaxation for the deltoids and biceps of the dominant/non-dominant hand. It is noted

that the trapezius muscle however has reduced reduction (marked as white) or increased physical effort as shown by the red marks for 2 participants. Overall, the assistance function performed equally effectively on both arms for all participants.

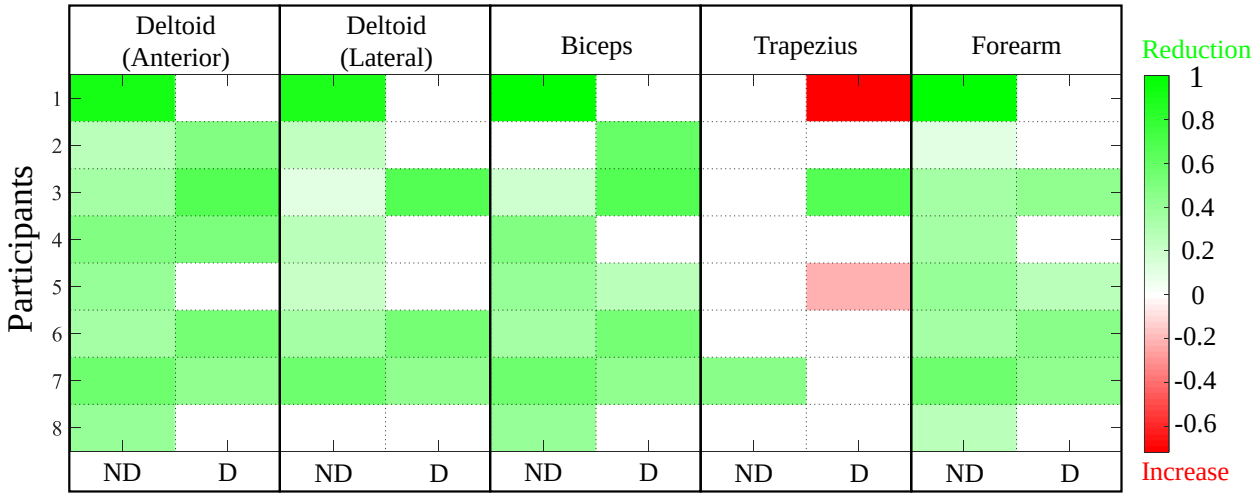



Figure 2.25: Comparison of physical effort across all muscles with dominant (D) and non-dominant (ND) hand.

The level of physical fatigue was computed using the fatigue index for all repetitions. Figure 2.26 presents the physical fatigue developed across all the muscle groups for a representative participant. The magnitude of the fatigue index was lower while using the teleoperation assistance for most of the repetitions for both dominant and non-dominant arms. The duration for each repetition of the task was relatively small and hence the fatigue index did not pass the fatigue threshold. However, these results can be used to predict the potential physical fatigue developed due to extended teleoperation duration and is a part of our planned future work. The area under the curve represents the total accumulated fatigue. As shown in Figure 2.27(a), most of the muscles had significantly less accumulated fatigue (non-dominant arm: anterior and lateral Deltoid, Trapezius and Forearm, $P < 0.01$; dominant arm: anterior and lateral Deltoid and Forearm, $P < 0.01$) while using teleoperation assistance.

We also used the weighted NASA-TLX scores to evaluate the teleoperators' perception of

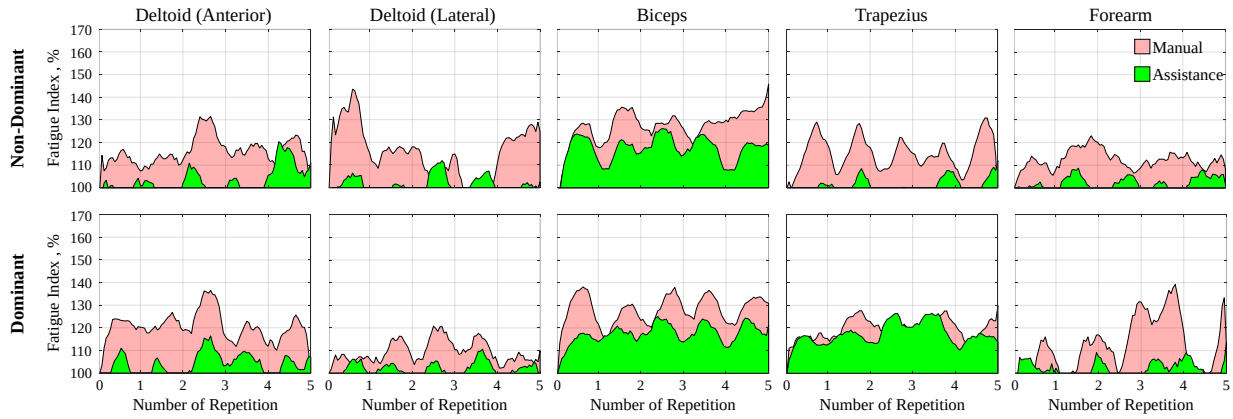



Figure 2.26: Representative participant result of physical fatigue across all muscles with dominant (D) and non-dominant (ND) hand.

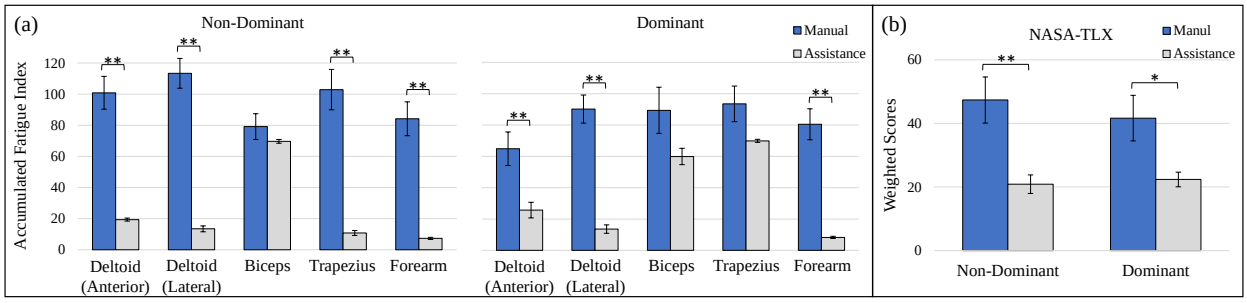



Figure 2.27: (a) Comparison of accumulated fatigue across all muscles in the dominant (D) and non-dominant (ND) hands. (b) The subjective workload from weighted NASA-TLX scores.

task performance and workload. The weighting coefficients were selected as follows: mental demand=4, physical demand=5, temporal demand=0, performance=2, effort=3, frustration=1. Shown in Figure 2.27(b), the teleoperators have answered the survey in support of the usability of the assistance function. Participants reported a significantly lower workload while using the teleoperation assistance for the non-dominant ($P < 0.01$) and dominant ($P < 0.05$) hands. The lower workload rating for the assistance function is understandable as there were no errors during operation and the need to manually execute the precise manipulation to perform grasping is eliminated. Additionally, as the assistance function reduces the duration of muscle contraction the mental fatigue incurred due to teleoperation also reduces. As a result, the operation times are reduced as there are no errors

and user motion is more efficient. The users may have reported reduced physical workload in their surveys because of these advantages.

**Preference for the Teleoperation Assistance —** In Experiment Session 2, participants could choose whether or not to use teleoperation assistance to pick and place objects. As shown in Figure 2.28, we found that (1) more participants prefer to use teleoperation assistance (16 times out of 24), (2) participants who used assistance more (more than two times out of three, P4-P8) had higher scores than the participants who performed the tasks more manually (P1-P3), (3) participants who completed the tasks more manually (55.3±7.4) reported higher subjective workload than using assistance (24.6±6.8).

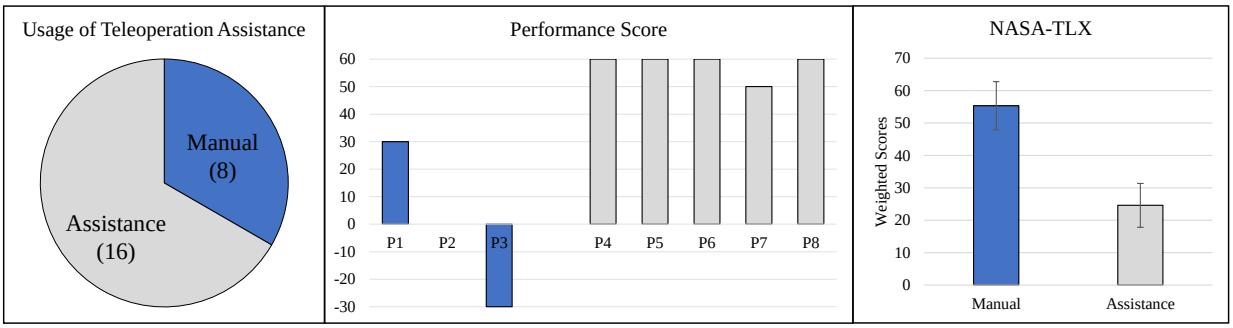



Figure 2.28: Performance of score system for collecting three objects.

After Experiment Session 2, participants rated in hindsight their preference for teleoperation assistance and manual control during robot teleoperation on a 1-7 Likert scale with 1 being the least and 7 being the most in terms of the agreement. As there was a greater preference for the assistive function (6.6±0.6) than purely manual control (3.6±0.7), the users were questioned on what factors made them favor teleoperation assistance more. They point out that teleoperation assistance can (1) increase the success rate; (2) reduce the task completion time; (3) reduce the cognitive workload; and (4) reduce the physical workload. The results highlight the participant's belief that teleoperation assistance improves performance.

## 2.7 Summary and Outlook

In this chapter, we present a complete framework that integrates interface evaluation and evolution in a closed-loop and potentially iterative process. We compare the motion mapping interface with several widely used tele-medical robot interface modalities. To the best of our knowledge, this is the first user study that evaluates nursing robot interfaces with nursing students and practitioners. We also design and evaluate robot autonomy for reducing the physical fatigue in robot teleoperation, as physical fatigue is a non-trivial problem when human motion tracking interfaces become more widely used for teleoperating co-robots in the near future. We further investigate the physical fatigue developed across all the muscle groups in the dominant and non-dominant hand while teleoperating with and without teleoperation assistance (shared autonomy) which can be used to objectively validate the effectiveness of interface design. We present a novel way of integrating the identification of the ideal teleoperation interface for healthcare/nursing duties through user studies and interviews, using EMG signals to identify the sources of physical discomfort and designing teleoperation assistance based on the findings from the EMG signals.

**Desirable Characteristics of Tele-nursing Robot Interface —** Tele-nursing robots should preferably be efficient and accurate while being intuitive with bi-directional communication. These components will increase the nursing workers' preference for using such robots as their proxy to perform repetitive daily tasks so that the risks of disease infection, physical strain, and injuries are reduced. However, the increased functionality also increases the complexity of the tele-nursing robot hardware and software. It is important to implement a suitable control interface for nursing workers who usually have limited experience with robot control or limited engineering expertise. Right now, the current and future population of nursing workers is pre-dominantly made of women and the female-to-male ratio is about 9:1 [157, 158]. The average age of the current nursing population is around 49 years [159]. About 50% of full-time nursing faculty are 50 years or older [160]. Research has

consistently shown that women and elders tend to perform worse in tasks that require spatial skill [161, 162], which is used to estimate robot teleoperation skills [163, 164, 165, 166, 167]. Elders also tend to have less experience with newer technology [168, 169] and are less willing to adopt them [170, 171, 172, 173, 174, 175, 176, 177]. Gender stereotypes are often perpetuated because women may not be included in the design process or test populations [178, 179, 180]. The lack of transparent and intuitive interfaces leads to not only low task performance of nursing tasks, but also creates intimidating cognitive and physical efforts for the users. The negative experience with traditional and contemporary robot interfaces may further reinforce the age and gender biases that discourage current and future nursing workers to envision the future of human-robot teaming in a nursing workplace, and integrating robots into nursing education.

We showed the advantage of using human motion mapping as the tele-nursing robot control interface by comparing it with a handheld gamepad and stylus-based device (User Study I). The lower learning effort among nursing workers new to robot operation and the intuitive freeform motion control could also make the teleoperation experience more immersive. Interestingly, from the post-study interview, the participants stated that they felt the operational effort was *reduced* because they *were able to* simultaneously control manipulation and navigation when using the motion mapping interface. The interviews also mentioned that participants appreciated being able to control the robot arms to perform complex orientations with ease. It is understandable that in a realistic patient-caring scenario, nursing workers often need to take care of multiple tasks simultaneously. Nevertheless, we noticed that the control interface using a gamepad is suitable for more structured tasks since it can precisely and slowly control each motion for each degree of freedom. Interviews of the participants indicated that the ability to use different buttons to control different functions of the robot made teleoperation simpler. They felt this would prevent unintended motions of the arms or the base. The participants also stated that the ability to move the robot in small discrete increments via the joystick input gave them more confidence while performing

delicate operations like picking up objects. This is unlike the motion control observed with the Vicon interface and the stylus-based interface where precise control is harder to achieve. On the other hand, when the workspace of the task is limited to a certain area, the stylus device will be a good fit to control the robot to perform tasks with small movements. Additionally, the buttons on the stylus hardware help integrate discretized base and arms control while also enabling teleoperation through intuitive motion mapping of the stylus. Compared to the gamepad interface the motion mapping of the stylus interface was reported as being more intuitive to use according to some participant surveys. However, for nursing tasks that involve an unstructured and large range of movement (e.g. laundry in a cluttered environment), freeform control (human motion mapping) is preferred.

The work can be enhanced by exploring other wearable/portable interfaces for whole-body motion mapping. Although motion capture systems are accurate for human motion tracking, the cost of hardware and the effort required to set them up makes them less desirable. We also noticed that the teleoperation performance and nursing workers' preference might be affected by lots of factors (e.g. age, gender, gaming experiences, spatial skills, etc.). Investigating the impact of each identified factor to further the development of desirable teleoperation interfaces is another way of improvement. A user study will also be devised to study how immersive each teleoperation interface will be as immersion will play a great role in improving the situational awareness of the operator while teleoperating. Through the three user studies presented in this chapter, we have verified the usability of our interface design and evaluation framework. To quantitatively evaluate the usability of the different interfaces, we will also work on developing a user study where the performance of these teleoperation interfaces for a diverse array of tasks will be analyzed.

**Teleoperation Assistance for Reducing Physical Workload —** Motion mapping as a teleoperation interface proves to be the most intuitive and preferred means of teleoperation. However, the physical fatigue developed from using this interface cannot be ignored and can result in the rejec-

tion of this interface as a means of everyday sustained use. As identified in this chapter (User Study II), physical fatigue is developed primarily in the Anterior Deltoid, Trapezius, and Biceps muscle groups due to teleoperation actions like steady arm postures for camera control and small-object manipulation. The squatting action was performed only when the operator had to pause teleoperation which only occurred at the start and end of the trials. Thus, the EMG signals of the leg muscles were not monitored for analyzing the physical workload. However, since standing for an extended duration might be a possible source of fatigue, monitoring the EMG signals to verify this aspect of teleoperation can be an interesting direction for future research.

We used a shared autonomous control interface to tackle the issue of physical fatigue. Nursing tasks require a lot of decision-making skills and occur in an unstructured environment. As a result, the entire task cannot be automated as current automation techniques do not capture the nuances of operator-controlled teleoperation. In this chapter (User Study III), we have proposed how automating reach-to-grasp reduces physical effort and fatigue in the operator and improves their perception of teleoperation. Aspects of teleoperation like locomotion and gross manipulation are left to the operator while the finer manipulation involved with object grasping is automated and can be triggered on and off base on the operator's needs.

We have demonstrated that augmenting the direct, freeform interfaces for robot control with a little bit of robot autonomy will lead to flexible and reliable robot control for complex tasks. Our proposed robot autonomy effectively reduced the operator's physical workload in the control of reaching-to-grasp motions. The work can be made stronger by developing a variety of robot autonomy to other fine motor skills necessary for quarantine patient care (see the fine manipulation tasks listed in [181]). We will also explore how to design robot autonomy to be fatigue-adaptive, such as triggering the autonomous function based on task context, inferred user intents, and estimated physical fatigue level.

**Evaluation Metrics for Nursing Workers and Tasks —** The general framework (in User Study

I) of the current human-robot teaming evaluation system we used in this chapter evaluates the robot task performance and human workload. The different levels of controllability, efficiency, accuracy, intuitiveness, and effort will affect the users' preference and attitude toward using tele-nursing technologies. Among the nursing workers, the weight of each teleoperation factor may change depending on the usage. For instance, most of the registered nurses from the post-study interview reported that the operational workload is their highest priority since they work in a tense environment where they handle multiple inputs and outputs in a nurse-patient interaction. In this chapter, we demonstrated the effectiveness of using evaluation-in-the-loop in the evolution of tele-operation interface design. We focused on the teleoperation interface with a lower mental workload (human motion mapping) and tried to learn more about its limitations by evaluating physical fatigue through the use of sEMG sensors.

The use of sEMG sensors helps us monitor the physical workload objectively. In this chapter, we showed the potential for sEMG as an offline analysis tool that can evaluate physical workload. It helped us to identify the fatigue-causing factors (User Study II) and helped us generate a novel objective index to evaluate physical fatigue. These results helped us identify the direction in which the shared autonomy must be designed (as seen in User Study III). Our work is similar to [8] in that we prioritize reducing muscle efforts in human-robot collaborative tasks. However, we focus on tasks that involve more complex motor skills and dynamic muscle contractions.

To reinforce the work, we will incorporate more advanced methods for the accurate estimation of physical fatigue. Traditionally, physical fatigue is measured using amplitude-based parameters and time-frequency distributions of non-linear parameters. These metrics are suitable for evaluating isometric fatigue. We will consider the novel approaches which are more suitable for assessing dynamic muscle fatigue. These approaches may utilize dynamic muscle fatigue model, differential equations, mechanomyography [182], inertial measurement unit (IMU) measurements [183], power spectral indices and kinetics and kinematics [184] from motion data.

# Chapter 3

# Perception-Action Coupling in Active Telepresence

## 3.1 Motivation

Contemporary tele-robotic systems (e.g., for nursing assistance [1], surgery [185], manufacturing [186], etc) are usually equipped with multiple active telepresence cameras to provide the teleoperator sufficient perception of the remote environment and the tasks. Deciding how to select and control them to acquire the desirable camera motion and viewpoint could be as difficult as controlling the freeform dexterous tele-manipulation, given that the remote cameras may be located at the robots head, the manipulators for task operation or camera assistance, the mobile base, or standalone in the workspace and can be moved as required (see Figure 3.1 [1] for example). When focusing on the tele-manipulation tasks, teleoperators often neglect effective control of the active telepresence cameras to avoid the additional cognitive workload. Although robot autonomy for camera assistance is necessary, ill-designed camera assistance, which *do not account for the natural preference of human visual perception and visual comfort*, may confuse and frustrate the teleoperators, and reduce their performance and trust in robot autonomy.

The remote control of active telepresence cameras is difficult because the robot teleoperators need to develop novel motor skills to control the unfamiliar viewpoint of the robots, which are different from human eyes and their viewpoint in their displacements, motion capabilities, depth

perception, and field of view (FOV) [84]. Controlling this foreign viewpoint of the robot is counter-intuitive to humans who are used to the location, perception capabilities, and natural viewpoint control motions of human eyes. To assist the teleoperators to utilize the active telepresence cameras better, prior research efforts have developed 1) interfaces (e.g., via head/gaze tracking [24, 187]) for intuitive camera viewpoint and motion control, and 2) robot autonomy for autonomous dynamic viewpoint selection and camera motion control [25]. However, the design of interfaces and autonomy for camera assistance is mostly hand-engineered and based on empirical experience, rather than *the in-depth understanding of human natural behavior and preference of perception-action coupling, which has a strong influence on how humans prefer to coordinate the remote camera control and robot motions and actions*. They are also mostly designed for single camera robotic systems and are not capable of handling active telepresence via multiple cameras.

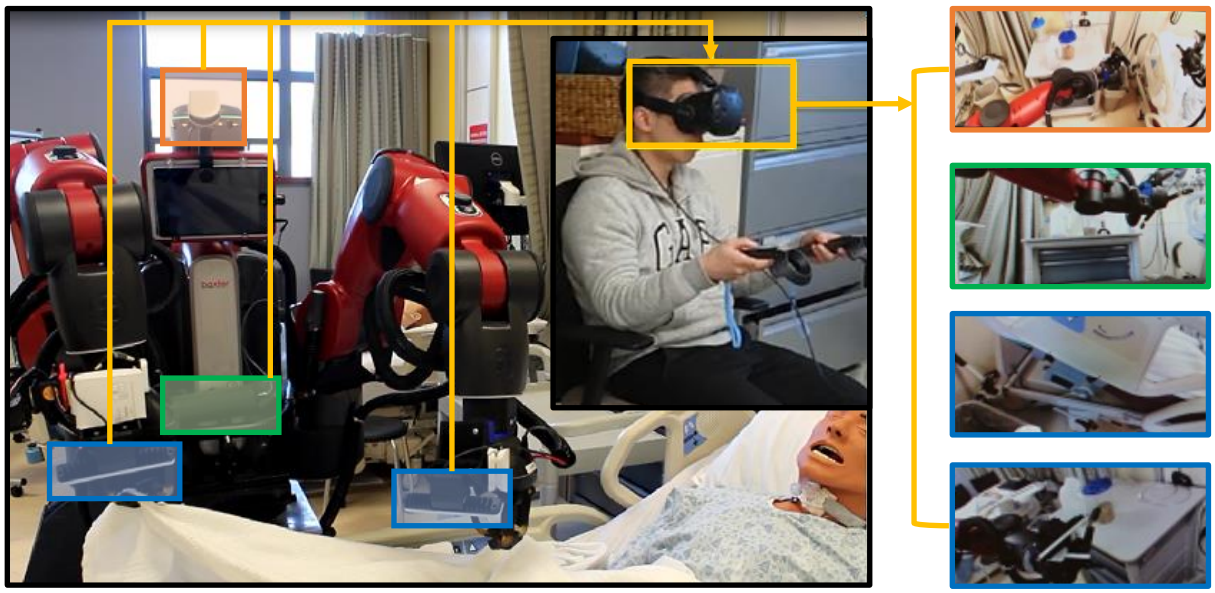


Figure 3.1: Nursing robot teleoperation via a freeform interface with feedback from multiple active telepresence cameras attached to head, torso, and wrists.

In this chapter, a novel experimental paradigm is proposed to study perception and action coupling in terms of vision-motion coupling, haptic-motion coupling, and vision-haptic coupling of sensory integration. Based on the results of the user study, it can be inferred that an effective active

perception camera control design would incorporate shared autonomy for camera selection, as well as an intuitive assisted teleoperation interface.

## 3.2 Literature Review

**Active Telepresence for Tele-robotic Systems —** The usage of multi-camera telepresence has enabled tele-robotic systems to operate in complex environments and perform tasks that require high dexterity and mobility while under the control, guidance, or supervision of remote human users. Many contemporary tele-robot systems integrate multiple cameras to increase the field of view, or to provide the additional viewpoint of robot, tasks, and environments [188]. For example, Nguyen *et al* recently integrated an array of four cameras to provide a wider field of view, such that remote users could assist with wheelchair navigation [189]. Compared to panoramic cameras, the integration of multiple telepresence cameras can provide a sufficiently wide camera view at lower cost and energy consumption. On the other hand, Whitney *et al* proposed to display the 2D video from hand cameras of a humanoid robot along with the point cloud from its head camera. Teleoperators use both the global and local task views to efficiently control the robots to perform dexterous manipulation tasks such as laundry folding [190]. An interactive detail-in-context telepresence interface displays the pan-and-tilt view from a narrow camera (in robot head) inside of a wider pannable view (attached to a pole extended from the robot back), such that the teleoperators can zoom in on details of a selected region [191, 192]. Indeed, a multi-camera telepresence system can integrate the displays from the cameras of different robots. For instance, De León et al proposed a design of multi-camera telepresence to increase the navigation capabilities of multi-robot systems in disaster response [191]. The robot primarily responsible for the mission is provided with the external viewpoints from the cameras of the other robot teammates, in addition to the onboard camera it carries. The feed from multiple cameras on the robot can be provided simultaneously or be relayed

as active camera feedback where the different viewpoints can be individually controlled. As presented by Seo et al [192], the ability to go back and forth between multiple camera views being relayed to them at the same time will let the operator get more information about the workspace at the same time and corroborate information about the workspace by going back and forth between views. However, displaying multiple camera views at the same time can cause information overload overwhelming the operator and thus affecting their ability to perform [193]. Additionally, to fully utilize the potential of simultaneous multi-camera feeds requires the ability to spatially correlate the events between different viewpoints. However, this ability is dependent on the spatial reasoning skills of participants which are highly user-specific and as a result can result in increased cognitive workload for operators with limited spatial reasoning skills [194, 195]. With an active multi-camera telepresence network there are improved remote perception capabilities of tele-robotic systems and improved situational awareness among the robot teleoperator or supervisors. However, tracking, managing, and controlling the feed from multiple cameras also demands additional cognitive and operational efforts. In general, related work in literature addresses this limitation by the design of: 1) *control interfaces* that use head motions and/or gaze for intuitive camera control [24, 196, 197], and 2) *robot autonomy for camera assistance* that autonomously adjust the camera viewpoint to track the object of interest [198, 199, 200], the robot end-effectors or tools [25, 31], or the features critical to task performance [201, 202, 203]. The autonomy for camera assistance can also be used to optimize the camera motions for visual comfort [204, 205, 206], or optimize camera viewpoint for information gain [207, 208], aesthetics [209, 210, 211], viewpoint familiarity [212] or other considerations. Nevertheless, these control interfaces and camera assistance are limited because: 1) they were mostly designed for single-camera systems, and 2) the strategy for camera viewpoint and motion control was proposed and evaluated case-by-case, based on empirical experience and hand-engineered criteria, instead of systematic understanding of human behavior and preference for the selection and control of active telepresence cameras.

**Vision-Motion Coupling —** If a human is subjected to a foreign viewpoint, like if the visual perspective was from their torso or hands, with limited haptic sensation from touch, then the human would have to adapt novel ways to use this foreign vision and haptic sensation to interact with the environment. Fortunately, we are confident that the human motor system is able to re-develop a "new normal" to best utilize the new perception and action capabilities, as seen in motor skill training [213] and rehabilitation [214, 215].

The temporal and spatial coordination of vision and movements, namely the visuomotor coordination, is essential to human motor control. The human behavior and underlying human motor control strategies of the vision-motion coupling [216] have been extensively investigated in various human motor skills. Specifically, many human factor experiments have studied the gaze pattern, visual control, or eye-hand/eye-foot/eye-head coordination in the tasks including active perception (e.g., visual search [217], target selection [218], target tracking [219, 220], scene viewing [221]), manipulation (e.g., reaching [222, 223], reaching-to-grasp [224, 225], grasping [226], interception [227, 228], bimanual coordination [229], object manipulation [230]), and locomotion (e.g., walking [231, 232], navigation [233, 234], driving [235]), tool and interface operation (e.g., laparoscopic surgery [236], video game [237]), and learning of motor skills (e.g., [238]). Such experimental studies reveal that human gaze and visual control in daily activities can be influenced not only by the salient features [239] and surprising stimuli [240] in the task environment, but also by the action and behavior goals [216] (and their associated intrinsic [241, 242] and explicit [243, 244] rewards), the benefits of collecting additional information to reduce the uncertainty in task environments [245, 246], the memory of task-relevant objects or context cues in the environment [216], and the predicted visual state in action control [247, 248, 249]. In more recent literature, frameworks such as probabilistic decision theory [216, 250], stochastic optimal control [251, 252] have been used to explain the vision-motion coupling of human motor control, while computational models are also developed to explain, predict, and render human (-like) gaze/visual attention/active per-

ception behavior (e.g., [253, 254]).

The natural behavior and preference of vision-motion coupling not only influence how humans perform various motor skills in daily activities (e.g., [255, 256]), but also influence how humans use robot teleoperation interfaces. In robot teleoperation, whether teleoperators can make motion control decisions depend on how well they can perceive, comprehend and predict the remote task being operated [257], which further relies upon how well they can select and control the remote cameras in coordination with their tele-actions [258]. In the usage of a teleoperation interface, teleoperators will have less cognitive workload and better situational awareness, if the telepresence interface allows them to control the remote cameras similar to their natural gaze control, and if the robot autonomy for camera assistance can provide camera viewpoints and motions can accommodate their needs for performing tele-action and visual comfort [259, 260]. Such interface and autonomy are also important to the learning of robot teleoperation interfaces because it facilitates the development of spatial skills, including spatial visualization (perceiving objects among cluttered environments), mental rotation (rotation and visualization of an object to form different configurations), and perceptive taking (visualizing objects in different frames of reference) [88, 261].

**Multi-sensory Integration —** Another important human factor we need to investigate is the multi-sensory integration in the usage and learning of robot teleoperation interfaces. Similar to vision-motion coupling, the integration of visual and haptic feedback [6, 262] are also natural and essential to human motor control. The effects of haptic perception and visual-haptic sensory integration have also been investigated in various human motor behavior and motor learning processes (e.g., [263, 264]). For example, prior research has shown that haptic perception can disambiguate visual perception of 3D shape [265] and facilitates the identification of objects [262, 266]. The integration of visual and haptic feedback also facilitates the learning of tool usage [267], laparoscopic surgery skills [268]. In many multi-sensory tasks (e.g., grasping small objects), visual and haptic inputs are weighted based on the reliability of individual cues [6]. The framework of Maximum Likeli-

hood Estimation (MLE) has been used to explain the weighted integration of multi-sensory cues (e.g., visual and haptic cues), in natural and synthetic environments [6, 269]. The haptic feedback provided by robot teleoperation interfaces, although limited in its accuracy, transparency, and sensitivity, can still be leveraged to compensate for lost information in the visual feedback via remote cameras.

## 3.3 Simulated Active Telepresence Setup

In direct robot teleoperation, natural perception-action coupling in human motor control cannot be preserved due to the dissimilarity of human and robot embodiment. The added complexity of controlling the robot and vision through a motion capture system [1] might make active camera selection and control during teleoperation harder. The strong spatial skills and high mental effort required to expertly perform vision control during teleoperation might set up high barriers to entry to teleoperation. As a result, we studied human perception-action coordination in a simulated telepresence setup, where participants wearing a head-mounted display received video feeds from cameras attached to their own bodies, thereby trivializing the manipulation component of the task to encourage active camera selection and control.


Figure 3.2: A representation of the experimental paradigm. The three images show the sequence of actions the subject uses to stack a cup while performing the experiment.

We present a novel experimental paradigm to study the perception-action coordination in the

usage of active telepresence through the stacking of cups as seen in Figure 3.2. The participants wore thick gloves to dampen haptic perception in the hands and hinder their sensation of friction for grasping and in-hand manipulations. This paradigm is designed to trivialize motion control for manipulation, locomotion, and active telepresence, while preserving the essential perception capabilities and challenges of remote robots (e.g., 2D display, limited visual range, haptic feedback, unnatural control of camera motions). Voice commands via a wireless microphone were used to switch cameras to make this operation as straightforward as possible without interfering with the experiments. The camera switch is automated based on the command received from the participant. Participants naturally reveal effective perception-action coordination strategies as they adapt to the camera configuration and discover their preferred camera selection and control. Since the task of stacking cups is simple and straightforward, experience or skill played little role in the successful completion of the task.

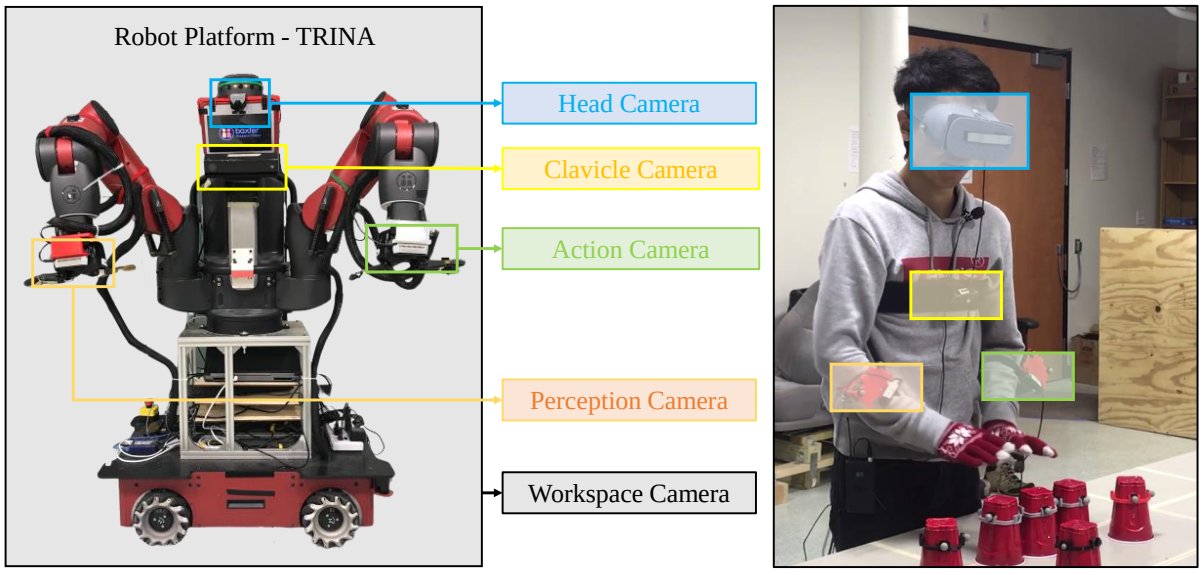



Figure 3.3: The camera set-up on the operator (right) is similar to the camera set-up seen on the TRINA humanoid robot (left). The two wrist cameras correspond to perception and action hand cameras. The gloves are used to dampen haptic perception in the hands while performing the experiments.

The participants were instructed to perform a cup-stacking task with the camera views from

various wearable and standalone cameras streamed to a VR headset. While the simultaneous display of visual feedback side by side from multiple telepresence cameras is a solution, the cognitive workload and distraction caused by this implementation can prove to be a major obstacle for teleoperators [270]. Shown in Figure 3.3, these telepresence cameras were chosen to simulate the perception cameras equipped on a mobile humanoid nursing robot, which can perform manipulation and navigation tasks under direct teleoperation [1].

The cameras available to the participants are shown in Figure 3.4 and detail listed below.

**Head Camera ($C_{head}$) —** This camera was attached to the front of the VR headset using a strap, matching natural human eyesight.

**Clavicle Camera ($C_{clavicle}$) —** This camera was attached to the chest above the sternum and between the underarms via a strap and mimicked the limited degrees of freedom and range of motion of a robot head camera.

**Action Camera ($C_{action}$) & Perception Camera ($C_{perception}$) —** These cameras were mounted on the 3D printing camera mount and then attached to the dominant hand primarily responsible for manipulation and the non-dominant hand that assists manipulation using straps, respectively.

**Workspace Camera ($C_{workspace}$) —** This camera was located across the workspace from the participant on a stationary tripod.

The head, perception, and action hand cameras were the Logitech C310 HD web-camera [271] which has a maximum resolution of 1280 x 720 pixels at 30 frames per second and a diagonal field of view of 60°. The workspace camera was a AUSDOM AW615 webcamera [272] which has a maximum resolution of 1280 x 720 pixels at 30 frames per second with a field of view of 65°. The Google Daydream VR headset with an iPhone 8 mobile phone for the display was used as the Virtual Reality headset for the experiment.

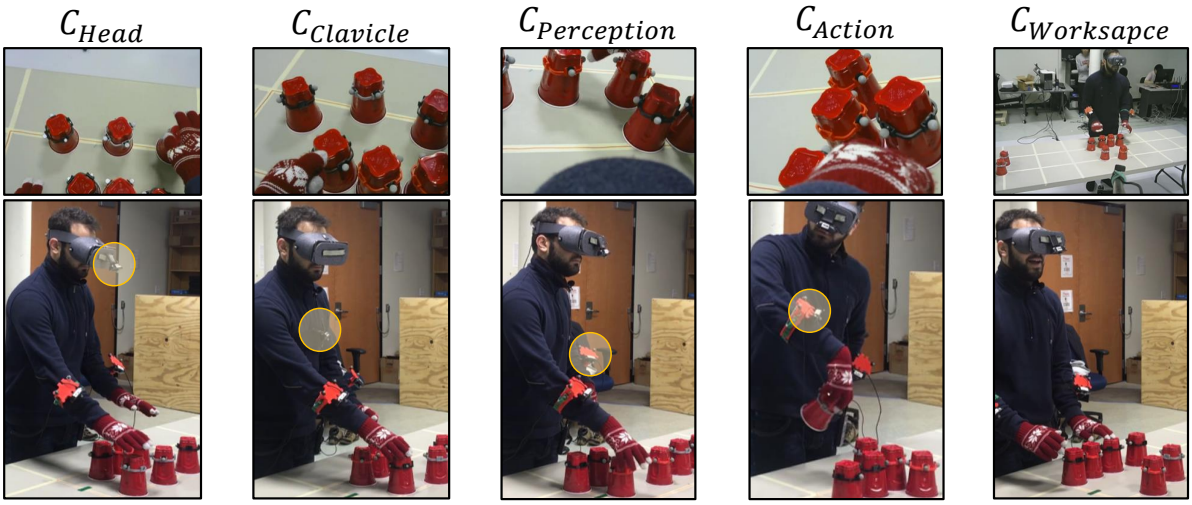



Figure 3.4: The demonstration of the video streams from head, clavicle, perception, action, and workspace cameras.

## 3.4 Human Experiment

### 3.4.1 Participants and Tasks

Our study recruited 16 healthy participants (8 males, 8 females, average age = 23.4 ± 3.6) including student and general populations. The experimental protocol was approved by WPI's Institutional Review Board.

We designed the task to be simple to understand and perform. The stacking task involved three distinct actions: (1) world exploration to observe the environment without interaction, (2) gross manipulation to reach for and carry objects, and (3) fine manipulation of objects with hands. These actions, and combinations thereof, span a wide variety of tasks a tele-manipulation system may need to perform. The cups were easy to grasp and manipulate, yet their low-friction surface and light weight made manipulation errors easy to observe.

### 3.4.2 Experimental Procedure

Before the experiment, the experimenter equipped the participant with wearable cameras, a VR headset, a microphone, and gloves, and introduced the task of stacking lightweight plastic cups into a pyramid. Also, participants were allowed to make small adjustments to the camera field of view to their preference. The available camera adjustments include:

**Head Camera ($C_{head}$) —** The angle between the front of the VR headset and the camera lens.

**Clavicle Camera ($C_{clavicle}$) —** The angle between the sternum mounting strap and the camera lens.

**Action Camera ($C_{action}$) & Perception Camera ($C_{perception}$) —** The location on the forearm (between the elbow and the wrist), the rotation of the mounting bracket around the forearm, and the angle between the mounting plate and the camera lens.

**Workspace Camera ($C_{workspace}$) —** The location of the camera tripod relative to the participant and workspace, the angle between the tripod mount and the camera lens, and the focal length of the camera image. The $C_{workspace}$ image was flipped horizontally based on user feedback during a pilot study.

Participants were first asked to stack six cups using the feedback from the telepresence cameras (single camera trials = 2 trials × 5 cameras). For each camera, a participant had a three-minute practice section to get familiar with the selected camera view. The first completed trial was extracted to represent the trial before practice (training phase). This second trial (performing phase) is used to evaluate the operator's skill and workload using the selected camera. The order of camera selection was randomized for each participant to minimize task-learning effects. Camera adjustment was permitted before, during, and after the practice trial, but the wearable camera locations and angles with the mounting point remained static during the performance trial. The participants were asked to prioritize the speed of completing the task (without compromising on comfort) and avoid

errors, like knocking over cups, and misaligning while stacking, when performing the task.

For the final trial, participants were instructed to stack ten cups and were able to use and switch the camera view at will (multi-camera trial). This trial did not include the head camera ($C_{head}$) because, in practice, VR telepresence systems may be uncomfortable to use for long periods of time like traditional healthcare worker schedules [84]; we used the $C_{head}$ condition to represent an ideal camera control baseline against which the other cameras can be compared. The participants were allowed to select the starting camera view of their own preference and were instructed to perform the final trial at a comfortable pace. Before the final trial, participants practiced using voice commands to switch cameras.

### 3.4.3 Data Collection and Processing

The methodology used while annotating the user study videos for identifying strategies developed with regard to perception-action coupling and human adaptation towards camera control while teleoperating will be expanded upon in the following sections. The annotation of user study videos involved two observers and one supervisor. The supervisor frequently held group discussions to address any conflicts in observations and converge on a conclusion.

**General Task Performance —** The overall task performance was measured by recording the duration of task completion, frequency of errors, and camera view utilization.

- **Task Completion Times:** The time taken to perform the experimental tasks during both the single and multi-camera trials was recorded. The task completion times help us get an objective evaluation of how a particular camera feed aids in performing a task efficiently and intuitively.

- **Number of Errors:** The number of errors that occurs during the practice and performance

phases of the single and multi-camera trials was recorded. These errors include misalignment of cups and knocking down of cups while stacking. Misalignment of cups implies cups were placed in the wrong location while stacking due to lost information from the camera's video feed. These errors help us objectively evaluate how a camera feed enables the correct performance of the tasks with sufficient visual feedback provided.

- **Camera Selection:** During the multi-camera trials the number of camera switches between the various camera views was counted. These results can help identify the preferences of the participant for completing the task and help objectively validate the responses provided by the participants' responses to the post-study survey.

**Human Behavior —** The observed behavior was classified into the following categories:

- **Instinctive Head Movement:** Participants tended to try and control their camera/vision using their head motion even when the camera is not connected to the head. The head motion was counted as a non-trivial rotation when it is along the transverse and longitudinal axes. The instances of these head motions were compared with task completion times to identify how the user instinctively desires to move their head or go for their natural mode of perception with the complexity of the camera view indicated by task completion time. These motions were counted for the training and performing phase of single camera trials and for all the camera feeds except the head and workspace camera.

- **Body Coordination:** In the performing phase of the $C_{clavicle}$ camera experiment, the participants moved their upper body or walked sideways to improve their field of vision. The instances of torso motion and walking motion were counted.

- **Bimanual Manipulation:** Bimanual Manipulation is the efficient way of performing tasks and thus the number of participants performing bimanual manipulation during the performing

phase of all the camera trials and in the multi-camera trial was counted. These motions were counted for $C_{workspace}$, $C_{clavicle}$, and $C_{head}$ when both hands were used to gather and stack cups.

- **Fixed Elbow:** During the performing phase of the $C_{perception}$, the time during which the perception camera was stationary while performing the experiment was recorded from the user study videos. This action usually involved the user holding a stationary pose for their elbow on which the perception camera was mounted with respect to their body.

- **Saccade Ahead:** We observed that some participants looked ahead at the location of the future cup placement before grasping it. This motion was counted for the training and performing phase in single camera trials and for all the camera feeds except the action hand and workspace camera.

- **Touch to Locate:** Even with limited haptic feedback, participants attempt to identify the cup location and the position of their hands using their ability to touch surfaces. We counted the number of times a hand was used to tap the bottom of the cup to identify the subject's reliance on haptic feedback. This motion was counted for the training and performing phase in the single camera trials for all the camera feeds and in the multi-camera trial.

- **Tentative Stacking:** Participants also tended to stack tentatively by tapping the bottom of the cup they are trying to stack against the surface where they intend to stack to precisely align their cup while stacking. This motion was counted for the training and performing phase in the single camera trials for all the camera feeds and in the multi-camera trial.

- **Slide Cup on Table:** We also counted the number of times the participants slid the cup across the table's surface rather than picking it up. This motion was counted for the training and performing phase in the single camera trials for all the camera feeds and in the multi-camera trial.

- **Touch for Alignment:** While in the stacking phase, we noticed that participants try to use one hand to hold the bottom cup and another hand to make the alignment. This motion was counted for the training and performing phase in the single camera trials for all the camera feeds and in the multi-camera trial.

**Subjective Survey —** The preference of participants for different camera views was verified by the time they spent using different camera views while performing the multi-camera stacking trial. A subjective camera preference survey was performed to record the participant's perceived preferences for different cameras while performing the various components involved in the stacking operation like a choice of the camera in exploring, reaching, grasping, and for the overall performance of the task. They were also asked to provide specific feedback about certain camera viewpoints and configurations and their thoughts on improving the system. Additionally, the participants also participated in a post-study interview at a later stage where the experimental video was replayed to them and questions pertaining to their reasoning for performing the action.

## 3.5 Perception-Action Coupling

We analyzed the human behavior from the performing phase in the single camera trial and combined the multi-camera trial to reveal the vision, haptic and motion coordination while performing the cup stacking task for each camera usage.

### 3.5.1 Vision-Motion Coupling

We noticed that people attempt to adjust the camera view using their head not only for the head camera but even for the action, perception, and clavicle cameras (see the head posture in Figure 3.5). The ANOVA analysis of the **Instinctive Head Movement** from the performing phase in

the single camera trial shows that using an action camera causes significantly more frequent futile head motion than the clavicle ($F(1,15)=24.4$, $p<0.01$) and perception ($F(1,15)=22.8$, $p<0.01$) cameras. We further examined the correlation between task performance (task completion time) and the instances of head movements (see Figure 3.5). A significant linear regression was found for clavicle ($F(1,13)=12.8$, $p<0.01$, with an $R^2$ of 0.5), perception ($F(1,13)=14.2$, $p<0.01$, with an $R^2$ of 0.52) and action ($F(1,13)=5.9$, $p<0.05$, with an $R^2$ of 0.32) cameras. Linear regression of this data predicts that the expected task completion time increases by approximately 9.5 (clavicle), 4.6 (perception), and 4.7 (action) seconds for each occurrence of head movements. Our interview reveals that not being able to control the camera viewpoint using their head movements caused a lot of frustration for every participant. Some participants were able to remind themselves that head movements are not effective for camera viewpoint control and try to suppress this instinct, while others only realized the head movements are ineffective for camera viewpoint control until they felt discomforts like motion sickness or physical fatigue due to activity. Overall, we found that it is more difficult for the participants to realize and suppress the instinctive head movements when the camera is considered more difficult to use.

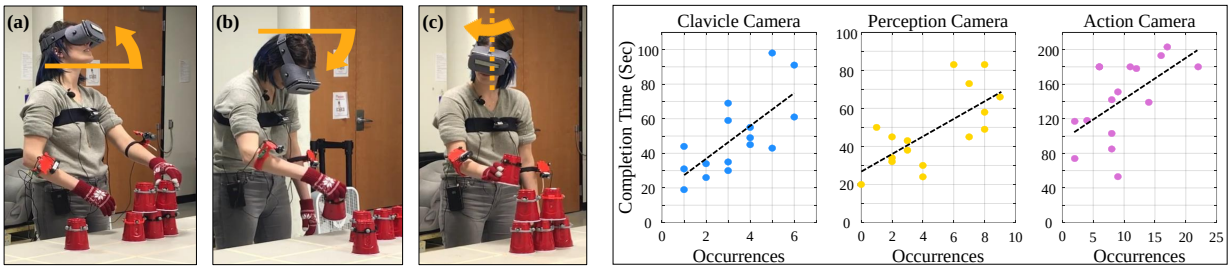



Figure 3.5: Compulsive head movement: (a) raise the head up; (b) hold head down; (c) turn head side way. Task completion time versus the occurrences of head movement for the clavicle, perception, and action hand camera.

Based on the usage of the clavicle camera ($C_{clavicle}$), we found that the participants can be separated into two groups by their **Body Coordination**. The result from the performing phase (Figure 3.6(a)) shows that one group of participants tend to explore the environment through torso

motions to control the camera view instead of walking around while the other group walked around in the workspace for the same.

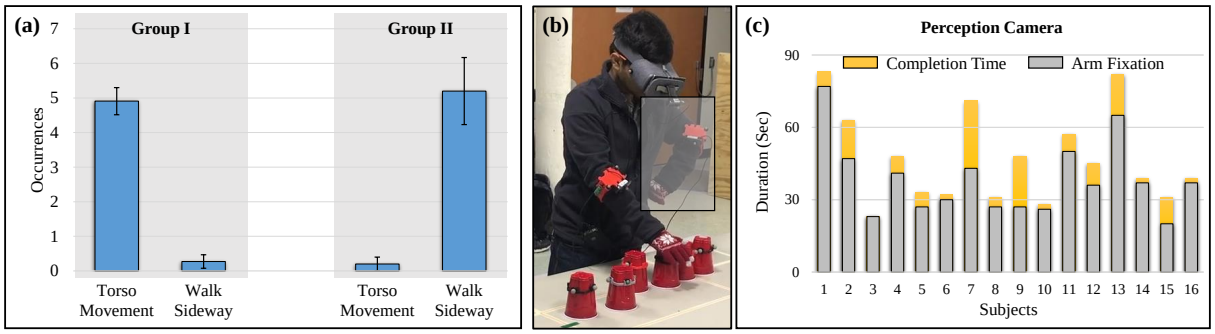



Figure 3.6: (a) Two groups of the body coordination while using clavicle camera; (b) The fixed elbow pose for perception camera control; (c) Duration of the arm fixation w.r.t. task completion in perception camera usage.

As shown in Figure 3.6(c), the proportion of task time that the participants employed a **Fixed Elbow Posture** (refer Figure 3.6(b)) while using the perception camera for the fixed camera trial was 82.9 $\pm$ 12.7 percent. We also found that the majority of the participants (11 of 16) tend to fix their shoulder joints and move their torso to control the perception camera viewpoint, thus limiting the perception hand camera motions with respect to the base frame of the torso. Our interview reveals that: most participants intentionally limit the elbow and shoulder motions of the perception camera arm to better remember the spatial relationship of the perception hand camera with respect to their body. This lets them coordinate the camera motions with the motions of their manipulating hand, object, and workspace. Some participants indicated that they unconsciously choose the elbow angle so that the perception camera is not too far away from their body, making it easy and comfortable to move and look around the workspace. Overall, the situational awareness of the perception camera pose with respect to their body is critical to the planning of coordinated perception and manipulation actions.

Whenever possible, participants preferred **Bimanual Manipulation**, to speed up the task and to increase their reach without moving the body. The usage of bimanual manipulation, in both sym-

metric and asymmetric forms, is observed when using the head, clavicle, and workspace cameras, for reaching to collect cups, and for placing/stacking the cups in the same row. We also found that bimanual control/manipulation is more frequent with the head camera (13/16 participants) than with the clavicle (3/16 participants) and workspace cameras (4/16 participants). Our interview shows that bimanual manipulations are more difficult when using the clavicle camera because reaching both hands forward to objects caused the torso to lean forward which reduces the viewpoint control of the clavicle camera. Compared to unimanual manipulation, bimanual manipulation is more efficient yet more complex to plan.

### 3.5.2 Haptic-Motion Coupling

Our experimental paradigm limited the haptic perception of the participants so that they had to rely mostly on the visual feedback from RGB cameras to perform the tasks. However, participants still learned to utilize the limited haptic feedback received through the thick gloves they wore to compensate for reduced visual feedback. Across all the participants and camera viewpoints, we observed the participants 1) **Touching to Locate** the cups to build the *contact* sensation, 2) **Sliding Cup on the Table** so that they can leverage the haptic perception of table *constraints* to better control the moving motions, 3) **Stacking Tentatively** to get the better placement location and 4) **Touching for Alignment** using the bottom of the cup.

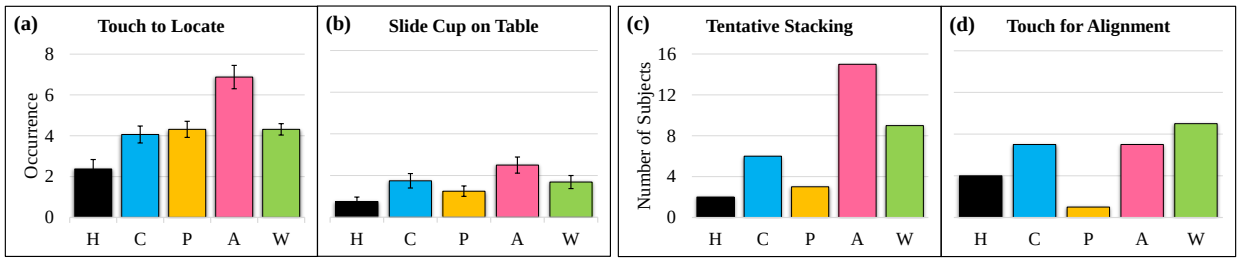



Figure 3.7: (a) Touching-to-locate, (b) sliding cups-on-table, (c) tentative-stacking and (d) touching-for-alignment actions observed in the usage of the head (H), clavicle (C), perception (P), action (A) and workspace (W) camera.

The haptic-motion coupling actions like touch to locate, sliding the cup on the table, tentative stacking, and touching for alignment were counted. Figure 3.7(a) shows the mean and standard deviation of touch-to-locate occurrences across participants for different cameras. The ANOVA analysis shows that using an action camera causes significantly more frequent ($p < 0.01$) touch-to-locate actions than all other cameras. Also, touch-to-locate actions occurred the least ($p < 0.01$) when using the head camera. These significant differences indicate that participants resort more to haptic feedback for the cameras more difficult to use (as indicated in our survey feedback). Both the observed human behavior and the interview feedback indicate that 1) touching-to-locate an object is the most necessary haptic perception to complement the loss of depth information and limited field of view while using active telepresence; 2) the haptic feedback does not have to be strong and realistic if it can provide a sense of contact. We hypothesize that this can largely reduce the mental workload and stress due to uncertainty in perception while improving task accuracy and efficiency.

In addition to touch-to-locate, participants also used touch-for-alignment when tentatively stacking, aligning, and sliding the cups on the table. Overall, haptic compensations were required for the cameras identified as non-intuitive and inefficient to use. In Figure 3.7(b), sliding cups on the table are observed the most in action hand camera usage. On the other hand, the tentative stacking actions are used by 15 of 16 participants when working with the action hand camera, and by 2 of 16 participants when working with the head camera (see Figure 3.7(c)). While in Figure 3.7(d), touch for alignment is observed in more than half of the participants for the workspace cameras followed by action hand and clavicle cameras. The interview feedback reveals that: 1) The gloves effectively damped most of their haptic perception; 2) the limited tactile sensing is still very helpful to the task in many cases.

### 3.5.3 Vision-Haptic Coupling

In the single camera trial, the participant used the video feedback from a single camera to perform the cup stacking task. This helps us compare the performance and human behavior across cameras. Figure 3.8(a) shows the concept of vision-haptic coupling, where information gathering while touching to locate is offset by the increased number of camera switches and vice-versa. Unlike the single camera trial that receives vision feedback from one camera viewpoint, the multi-camera trial allows participants to switch the viewpoint across cameras. The need for haptic compensation like loss of depth information can be compensated by switching the camera view to a different view like the perception hand camera. However, our interview feedback suggests that the cognitive workload increases when having to involve more camera switches thus limiting the bandwidth to perform other actions.

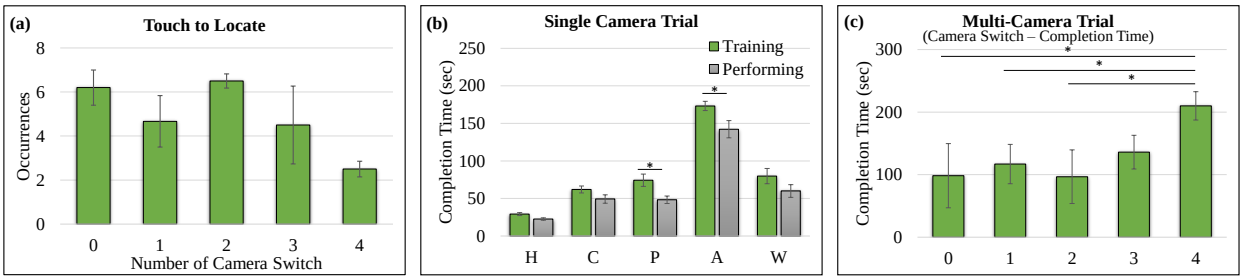



Figure 3.8: (a) The correlation between camera switches and touch to locate the action in the multi-camera trial; (b) The comparison of the completion time between training and performing phase in single camera trial. (c) The comparison of the completion time across the number of camera switches in the multi-camera trial.

## 3.6 Human Adaptation

We performed the analysis of human behavior in the single camera trial to investigate perception-action coupling while using the active telepresence cameras. In this section, we further compare the performance and the human behavior between the training and performing phases in the single

camera trial to disclose the impact of the practice and motor learning process. We also combine the information from the multi-camera trial to better understand how the skill sets learned from the single camera trial transfer to the multi-camera trial.

### 3.6.1 General Performance

We compared the performance in terms of *task completion time* and *number of errors* between the training and performing phase in the single camera trial. An error can occur when the cup drops due to: 1) misalignment during stacking, and 2) collision while moving hands around. Figure 3.8(b) shows the comparison of the task completion times between the training and performing phases in the single camera trial. The ANOVA analysis shows that the completion time had significantly reduced after practice while using the perception ($F(1,30)=7.3$, $p<0.05$) and action ($F(1,30)=5.6$, $p<0.05$) hand camera. These significant differences indicate that the comprehensive practice section is necessary for difficult cameras. In addition, Figure 3.8(c) shows the correlation between task completion time and the number of camera switches in the multi-camera trial. Based on the ANOVA analysis, the time needed to complete the task is significantly longer when participants had the most number of camera switches than switches twice ($F(1,4)=11.3$, $p<0.05$), once ($F(1,3)=12.5$, $p<0.05$) and none ($F(1,5)=8.1$, $p<0.05$).

Figure 3.9(a) shows that trials using the action hand camera had more participants who misaligned cups compared to the other camera views and displayed limited improvement after practice (11/16 to 9/16 participants). This implies the non-intuitive camera usage in terms of loss of depth information which may lead to failure of the task despite the practice session. In Figure 3.9(b), the action hand camera still caused most participants to knock down the cup while moving their hands around. However, the practice helped prevent the collision with the cup when using: perception (3/16 to 0/16 participants), action (8/16 to 4/16 participants), and workspace (4/16 to 2/16 partici-

pants) cameras. This implies the narrow field of view and complex camera control can be adapted to by practice.

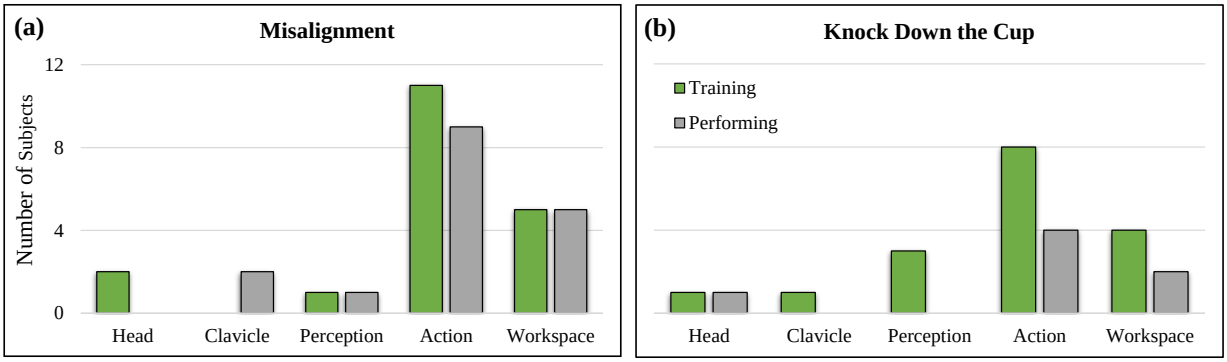



Figure 3.9: The errors occurred in the single camera trial in the type of (a) misalignment; (b) colliding with the cup.

### 3.6.2 Motor Learning

We identified several actions that we constantly observed from both the training and performing phases in the single camera trial. As shown in Figure 3.10(a), we found no significant differences for the **Instinctive Head Movement** in the clavicle, perception, and action hand cameras between the training and performing phase. This solidifies that it is difficult to suppress the head movement though participants are able to realize that the camera cannot be controlled by the head during the training phase. Figure 3.10(b) shows the duration (the proportion with respect to task completion time) of fixing the elbow in a certain posture while using the perception hand camera (including the training and performing phase). We found that the duration of the **Fixed Elbow posture** significantly reduces ($F(1,30)=13.5$, $p<0.01$) after practice. This implies that the training session helped improve the participant's understanding of the spatial relationship of the perception of hand camera with respect to their body. We noticed that some participants made a **Saccade Ahead** of the cup, just before grasping it, to a location on the future placement. In Figure 3.10(c), there is a noticeable increase in the participants (from 3/16 to 9/16) who looked ahead when performing pick-and-place

motion in the perception hand camera trial after the practice session. This observation implies the practice section can improve the cognitive bandwidth when controlling a non-intuitive camera.

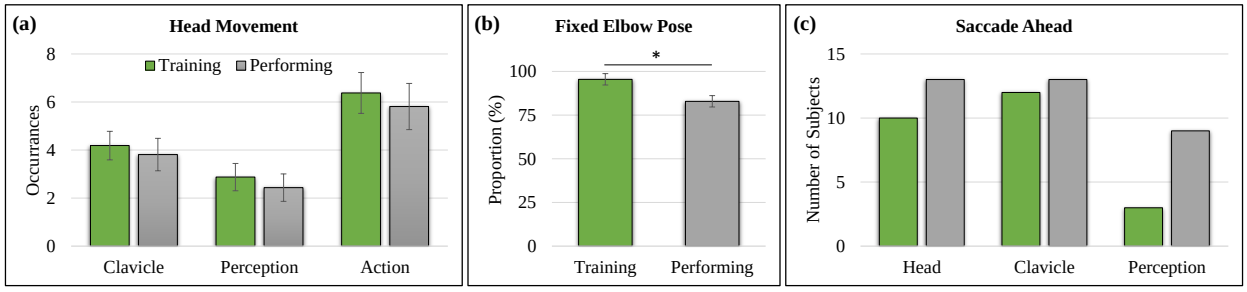



Figure 3.10: The comparison of the human behavior between training and performing phase for: (a) head movement, (b) arm fixation, and (c) looking ahead while using active telepresence camera.

We divided the **Bimanual Manipulation** into gathering (maneuvering multiple cups in the workspace), picking/stacking (picking up and stacking actions of the cup in the workspace), and holding the cup (holding and carrying a cup). Figure 3.11(a) and Figure 3.11(b) show the increase in the number of participants who performed the bimanual gathering and picking/stacking after practice in the head camera trial which was identified as the most intuitive camera view to control. However, lesser participants used both hands to hold the cup after practice while using the workspace camera (see Figure 3.11(c)). Our interview feedback indicates that they try to eliminate the mirror effect that occurs while the workspace camera by holding a cup with both hands.

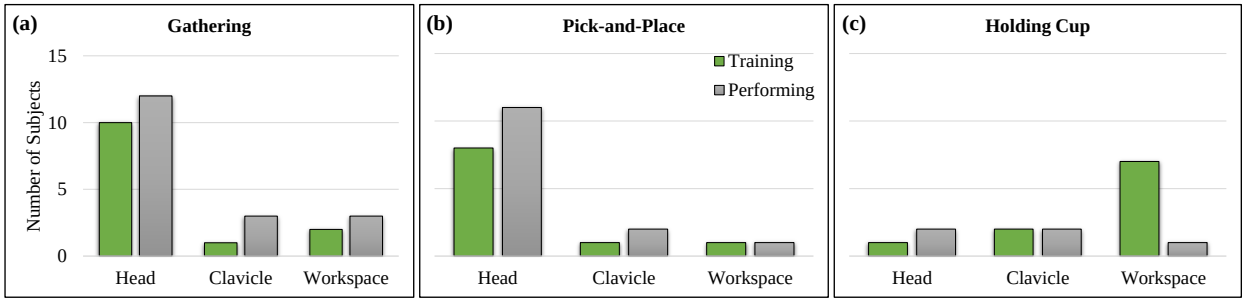



Figure 3.11: Bimanual manipulation for: (a) gathering, (b) pick-and-place, and (c) holding a cup.

We also investigated the differences in **Haptic Compensations** between the training phase and performing phase to better understand how humans adapt to the different active telepresence

cameras usage. Figure 3.12(a) and Figure 3.12(b) show the mean and standard deviation of touch-to-locate and slide-cup-on-table occurrences across participants for different cameras in the training and performing phase. The ANOVA analysis indicates that using the perception ($F(1,30)=9.5$, $p<0.01$) and action ($F(1,30)=7.1$, $p<0.05$) hand cameras significantly reduces the touch-to-locate actions and perception hand camera significantly reduces ($F(1,30)=5.9$, $p<0.05$) the slide-cup-on-table action after practice. These differences imply that haptic feedback helped improve the usage of the more difficult, limited field-of-view cameras by virtue of being close to and focused on the object of manipulation. Figure 3.12(c) and Figure 3.12(d) show the number of participants who performed the tentative stacking and touch-to-alignment actions for different cameras in the training and performing phase. We found a slight decrease in the participants who performed the tentative stacking while using the perception hand camera and an increase in the participants who performed the touch-for-alignment in the action hand camera after practice.

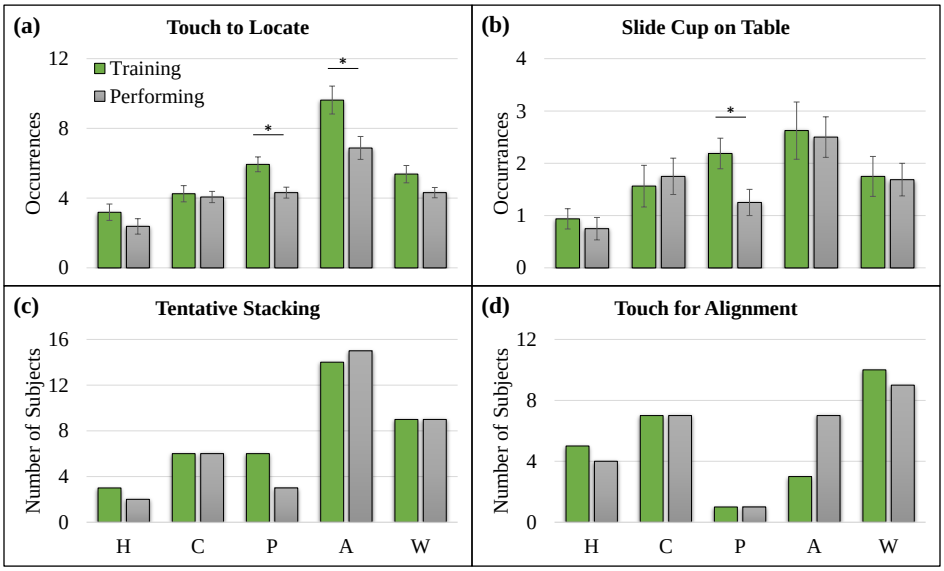



Figure 3.12: Comparison of the haptic compensation between training and performing phase including: (a) touch-to-locate, (b) slide cup on the table, (c) tentative stacking, and (d) touch-for-alignment.

We further analyzed the identified actions including head movement, bimanual operation, and haptic compensations in multi-camera trials to investigate the process of human adaptation in the

usage of active telepresence cameras. As shown in Figure 3.13, more than half of the participants use the touch-to-locate (16/16 participants), slide-cup-on-table (15/16 participants), bimanual manipulation (13/16 participants) and touch-for-alignment (9/16 participants) actions while using the active telepresence cameras. Furthermore, we observed 13 out of 16 participants still tried to control the camera viewpoint using their heads.

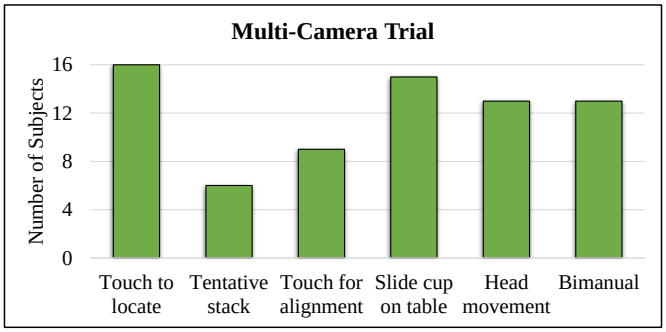



Figure 3.13: The number of subjects who performed the haptic compensation, head movement, and bimanual manipulation in the multi-camera trial.

## 3.7 Camera Selection and Preference

We conducted the analysis of the camera preference as indicated by camera selection while performing the multi-camera trial and post-study survey. Figure 3.14(a) shows the correlation between the duration of camera usage and the number of camera switches in the order of total task completion time. We found that fewer camera switches and participants who had a higher proportion of clavicle camera usage led to better performance (in terms of task completion time). These observations aligned with the recent study of multi-view interface design where it was observed that autonomous switching might ease the control effort [273]. The camera preference survey indicates that the workspace camera is preferred while exploring the environment and the perception hand camera for gross and fine manipulation followed by the clavicle camera (see Figure 3.14(b)). It is to be noted that the action hand camera was never selected during the multi-camera experiment.

From a human action perspective, these results aligned with the findings from the recent design of the adaptive viewpoint in telemanipulation where the performance was improved with the shared-autonomous camera control [31].

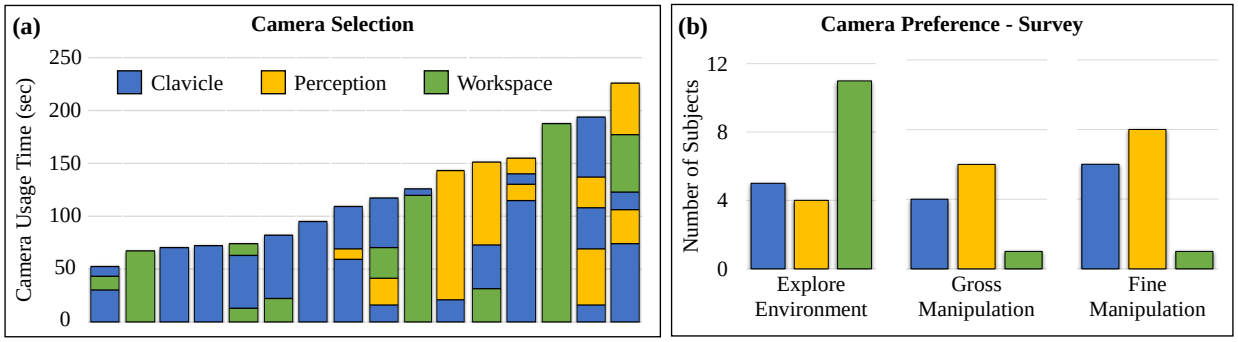



Figure 3.14: The subjective assessment about the camera selection and preference from the multi-camera trial.

## 3.8 Summary and Outlook

In this chapter, we investigate how humans coordinate perception-action coupling during active telepresence through a novel experimental paradigm that emphasizes limited haptic feedback. The results from participant task performance, human motion analysis, and user feedback reveal the integration of vision, motion and haptic feedback, human natural motor learning, and preferences. The key findings and suggested designs are:

(1) **Intuitive Control of Multi-Camera Active Telepresence** — The instinctive head motion we constantly observed to not only control the head camera but the cameras attached to their torso and hands indicates the head should control for any camera selected for telepresence.

(2) **Active Telepresence Assistance for Supervisory Control** — The participants intended to maintain the arm posture for better spatial awareness of camera pose implies that the motions

of the shared autonomy camera should follow the simple translation or rotation to make it easier to understand and predict by the users.

(3) **The Need for Visuo-Haptic Sensory Integration** — People tend to resort to using every possible haptic sensation to compensate for the limitation of the visual feedback reiterating the importance of integrating vision and haptic feedback in robot teleoperation interfaces.

We further discuss the implication of suitable camera control and selection as well as the preferable robot teleoperation interface design.

**Desirable Characteristics of Viewpoint Control and Selection —** Tele-nursing robots need different viewpoints from strategically placed telepresence cameras to provide a comprehensive view of the environment and the task workspace. A natural approach to control and select the cameras are necessary to reduce the cognitive workload and increase the transparency in robot teleoperation. The findings from our human motion analysis identify several components for camera control and selection so that the perception-action coupling complies with the natural human motor control.

As human tracking technologies become more accurate, portable and affordable, head- and gaze-control are getting increasingly adopted for the control of eye-in-hand cameras of manipulator and continuum robots [274], and the head camera of mobile and humanoid robots [275, 276]. While matching human eyes to robot eyes is considered to be a natural design choice, it is also not rare to see remote cameras controlled by robot hands. When multiple cameras are available (as on many commercial and prototype humanoid robot platforms [1, 3]), head and hand control are usually only used for the head and eye-in-hand cameras, respectively. When a teleoperator switches their primary viewpoint (i.e, the camera view they primarily rely upon to perform the task) from the head-to-hand camera, adapting to control of camera viewpoint via hands always causes interruption of task performance. Lessons learned from (tele-robotic) laparoscopic surgery training also indicate that it takes much more training effort to learn to use hand-controlled cameras [277]. The intuitive

nature of the head motion observed in the clavicle and hand camera trials, highlights the need for egocentric control (usage of head to control gaze) to control any camera viewpoint selected as the primary viewpoint. This head-controlled dynamic viewpoint aligned with the recent mobile manipulator implementation [278].

In direct teleoperation, understanding the camera pose and motions is critical to control the robot action components (e.g., end-effectors, mobile base). Even in supervisory control, lack of spatial awareness due to sub-optimal camera pose will reduce the operator's situational awareness and capability to intervene if the robot autonomy is not reliable [279]. The elbow joint fixation we observed from the single-camera trial highlights the strategy that humans adopted to maintain the spatial awareness of the camera poses with respect to their bodies. A preferable method for camera control thus should limit the degree of freedom to be controlled by simple translation or rotation. In the case of supervisory control, the trajectory of the autonomous camera system should be easy to understand and predictable for the operator. Learning preferences for camera viewpoints for specific tasks increases situational awareness crucial for supervisory control of remote robots.

In the fixed camera usage, our study reveals that the camera which is intuitive to use is preferred which leads to better performance (faster completion time and fewer errors) and lower cognitive workload. When people have more cognitive bandwidth, they are able to perform complex motions. This is supported by the fact that most participants perform bimanual operations and look-ahead motions to place the cups when using the head camera. On the other hand, our camera preference survey indicates the correlation between the purpose of the action and the preferred camera for this action in a multi-camera setup. However, the camera choice in the multi-camera trial shows a large variance with no consistency across participants. These outcomes imply that customization of autonomous camera selection with respect to user groups, or even personalization, is necessary.

As part of our future work, we will further develop an intuitive method to control multi-camera active telepresence. A user study will also be devised to investigate if the perception and action

hand camera could be controlled using the head, hand, or a mixture of head and hand motions as well as to understand the human behavior, preference, and rationals in the usage of a multi-camera active telepresence system. The entire experiment was performed in a motion capture enclosure with motion capture markers located on the VR headset, wrist camera mounts, and cups. However, the motion capture data did not yield any significant results due to the lack of quality. We will further utilize the VR trackers to get meaningful data to investigate human behavior objectively. We will also explore the use of autonomous camera control and selection to reduce operator workload and improve task performance in supervisory control.

**Design Philosophy for Multi-sensory Integration —** The experiment paradigm enables participants to manipulate the object with their own hands, which is more capable of moving and sensing through proprioception. In object manipulation, the benefit of proprioception is limited because visual information is still required to locate the target and a freely moving arm will not help in locating the object. The feedback from participants also indicated that they need to place their hand in the view to better understand the relationship between the arm and the target object implying the limited usage of proprioception during object manipulation. However, if proprioception combines with the human's memory of the workspace, it will indeed ease the effort in object manipulation because the direction towards the target can be identified.

Our human motion analysis indicates that people tend to use haptic feedback to compensate for the loss of depth information and narrow limited vision of the visual feedback via active telepresence. The desire for haptic feedback ranges from precise or gross manipulation to general environment exploration. Indeed, human motor control has the instinct to pursue visual-haptic sensory integration when they perform tasks with their own bodies as well as via teleoperation interfaces. Unfortunately, state-of-the-art haptic feedback rendering technologies cannot enable the teleoperation interface to provide the most realistic haptic perception. The leverage between the complexity and what is the suitable level of the haptic feedback to compensate for the limitation of active telep-

resence visual feedback needs to be studied. Our study reveals that: 1) human motor control can achieve very effective visual-haptic sensory integration with active telepresence visual feedback and limited haptic feedback; 2) for general-purpose manipulation tasks, adding a little bit of haptic feedback to indicate the contacts with the remote physical environment will be much more simple and effective than fabricating complicate strategies for the optimization of camera control and selection.

Inspired by findings from our study, we propose a philosophy for visual-haptic sensory integration to re-establish the perception-motion coupling with the perception and action capabilities of the remote robotic system. From a high-level perspective, there are three strategies to achieve this goal. Take several designs in literature and our prior work for example: 1) we may **restore** the lost haptic perception by adding vibrotactile feedback to indicate contacts with the remote environment [280]; 2) we may also **replace** haptic display with augmented reality visual display [281]; 3) we may **delegate** the task components that heavily rely upon haptic feedback to reliable robot autonomy, to eliminate the need for remote perception-action coupling [282]. Our future work will implement the proposed philosophy and conduct a user study to compare the efficacy of each haptic compensation approach and user acceptance as well as preference for the use of robot teleoperation.

**Considerations —** The limitations of our current testing suggest many extensions.

- **Advanced Gaze Analysis:** There was no gaze detection used in this paper. Our future work regarding perception studies will involve the utilization of a gaze tracker to accurately track human gaze motion and collect more reliable data. This will help us accurately determine where the operator is looking at different parts of the task improving our ability to draw information regarding camera view usage.

- **Influence of Human Sensation:** The participants felt that the usage of multiple gloves effectively reduced the haptic feedback while performing the task. However, a systematic manner

to dampen the haptic sensation was not implemented. Studying the impacts of varying levels of haptic dampening and their impact on camera interface control will be an interesting avenue for future research.

- **Integration with Teleoperation Systems:** Ideal teleoperation needs to consider both remote perception and robot control. As the first step, the proposed experiment paradigm provided the simulated telepresence setting and retained the human's ability to manipulate the object which relaxed the control effort and focused on the investigation of active telepresence design in remote perception. A further investigation of teleoperating the robot with the preferred active telepresence design will be conducted along with the suitable robot control interface.

# Chapter 4

# Human-preferred Remote Manipulation Assistance

## 4.1 Motivation

Contemporary motion tracking interfaces (e.g., HTC Vive virtual reality system [34]) enable manipulator robots to track the natural human arm and hand motions to perform more dexterous, freeform manipulation. While human operators can efficiently and intuitively control gross manipulation (e.g., reaching to or moving an object), they may experience significant cognitive and physical workload when controlling precise manipulation, such as carefully adjusting the robot end-effector near an object for grasping or placing. This is usually because human operators can not acquire the necessary sensory information (e.g., visual or haptic) to perceive and control the remote tasks [35, 36]. For example, the operators may need the camera viewpoint from a different perspective to perceive the depth information not available in the primary camera viewpoint. They may also need proprioceptive and haptic feedback to control the end-effector's motions or postures precisely. Besides the remote perception problems, the cognitive and physical workload may also come from the difficulties in remote robot motion control. Freeform manipulation tasks typically involve both gross and precise manipulation which can be difficult to perform efficiently through interfaces designed around motion tracking. Changing the interface mapping and scaling (from human-controlled inputs to robot motion outputs) based on task state or user input will be

required to control the robot with efficiency and precision. Related work has proposed to various approaches assist human teleoperators' remote perception and motion control. The existing solutions regarding the design of the tele-manipulation interfaces include the methods to: 1) improve the capabilities of the interface to display additional sensory information (e.g., multi-camera viewpoints, haptic interfaces [1, 283]); 2) Resort to alternative sensory feedback to present the missing information, such as using augmented reality (AR) visual cues to represent the remote contact or force feedback [284]; 3) Delegate the part of the tele-manipulation task difficult for humans to capable and reliable autonomy [28] so that the interface only needs to present feedback on the autonomy's performance instead of the detailed sensory information. However, related work in the literature mostly compares the same types of approaches to validate the effectiveness of their proposed methods. There is still no work to compare different types of approaches, to inform how to choose among or integrate them when multiple types of approaches are available.

Another problem we are concerned with is: how to combine human operator and robot autonomy to optimally control tele-manipulation? Our insight from the related work in the literature is that: shared autonomy to assist tele-manipulation can be more effective if it is designed to enable an appropriate division of task and effort between the human and robot. Such task division should allow humans to have sufficient freeform control to perform the unstructured parts of the task and allow robot autonomy to handle the structured parts of the task with desired performance (e.g., speed, accuracy, reliability). In recent related work, the shared autonomy to assist remote manipulation is mostly designed to assist as much and as early as possible, based on the prediction of human intents (e.g., target object [285], expected motion trajectory [286]). These shared autonomy designs are designed to minimize control inputs and efforts, and may not always be necessary and effective to assist the operators who could prefer to have more freedom than assistance to control gross manipulation. While humans can easily perform freeform reaching motions to clearly indicate the object they intend to grasp, the most effective way to reduce the human workload is

to provide autonomy only to the part of the task that causes humans high cognitive or physical workload. This chapter delves into the topic of reducing interface complexity and improving communication between humans and robots during remote manipulation. To achieve this, we present the effective techniques, which are:

- **Multi-sensory Integration:** This technique involves the use of haptic and augmented reality visual cues to enhance the perception of the robot's movements and the workspace.
- **Perception Augmentation:** By providing additional viewpoints, this technique allows for different perspectives on visual feedback, thereby increasing the accuracy of the human's perception of the robot's actions.
- **Action Augmentation:** This technique involves the use of assistive autonomy and constrained/scaled motion to enable more precise control over the robot's actions.
- **Augmentation Integration:** By providing appropriate AR features tailored to varying levels of robot autonomy, this technique enhances transparency in human-robot collaboration.

## 4.2 Literature Review

**Multi-modal Sensory Integration —** For decades, various haptic and AR visual cues have been designed, evaluated, and compared for a wide range of robot teleoperation tasks. Overall, haptic feedback (in terms of force and vibrotactile display) is usually used to communicate continuous feedback that requires a time-critical response, yet the precision and the amount of information encoded are limited. On the other hand, AR visual cues can communicate very rich, detailed information using a variety of colors, shapes, and displayed text. However, it takes more time for a human to respond to visual feedback (around $500$ ms) and may cause visual crowding and distract the user's

attention. As a result, haptic feedback has been widely used for driving assistance (e.g., safe lateral control [287], braking assistance [288], cruise control [289], curve negotiation [290], avoiding collision with obstacles and pedestrians [291, 292], etc) in terms of shared haptic control [293], as well as teleoperation assistance for mobile/wheelchair robots (e.g., trajectory guidance [294, 295], cooperative navigation [296, 297], etc.). For tele-manipulation tasks, haptic feedback in terms of virtual fixtures [298, 299] has many benefits in terms of control precision and manipulation safety [280, 283], user comfort [26], and coordination of multi-robot systems [300, 301, 302]. The fusion of force and tactile feedback can also enhance the remote perception for unknown object identification [303]. Meanwhile, AR visual cues are preferred to assist the estimation of spatial relationship (e.g., gap estimation for driving assistance [304]), to direct and enhance visual attention (of drivers [305, 306], and video game players [307]). More recently, AR visual cues are used to communicate the intent between humans and robots [7, 27, 308] and enable robot autonomy to become more explainable [309]. The various designs of haptic and AR visual cues also enable the integration of multi-modal sensory feedback, which is natural to human sensorimotor control. Indeed, the integration of visual and haptic feedback is critical to the precise manipulation tasks, because it plays an important role in the perception of object shape [262, 265, 266, 310] and object dynamics [311], and in the performance of precise pointing and grasping movements [265, 312], and in the usage of (articulated) tools [313, 314]. Related work in the literature has investigated the dominance, weighting, and the roles of attention in human integration of visual and haptic feedback [6, 315], and compared haptic and visual feedback for their effectiveness on body posture guidance [316], on the control of dexterous tele-manipulation [317], and motor learning [318, 319, 320]. Recent designs of sensory integration have integrated AR visual displays which visualize robot model and end-effector, or interaction force, with the haptic cues which indicate task guidance or constraints [301, 321, 322, 323, 324, 325, 326, 327, 328, 329]. Although these designs of haptic and AR visual cues, as well as their integration, have been validated to be effective in user studies, these designs and validations are mostly specific to tasks and robots. There is limited work

that systematically investigated how the design of visual-haptic sensory integration depends on the nature of conveyed information and the modality that the information is conveyed through [329]. It is also unclear how the design of sensory integration should be changed if the robot teleoperator needs to perform secondary tasks that demand additional visual attention, awareness of the haptic/vibrotactile feedback from other wearable interfaces, or efforts for critical thinking.

**Action and Perception Augmentation —** Table 4.1 categorizes the conventional and contemporary control interfaces for assisted tele-manipulation that represent the state-of-the-art. Compared to conventional interfaces, contemporary interfaces tends to: 1) improve the control *dexterity* of high degrees-of-freedom motion coordination (e.g., multi-finger coordination, hand-arm coordination), and simultaneous position and orientation control; 2) improve the *intuitiveness* of manipulation control, either by mapping natural human motions to robots, or replicating/representing the controlled robots or the manipulated objects (e.g., using 3D-printed prototypes or virtual reality). 3) improve control *accuracy* by providing (shared) autonomy with/out haptic feedback. Considering the tele-manipulation assistance in recent related work, we also found that: while the **action support** that (partially) automates task-specific manipulation actions can improve the control accuracy and are used more for structured tasks [330, 331], **control augmentation**, such as the design of *interface mapping* and *scaling*, can better enhance the dexterity and intuitiveness, can be generalized across various interfaces, and are more used for freeform manipulation [332]. While being intuitive, motion tracking interfaces are generally limited in their control efficiency. This is a result of the limited accuracy of human motions, and interference of intended and unintended motions due to simultaneous control of many Degrees of Freedom (DOFs). The efficiency of the controlled motions can be improved by introducing constraints in terms of virtual fixtures [299] or autonomy for teleoperation assistance (e.g. collision avoidance [325], motion guidance toward intended goal [9]). From a more general perspective, the interfaces that map gestures or point-and-click actions to autonomous robot motions or movement primitives [333] can all be considered as some kind of

constraints that limit the extent to which human operator can control the robot motion freely. In addition, motion constraints can also be introduced by the separation of degrees of freedom (DOFs) in the design of interface mapping. For instance, people may use separate controllers to manipulate a 3D object's position and orientation, to avoid the interference of intended and unintended motion control [333]. Some interface hardware, such as the trackpad of hand-held controllers, and the joystick of gamepads, are naturally suitable for the separation of DOFs as they can clearly distinguish the control inputs for different motion directions based on the controlled axes. For screen- or projection-based interfaces, interactive avatars such as the ring-and-arrow markers [334], the virtual handlebar [335] enable the independent control of individual DOF(s) of the manipulated (virtual) object or robot end-effector. The motion scaling ratio of the interface affects both the control efficiency and accuracy of the tele-manipulation tasks. Scaling up the control motion can increase robot motion speed and range but may compromise the accuracy of motion control. On the other hand, scaling down the control motion will increase the motion control accuracy in the concerned small-scale workspace, but may also improve efficiency by reducing operational errors. The scaling ratios can be fixed (commonly used by tele-robotic laparoscopic or eye surgery interfaces [336]), or vary with the user's operating speed (e.g., PRISM method [337]) or regions of operation [332]. Both fixed and varying scaling have pros and cons. Interfaces with varying scaling ratios can better adapt to the control of fast and slow motions in a large or small workspace. In contrast, interfaces with fixed scaling ratios tend to be more stable and predictable to the teleoperator. The major concern in the design of interface scaling ratio is how to achieve the trade-off of control *efficiency*, *adaptability*, and *predictability*. Related work in the literature proposed several solutions to achieve such trade-off, which allows the user to: a) manually switch among several pre-defined scaling ratios (suggested by task experts) depending on the types of operation or size of the workspace [336]; b) manually adjust the scaling ratio as a continuous control parameter (e.g., by changing the distance between the controlling hands [335, 338]); c) autonomously adjust the scaling ratio (e.g., according to the location of operation in the workspace [339]). While these rep-

resentative designs balance the performance objectives to some extent, it was also revealed that: 1) Being able to adjust the scaling ratio (manually or autonomously) is useful overall but may lead to a complicated interface design hard for users to learn; 2) The manual switching of scaling ratio modes leads to more predictable interface behavior, but may increase the control efforts and mental workload; 3) Autonomously adjusting the scaling ratio reduces the control efforts and cognitive workload, but has to be carefully designed to the nature of the task and the preference of users.

Table 4.1: Conventional and contemporary control interfaces for assisted tele-manipulation.

| **Type** | **Representative Interfaces** |
|---|---|
| | **Conventional** |
| Desktop | Keyboard + mouse + point-and-click graphical user interface [22, 333, 340, 341] |
| Hand-Held | Xbox gamepad [342, 343, 344], (Haptic) Joystick [198, 283, 333, 345, 346, 347], Customized teleoperation console (e.g., Da Vinci Surgeon Console [348, 349]) |
| | **Contemporary** |
| Wearable | Arm/hand exoskeleton [350], Data glove [351, 352], Soft haptic glove [353, 354] |
| Hand-Held | Hand-held controllers of virtual reality systems (e.g., HTC Vive, Oculus) with trackpad and buttons, Robopuppet [355], Chopstick [356], Haptic tweezer [357], Tangible interface [358] |
| Motion / Gesture | **Touchless**: Arm/hand motion tracking (e.g., vision-based [67, 333, 339, 359], marker-based [18, 360], and IMU-based [122]), Mid-air gesture control [361, 362];<br>**Touch-based**: Touch-screen gesture control [363, 364, 365] |

Human operators need to have visual information regarding the global, spatial relationship in the workspace, and the detailed, local visual information of region-of-interest critical to the precise operation. Realistic tele-manipulator robots also need to have multiple cameras dedicated to providing global and local viewpoints, or a single camera with sufficient mobility and motion of range to serve both purposes. Related literature and our prior work proposed to present both (or switch between) the global viewpoint from an onboard or standalone workspace camera, and the local viewpoint from a high-mobility, eye-in-hand camera to focus on the region of interest (e.g., using detail-in-context display [192, 366, 367]). However, for gross manipulation, humans may need to

reach beyond the workspace covered by the workspace cameras. Providing a much-too-large camera FOV is not efficient for limited communication bandwidth and may compromise the resolution in visual display for the workspace of frequent manipulation operations. For precise manipulation, an additional camera viewpoint from a different perspective may be necessary to confirm if the manipulation motion satisfies the task constraints in multiple degrees of freedom. Another problem we have to address is the dilemma to retain human authority and freedom to control the camera while reducing the human effort for camera control. The autonomy for dynamic camera viewpoint control and optimization can reduce human camera control efforts, yet it may move the camera in an unexpected and unpredictable way and disrupts the operator's manipulation.

**Physical and Cognitive Workload in Remote Manipulation —** Maintaining human freedom in motion control is essential to the freeform teleoperation of unstructured, unpredictable manipulation tasks (e.g., including tele-robotic surgery [368], nursing assistance [1], manufacturing [369], hazardous material handling [370], explosive ordnance disposal [371]). Such tele-manipulation tasks are usually not feasible or error-prone for high-level robot autonomy (refer to the review on the level of autonomy [63] ), and heavily depend on human knowledge, expertise, and robot control dexterity. To assist freeform tele-manipulation, it is preferred to enable humans to *efficiently* and *intuitively* control the remote robot and cameras, while having some low-level robot autonomy for perception or action support to reduce human operator's *cognitive and physical workload*. Consider the various motion mapping interfaces (e.g., soft/hard exoskeletons [372], camera/IMU-based motion capture systems [18]) that can map the human body, arm and hand motions to efficiently and intuitively control the freeform motions of manipulator robots. These interfaces tend to cause non-trivial cognitive and physical workload [18], because precise control of manipulation motion or posture could be difficult without the necessary haptic or proprioceptive feedback [194, 373]. The remote visual perception problems, including the limited field of view, loss of depth information, and unnatural camera viewpoint control (typically for eye-in-hand cameras), also contribute to the

cognitive and physical workload. This cognitive and physical workload not only fatigues the teleoperator if they use the interfaces for hours but may also lead to work-related musculoskeletal injury for the operator who uses the teleoperation interfaces on a daily basis. For the tele-manipulation tasks that are designed for manipulation dexterity rather than handling heavy payload, the "assist-as-need" action augmentation helps humans to efficiently and reliably complete the manipulation actions clearly indicated by humans (e.g., by moving the end-effector close enough to the target location or object). This will be more effective than the autonomy predicting human intents (e.g. [28, 346, 374]) to assist as early and as much as possible. To effectively reduce the cognitive workload, perception augmentation can present additional visual information using the camera viewpoint from a different perspective, or present AR visual cues to communicate the high-level task and robot states so that human operators do not need to perceive and comprehend the low-level feedback from various sensors [375].

**Transparency in Human-Robot Collaboration —** In the field of telemanipulation and human-robot interaction, researchers have proposed various designs for robot autonomy that aim to address the challenging aspects of dexterous manipulation [376] or take over repetitive tasks [377], to improve efficiency, accuracy, safety, and ease effort from human operators. Such robot autonomy ranges from simple action support to full automation (see the categorization in [63]) and is able to adapt to a wide range of applications (e.g., healthcare [378], exploration [379], and manufacturing [377]). The design of various levels of robot autonomy may leverage the capabilities of both humans and robots to better sense the workspace, make action decisions and execute the planned motion. In such human-robot collaboration, effective communication is crucial, as a lack of transparency can lead to confusion, resulting in reduced control efficiency and increased operational workload. Researchers have proposed varying sensory feedback to promote transparency through auditory cues [330], haptic feedback [380], and AR visual displays [380], which can provide real-time information to the human operator. Unlike the limited channels in auditory and haptic

feedback, the AR visual cues can convey rich and detailed information, such as the workspace environment [381] and the robot's state, actions, and intentions [382, 383]. We have built our AR visual cues by leveraging the existing research and solutions that have addressed similar challenges. Our approach incorporates three distinct representations to indicate the activation of autonomy and two different types of information to illustrate the intent of autonomy. In addition, some situations may require more human input whereas robot autonomy is impractical. Therefore, we extend our AR visual cues to assist both large-scale movements or fine-tuned adjustments, while also allowing for obstacle avoidance. Our insight from related work in literature is that telemanipulation with various levels of robot autonomy can be advanced by introducing suitable AR visual cues to better inform the autonomy intentions and aid human control when necessary. Conducting a systematic investigation to identify the deciding factors for each type of AR feature at varying levels of robot autonomy is crucial to enhance human-robot collaboration.

## 4.3 Haptic and Augmented Reality Visual Cues

Robot teleoperation benefits from motion capture interfaces, which map human motions to the teleoperated robots for freeform control. While being intuitive and efficient, such teleoperation interfaces are limited when required to control precise tele-manipulation tasks. Specifically, the precise control of the position and orientation of the robot end-effector and the manipulated objects usually lead to a significant cognitive and physical workload and may exhaust and frustrate the users' novice to robot teleoperation. To increase the control precision, haptic and augmented reality (AR) visual cues can be introduced to communicate information critical to task performance, such as direction and distance to target, contact with objects, and environmental constraints. Although related work has proposed, evaluated, and compared various designs of haptic and AR visual cues, as well as their integration, there is still a lack of design theory about what (types) of information are

preferred to be communicated in which modality of sensory feedback. Moreover, it is also unclear how this preference will change in the presence of secondary tasks that demand additional cognitive workload (in terms of visual attention, haptic/proprioceptive perception, or critical thinking). Such design theory will contribute to the fundamental science of the design of multi-modal sensory feedback, and enable the design of haptic and augmented reality visual feedback to be generalized across robots, tasks, and interfaces. We first focus on the design of haptic and AR visual cues for assisting the dexterous freeform tele-manipulation tasks, performed in the context of tele-nursing assistance.

### 4.3.1 Teleoperation Interface and Assistance

**Motion Tracking Interface via HTC Handheld Controller —** Shown in Figure 4.1, we used the hand pose tracked by the handheld controller of the HTC Vive virtual reality system to control the end-effector of a Kinova Gen 3 robotic manipulator. The hand-held controller provides programmable vibrotactile feedback, while AR visual cues augment the video stream of the remote workspace camera.


Figure 4.1: Tele-nursing assistance tasks may involve freeform, and dexterous manipulation (e.g., inserting a straw into the beverage container). Haptic and AR visual cues can be leveraged to communicate task-critical information.

**Haptic and AR Visual Cues —** We proposed the four types of tele-manipulation assistance based on the findings from the human factor experiments [367, 384] presented in Chapter 3, which investigated the visual-haptic sensory integration in the usage of active telepresence. We observed how the participants perform a general-purpose manipulation task that integrates reaching, moving, grasping, and stacking actions, after adapting to the new visual and haptic feedback. We found out that: 1) after some practice, participants can seamlessly integrate the usage of active telepresence cameras with the actions to perceive haptic feedback; 2) the typical actions to leverage the limited haptic feedback include touching-to-locate objects, moving the objects on the table (instead of above the table) to take advantage of the table constraints; 3) when performing the manipulation tasks, participants leverage active telepresence cameras and haptic feedback to locate the target, detect environment constraints, examine possible grasp, and confirm contact with an object.

These observations inspired the design of haptic and AR visual cues for sensory augmentation (as shown in Figure 4.2), which include:

- **Target Locator** indicates the robot end-effector's movement direction and distance to the goal (i.e., the targeted object). As the robot end-effector approaches the targeted object, the haptic cue decreases the strength of its continuous vibration, while the AR visual cue uses a green vector to visualize the ideal direction of motion of the robot end-effector (the blue dot) and its distance to the target (the distance in centimeters at the bottom left of the screen). The marker on the robot end-effector will turn red and increase the diameter to alert the user that the robot is out of view and provide the exact direction to support the self-correction.

- **Constraint Alert** alerts the teleoperator if the end-effector is about to violate any environment constraints (e.g., hitting the table), in which case the haptic cue vibrates the hand-held controller, while the AR visual cue monitors the robot end-effector's height from the table and turns from green to red if too close to the robot. The height indicator bar also fills up

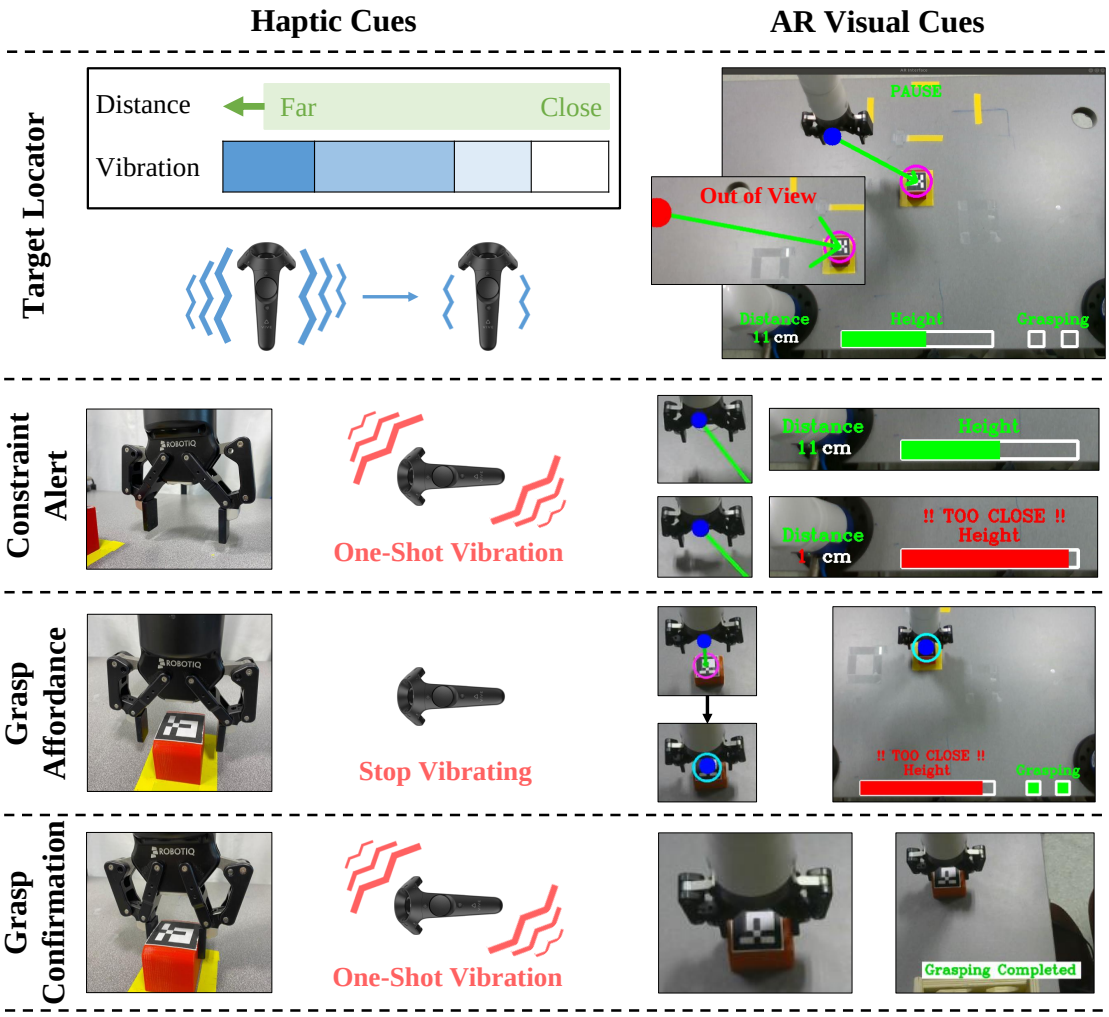



Figure 4.2: Design of equivalent haptic and AR visual cues, for the target locator, constraint alert, action affordance, and grasp confirmation.

with increased proximity to the table.

- **Grasp Affordance** indicates if the end-effector is posed to be ready to grasp the target object, which can be generalized to indicate if the robot end-effector is ready to afford the action to be performed. When the robot end-effector is ready to grasp, the hand-held controller will stop to vibrate, while the AR visual cue will highlight two boxes if the end-effector is within the target region and the height is low enough.

- **Grasp Confirmation** indicates that the robot end-effector has successfully grasped the target

object, using a one-shot vibration of the hand-held controller, while the AR visual cue will display "Grasp completed" and hide all the other AR visual cues. This indicator can be generalized to confirm any completed action.

### 4.3.2 Experiment

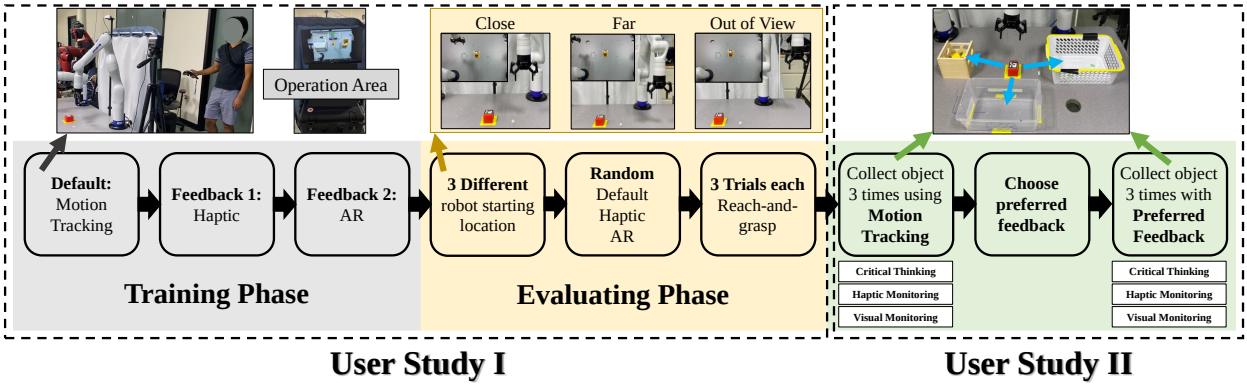



Figure 4.3: Experiment procedure.

**Participants —** We conducted two user studies (with the same 8 participants, 5 males, 3 females, age = 30.5±3) to investigate: 1) How effective these haptic and AR visual cues can support tele-manipulation? 2) Which modality of sensory feedback do people prefer to use to communicate what types of information for tele-manipulation support? 3) How this preference will be influenced by the different types of workload introduced by the secondary tasks performed along with the robot teleoperation?

**Experimental Procedure and Conditions —** Figure 4.3 shows the experimental procedures for User Study I and II. The User Study I includes a *training phase* and an *evaluating phase*. During the training phase, participants first learned the baseline interface (i.e., tele-manipulation via hand motion tracking), the haptic and AR visual feedback augmented interfaces from the verbal instructions and demonstrations of an experimenter and then practiced for five minutes each. During the following *evaluation phase*, participants performed three sections of a reaching-and-grasp task, in

which they controlled the remote robot manipulator to grab a small wood block in the workspace (see Figure 4.1). We randomized the order of the sections using different modes of sensory feedback, namely no feedback, haptic, and AR visual feedback. In each section, the robot end-effector was set to be at three different starting locations, which were close to the target, far away from the target, and out of the camera view, respectively. The participants manipulated the object for three repetitions for each robot's starting location. In User Study I, the participant performed a total of 27 trials (3 modes of sensory feedback $\times$ 3 robot starting locations $\times$ 3 repetitions).

In User Study II, participants teleoperated with the robot to perform the primary task of picking-and-place an object into three different bins (shown in Figure 4.3), for two sessions. The participants performed the task first in Session 1 using the baseline interface (hand motion tracking without haptic or AR visual feedback) and then performed the task again in Session 2 using their preferred form of sensory feedback for the four types of tele-manipulation assistance. In both Session 1 and 2, we introduced three types of secondary tasks to robot tele-manipulation in a randomized order, to understand how they may change the performance, workload, and preference of sensory feedback. As shown in Figure 4.4 (top), the *haptic monitoring* task requires the participants to press the button of the controller held in their non-dominant hand within 1 second if they detect any vibration from it. The one-shot vibration occurs randomly every 7 to 13 seconds during the tele-manipulation task. The *visual monitoring* task (bottom in Figure 4.4) requires the participants to monitor the simulated vital sign profile displayed in the second user interface. About every 5 seconds, the displayed vital sign goes beyond its normal range (60 to 100) and turns red, and the participants have 1 second to press the button of the controller held in their dominant hand (same controller for robot operation) to record this event. Moreover, we designed a *critical thinking* task which requires the participants to verbally respond to simple math problems (only includes addition and subtraction of one-digit numbers) continuously while performing the tele-manipulation task. These secondary tasks were designed to simulate the additional cognitive workload that healthcare

workers may experience when performing patient care tasks along with nursing robot teleoperation, including the haptic perception for standing by the emergency on-call, the visual attention for monitoring patient status, and the critical thinking for patient evaluation.

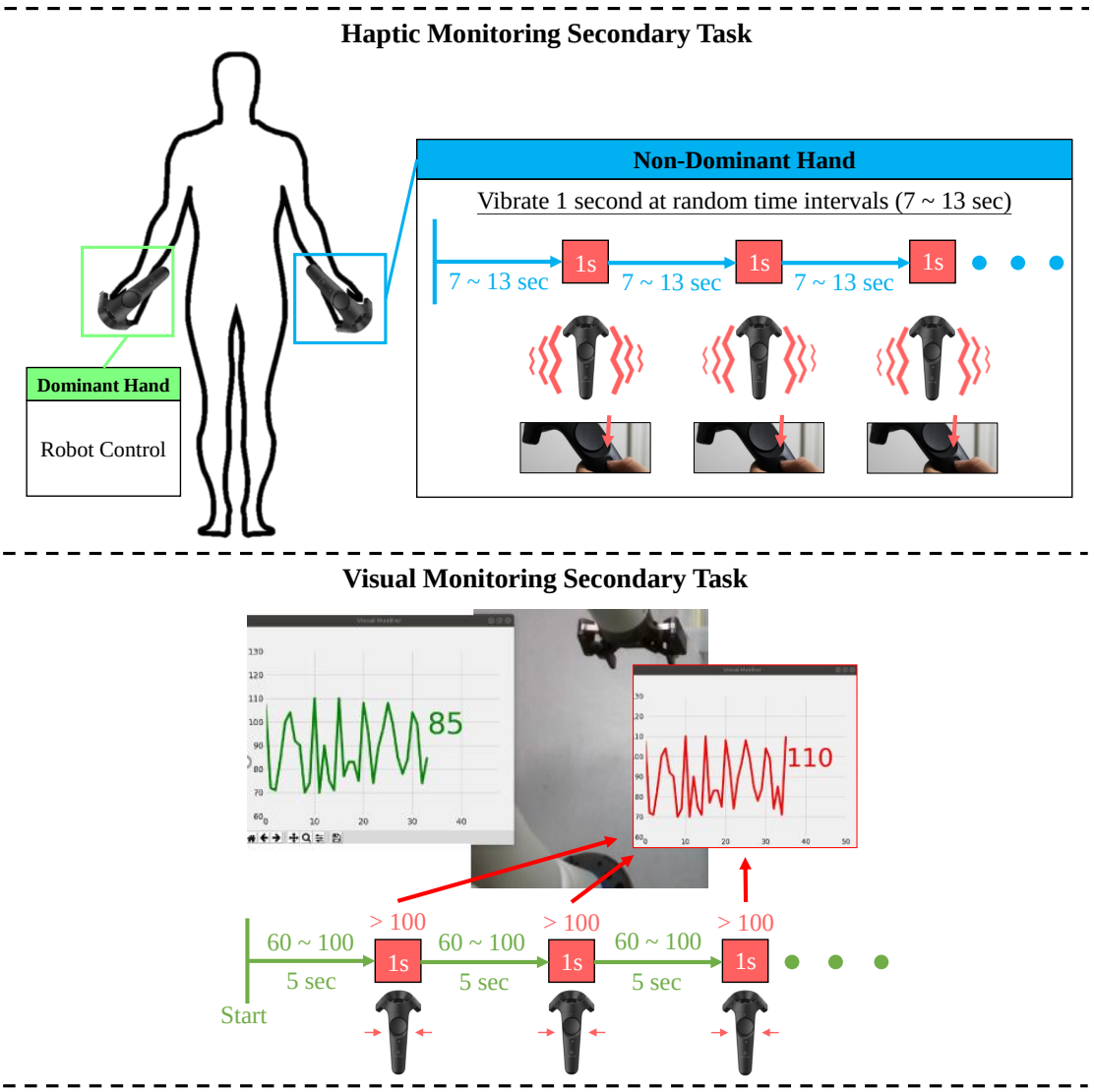



Figure 4.4: Secondary tasks that introduce additional cognitive workload for haptic monitoring (top) and visual monitoring (bottom).

### 4.3.3 Comparison of Sensory Feedback

Figure 4.5 to 4.7 compares tele-manipulation tasks performed with and without haptic or AR visual cues. Figure 4.5 shows the mean and variance of the task completion time and trajectory length across all the trials and participants, for different starting locations of the robot end-effector. To some extent, the AR and haptic cues reduced the task completion time. Based on the one-way analysis of variance (ANOVA), the haptic cues and AR visual cues significantly reduce the task completion time for far-away ($F(1,46)=5.1$, $p<0.05$) and out-of-view starting point ($F(1,46)=6.68$, $p<0.05$), respectively. ANOVA analysis showed a significant difference in trajectory length between baseline and with haptic feedback ($F(1,46)=42.99$, $p<0.05$) for close starting location. While the AR visual cues had significantly shorter trajectory length than baseline and haptic feedback for close ($F(1,46)=63.49$, $p<0.05$; $F(1,46)=17.63$, $p<0.05$), far-away ($F(1,46)=4.87$, $p<0.05$; $F(1,46)=3.39$, $p<0.05$) and out-of-view ($F(1,46)=5.2$, $p<0.05$; $F(1,46)=3.45$, $p<0.05$) starting locations.

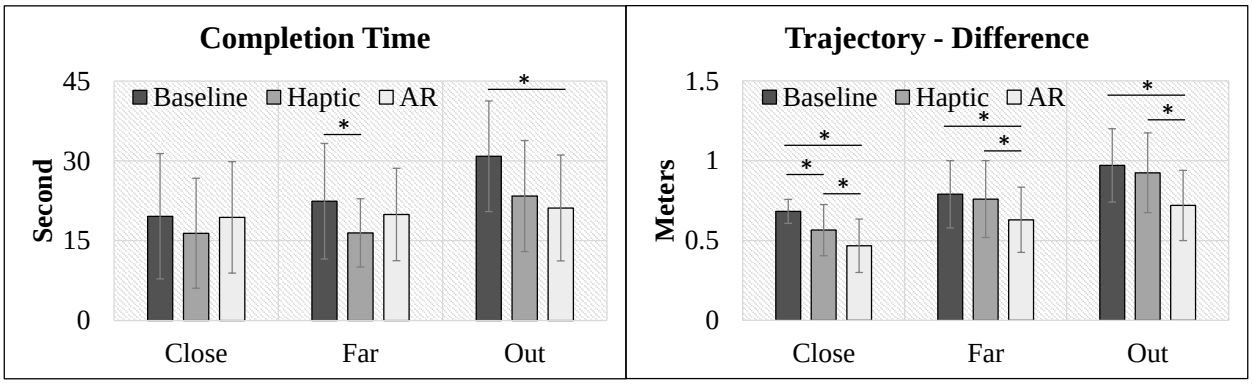



Figure 4.5: Task completion time and trajectory.

In Figure 4.6 we compared the total occurrence of errors from all the participants and trials because the mean and variance of the error across participants and trials were consistently small. For the table collision, the mean and standard deviation of occurrences for baseline, haptic and AR visual cues are (close: $0.75\pm0.43$, far-away: $0.88\pm0.44$, out-of-view: $0.67\pm0.47$),

(close: 0.13±0.33, far-away: 0.08±0.27, out-of-view: 0.17±0.37) and (close: 0.42±0.49, far-away: 0.54±0.49, out-of-view: 0.46±0.49). While for the hitting target object, the mean and standard deviation of occurrences for baseline, haptic and AR visual cues are (close: 0.33±0.47, far-away: 0.42±0.49, out-of-view: 0.33±0.47), (close: 0.2±0.41, far-away: 0.38±0.48, out-of-view: 0.29±0.45) and (close: 0.04±0.19, far-away: 0.08±0.38, out-of-view: 0.04±0.2). Across all the starting points, the haptic cues consistently performed better than AR visual cues in avoiding the collision with the table, while the AR visual cues can better prevent hitting the targeted object compared to the haptic cues. This implies that AR visual cues may be more suitable to avoid errors in the operation of the target objects, which the teleoperator needs to continuously track with visual attention, while the haptic cues are preferred to alert the teleoperator of the environmental constraints as needed to avoid visual distraction.

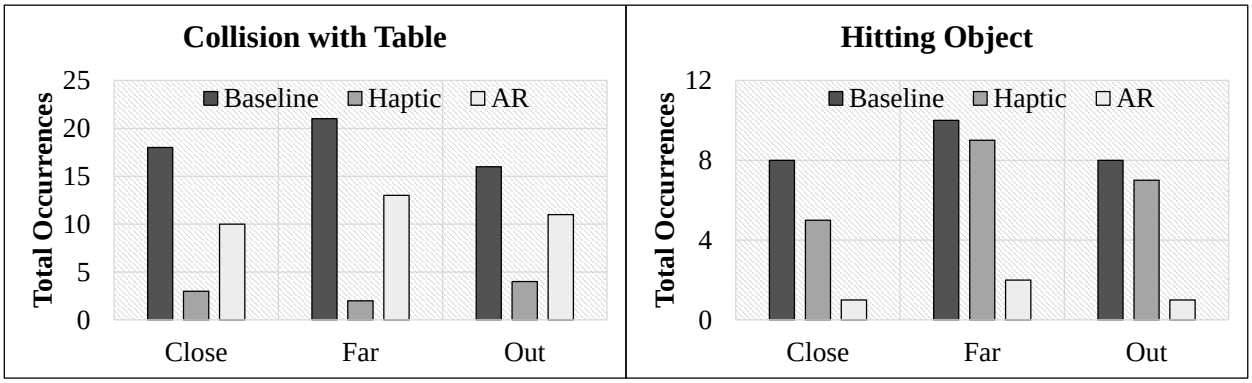



Figure 4.6: Occurrences of collision with table and hitting object.

We use the weighted NASA-TLX scores to measure the subjective workload across baseline, haptic, and AR visual cues. The weighting coefficients were selected as follows: mental demand=5, physical demand=1, temporal demand=0, performance=4, effort=3, frustration=2. Based on the Wilcoxon rank sum test, the feedback from the NASA-TLX and the SUS survey (Figure 4.7) further indicated that both the haptic and AR visual cues had significantly lower workload ($p<0.05$) and higher usability scores ($p<0.05$) than the baseline condition without any sensory augmentation cues. We also noticed a consistency in the preference of sensory feedback for different types of

tele-manipulation assistance: 6 out of 8 participants prefer to use AR visual cues for the *target locator* and *grasp affordance*, and use haptic cues for the *constraint alert* (7 out of 8 participants) and *grasp confirmation* (6 out of 8 participants).

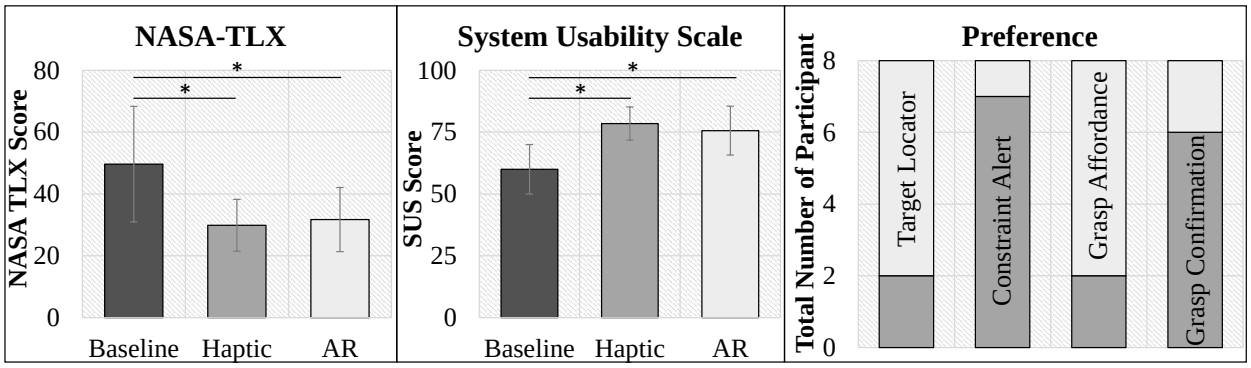



Figure 4.7: Feedback of NASA-TLX, SUS and user preference survey.

### 4.3.4 Adaptation to Secondary Tasks

Figure 4.8 and 4.9 compared the performance of the primary and secondary tasks without the augmented sensory cues, and with the user's preferred sensory cues, for tele-manipulation performed with different types of secondary tasks. Note that the participants can choose different sensory modes for each of the four types of tele-manipulation assistance. Figure 4.8 shows the mean and variance of the task completion time, which indicates significant performance improvement ($p<0.05$) for all the secondary tasks using the preferred sensory modes. Moreover, the preferred sensory mode reduced the total errors from all the participants in all the trials. The reduction for the collision with the table is much more obvious compared to that for hitting the target object.

Figure 4.9 shows the preferred sensory modes also significantly improved ($p<0.05$) the performance of the secondary tasks by 1) reducing the time to respond to the math problems simulating the critical thinking workload, and 2) improving the success rate in the tasks of visual and haptic monitoring.

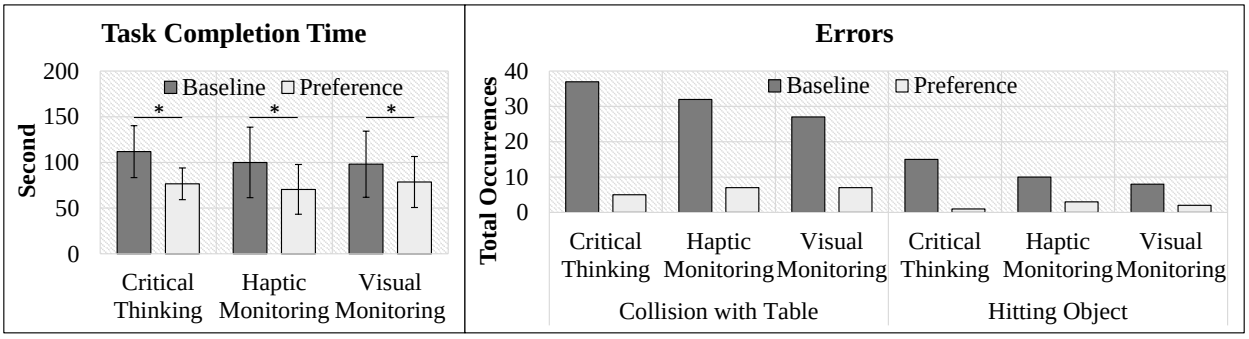



Figure 4.8: Primary task completion time and error occurrences.

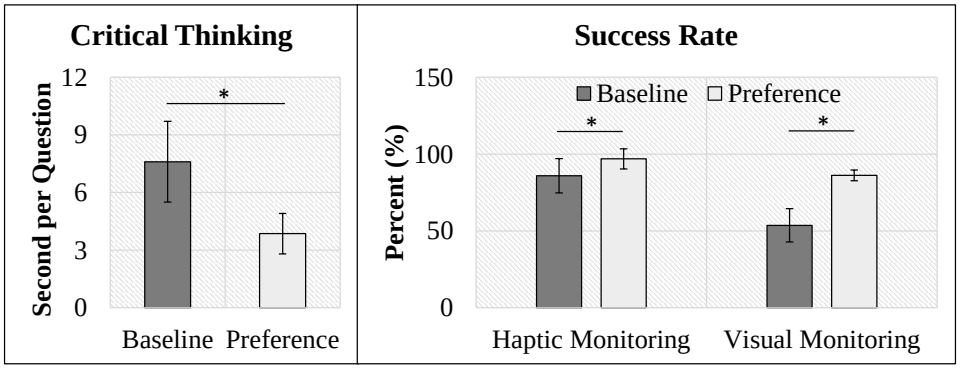



Figure 4.9: Performance of secondary tasks.

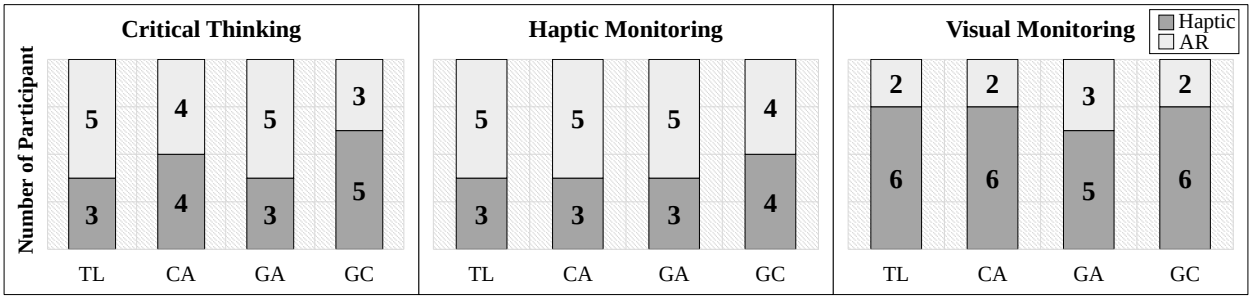



Figure 4.10: Preferred sensory modes for target locator (TL), constraint alert (CA), grasp affordance (GA), and grasp confirmation (GC).

Figure 4.10 shows the number of users that prefer haptic versus AR visual cues for each tele-manipulation assistance when the tele-manipulation was performed with each secondary task. We noticed that for critical thinking and haptic monitoring, the users' preferences fell into two categories with a 5:3 or 4:4 ratio, while for visual monitoring there is a tendency that the users to prefer haptic cues more than the AR visual cues, with 6:2 ratio for three out of four tele-manipulation

assistance. This tendency in the preference may be because the haptic cues do not compete with the primary task for visual attention.

## 4.4 Perception and Action Augmentation

Consider a comprehensive tele-manipulation task such as workspace organization, which may involve control of reaching, grasping, moving, and placing of various kinds of objects. This task is mostly unstructured and requires freeform control because it does not follow (and requires) any procedure; What, when, and how to handle each object will be decided by the user on-the-fly as the task goes on. Sufficient freeform control will also allow humans to improvise based on their knowledge and experience (leveraging environmental constraints, physical properties of handled objects, etc) to facilitate or enable some manipulation. To effectively assist tele-manipulation without compromising the human's control authority and freedom, the robot autonomy can provide an additional camera viewpoint from a different perspective or AR visual cues, to augment their remote perception and enable them to control the manipulation themselves. The robot autonomy can also be different interface mappings or autonomous actions that can effectively perform the task. As the human has moved the robot end-effector close enough to the target object or container to place, it is easier for the robot to infer the human's intent and assist in the structured component (placing object) of the unstructured task using simple but effective autonomy.

To this end, we propose systematic approaches for action and perception augmentation. The *Action Augmentation* allows humans to control the robot motions using hand pose tracking and a trackpad available on the hand-held controller, for freeform or constrained motion control. It is also implemented by dynamically adjusting the scaling of the operator to robot interface mapping to support both gross and precise manipulation. For *Perception Augmentation*, we provide AR visual cues to convey the visual information difficult to perceive in 2D cameras (e.g., the task

and robot status, interface control mode, and autonomous action affordance). We also provide a complementary camera viewpoint from a significantly different perspective, in which missing visual information, like loss of depth perception, can be more easily perceived.

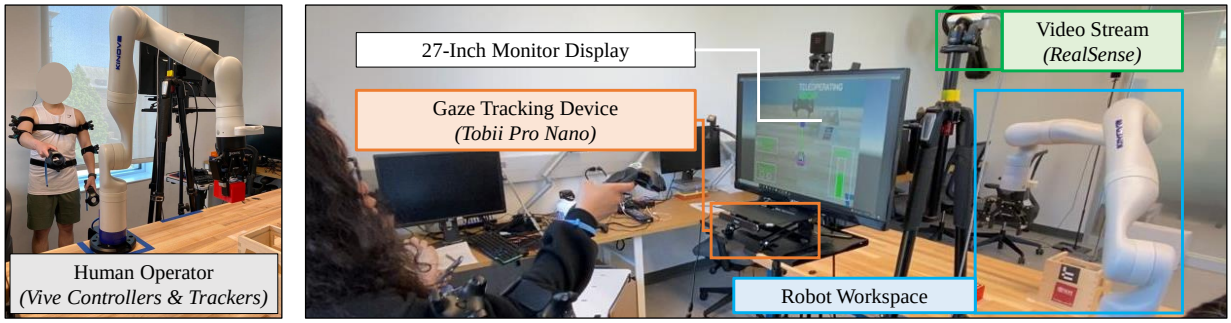



Figure 4.11: Overview of the tele-manipulation system.

We have implemented the proposed perception and action augmentation on a representative interface. As shown in Figure 4.11, we used the HTC Vive hand-held controller to control robot motion and used the desktop monitor to display the remote camera viewpoints and AR visual cues. The robot could provide autonomous actions (e.g., grasping and placing actions) or switch to constrained motion control using a trackpad when humans operated the robot end-effector near the target object or location to place. The implementation can be generalized to the teleoperation system using various contemporary tele-manipulation control devices (e.g., hand-held, touch-based, wearable), display (e.g., screen-based and head-mounted visual display). Note that we only provided the autonomous actions and motion constraints needed for the pick-and-place tasks, because this chapter does not focus on how to predict human intents, or how to enable autonomy.

### 4.4.1 Interface and Evaluation Design

Here we present our proposed approaches for perception and action augmentation, and their implementation on a representative tele-manipulation system and for a general-purpose pick-and-place

task. We further propose methods for the estimation of eye-tracking-based cognitive workload, and motion-tracking-based physical workload to enable the evaluation of integrated interfaces in the formal user study. We conducted three pilot user studies to (1) provide the reasoning for complementary viewpoint design; (2) determine the effective integration of the perception and action augmentation; (3) validate the physical workload estimation with sEMG data.

**System Overview —** Figure 4.11 shows the tele-manipulation system [380], which enables the development of the perception and action augmentation proposed for this project. The **robot platform** is a 7-DoF Kinova Gen 3 manipulator with a two-fingered Robotiq gripper that can detect contact with the grasped object. Two RealSense Cameras (D435) were standing alone in the workspace for primary and complementary remote perception.

For **robot motion control**, we use an HTC Vive hand-held controller (referred to as "controller" in the rest of this paper) that allows human operators to control the freeform robot motions using their natural hand motions and constrained motion using the controller's trackpad. By default (i.e., Mode 1 of Figure 4.13), the linear velocity of the controller will be mapped to the linear velocity of the robot. The input-to-output motion mapping ratio is 1:5 along the x-axis and 1:3 along the y- and z-axis. We locked the robot's rotational motions because this chapter focuses on developing and comparing different *modalities* of teleoperation assistance instead of the capabilities of robot control. To perform a tele-manipulation task, the operator will: 1) press the menu button on the controller to send the robot to the home configuration, 2) press the menu button again to get the robot ready, and press the grip (side) button to initiate the control.

For **visual feedback**, we used 1440 $\times$ 1080 pixel resolution Unity 3D window on a desktop monitor to display the remote camera video stream (at 30 Hz frame rate) and to display augmented reality visual cues. By default, the graphical user interface (GUI) will only display the robot's operation state. Specifically, the GUI will display: 1) "*WAITING*" when the tele-robotic system is ready for operation and waiting for a control command; 2) "*SENDING HOME*" when the operator

presses a controller button to set the robot to the default pose; 3) "*READY*" when the robot is posed at the start position for the current task; 4) "*TELEOPERATING*" when the robot is being teleoperated; 5) "*PAUSED*" when the robot is paused by the teleoperator. Figure 4.12 shows the control architecture and data communication pipeline of the tele-manipulation system. A screen-based eye tracker (Tobii Pro Nano) was attached below the monitor to track the human operator's gaze and eye movements (e.g., pupil diameter) at 60 Hz. The autonomy for perception and action can detect the ArUco tags attached to the objects, container, and counter workspace [385, 386], to estimate the information for the AR visual cues and control the robot's autonomous actions for precise manipulation (e.g., object grasping and placing).

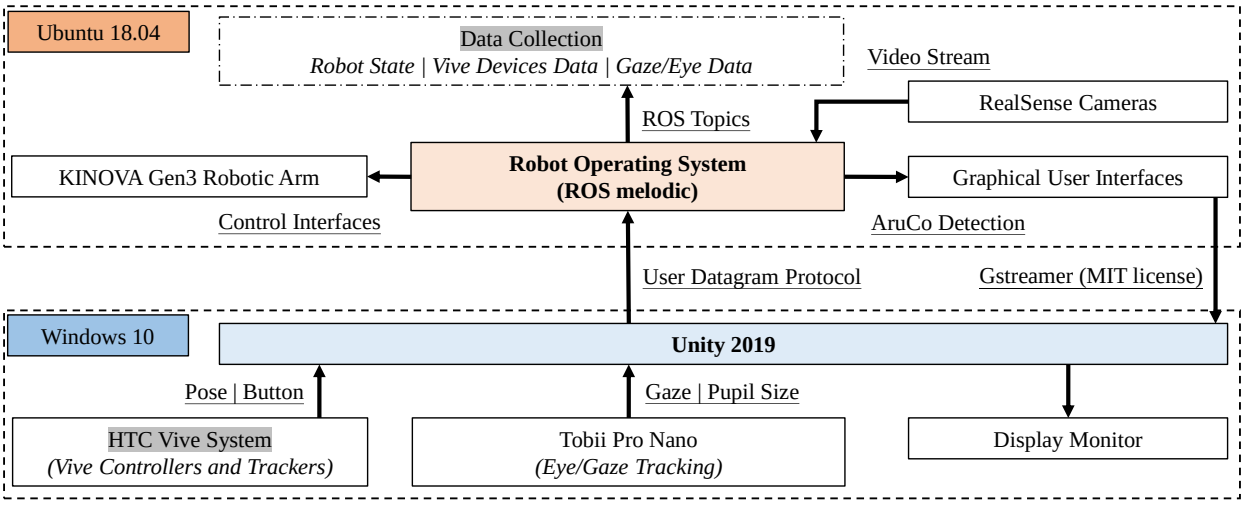



Figure 4.12: System architecture.

**AR Visual Cues and Assistive Autonomy —** To assist robot remote manipulation, we implemented systematic AR visual cues and user-triggered autonomous actions as the baseline representing the common solution for remote perception and action problems. We then develop, integrate and compare different types of perception and action assistance upon the baseline AR visual support and assistive autonomy, to discover new knowledge on optimal human-robot collaboration for freeform tele-manipulation.

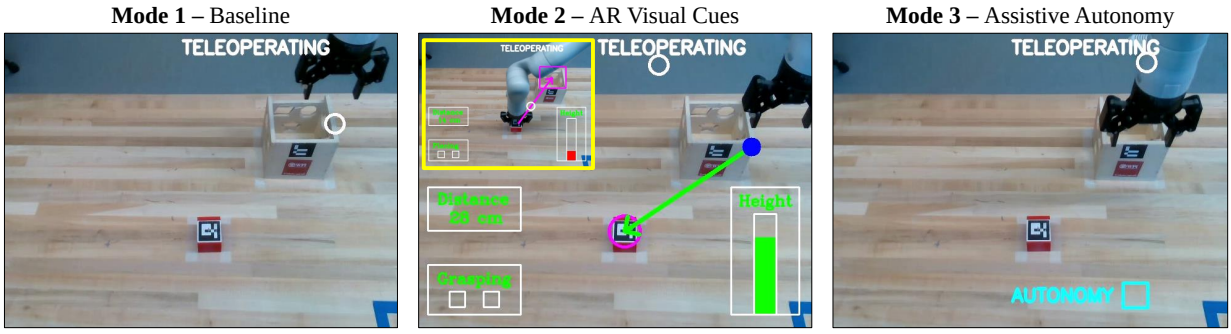



Figure 4.13: Visual interfaces for baseline, AR visual cues, and assistive autonomy.

For **AR Visual Cues**, we have presented four types of AR visual cues for freeform teleoperation assistance, including: 1) the ***Target Locator*** to indicate the robot's movement direction and distance to the targeted object or goal pose; 2) the ***Action Affordance*** to indicate if the robot is ready to afford the action to be performed (e.g., grasping or stacking an object, with a good chance of success); 3) the ***Action Confirmation*** to indicate that the robot has successfully performed an appropriate action; and 4) the ***Collision Alert*** to alert the teleoperator if the end-effector is about to violate any environment constraints (e.g., hitting the table). Figure 4.13 (Mode 2) shows the implementation of these AR visual cues to assist a pick-and-place task.

- The ***Height*** indicator shows the robot's distance to the table surface. Besides the display of numerical distance, the height bar display also turns from green to red if the robot is too close (within 0.1 m ) to the table.
- The ***Alignment*** indicator (displayed as a dot-in-circle) shows if the robot is aligned with the object to grasp or the container to place the object, in x- and y-direction. Once the blue dot moving with the robot is aligned with the pick circle displayed on the object or containers, the pink circle will change its color to light blue to indicate the operator can reliably close or open the gripper to reliably grasp or drop the object into the container.
- The ***Grasping/Placing Hint*** includes two square-shape that turn on and off to show whether

the robot is aligned with the object or the container so that the operator can confidently close or open the gripper. It is designed to confirm the critical information conveyed in the ***Alignment*** and ***Height*** cues.

- The ***Arrow with Distance*** indicator shows the distance (in cm) and direction (using the green and pink arrows, respectively) to show the target object to grasp or container to place.

The proposed implementation of AR visual cues is refined based on the prior design and evaluation results [380]. Specifically, we have: adjusted the ***Height*** indicator to be vertical instead of horizontal for a more intuitive visual display. We grouped the ***Grasping/Placing Hint*** into a white box and highlighted the boundary of the container to make them easier to spot at a glance. We also extended the AR support to picking-and-placing as well.

For **Assistive Autonomy**, we also provide autonomous actions to assist the operators to perform precise manipulation (e.g., to pick and place an object) as shown in Figure 4.13 (Mode 3). The robot autonomy can detect the human's goal and action intents based on robot states, including the distance to the object or container, and whether the gripper is open or closed. When the gripper is open, we predict human intent to grasp the object if the robot is within the predefined distances to the center of the object (0.05 m, 0.08 m, and 0.13 m in the x-, y- and z-direction). A hint of "AUTONOMY" will be displayed to show the robot has detected the human's goal and action intent by filling the box when the robot can reliably perform the action. Humans therefore can press the controller's trigger to confirm the execution of the autonomous action, after which the robot will autonomously reach to grasp the object and lift it to 0.2 m above the table surface. To place an object, the operator needs to move the robot to be within a predefined distance (0.08 m, 0.08 m, and 0.15 m in the x-, y- and z-direction) to the center of the top of the container. Once confirmed by the human, the robot will autonomously move to the top of the container and drop the object into it reliably. Our proposed visual and action augmentation depends on the robot

autonomy to predict human intents, determine action affordance and success, and detect and avoid the collision. Here we implemented a simple design of autonomy that predicts the human's intent to grasp or place an object based on the robot's state. Object detection, action affordance, and collision are also simplified given that we know the location and geometry of the object and the environmental constraints. Note that more advanced methods to predict human intents, from human control inputs [28, 67, 387], gaze [388, 389], or their fusion [390], can be integrated with our proposed visual and action augmentation for more complex manipulation tasks. Advanced methods to detect objects and their action affordance (e.g., using Sim2real approach [391], for unknown objects [392]) can also be incorporated to enable more complicated precise manipulation and the delicate control of interaction forces. Collision in dynamic and cluttered environments can be detected using an advanced method such as generalized velocity obstacles [393].

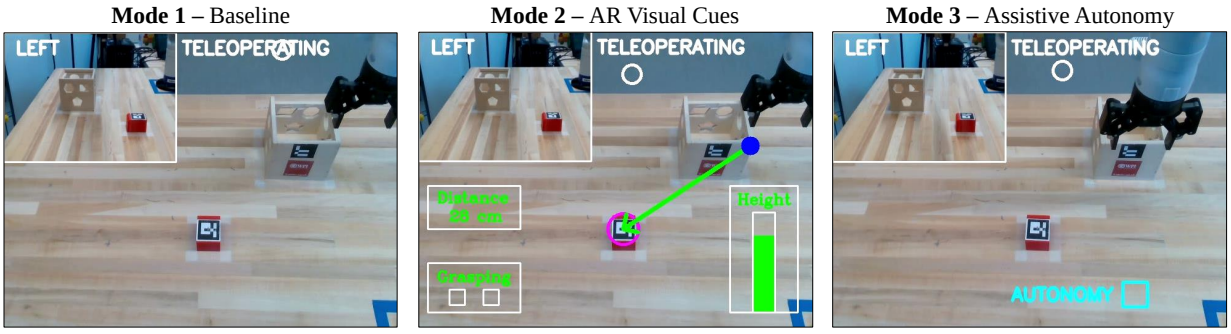



Figure 4.14: Complementary viewpoints in different interface modes.

**Complementary Viewpoint for Perception Augmentation —** We propose to leverage an additional workspace camera to provide a complementary viewpoint in which the operator can better perceive the information missed in the primary workspace camera viewpoint. Shown in Figure 4.14, the GUI presents a picture-in-picture (PIP) display to embed the complementary viewpoint into the primary viewpoint. The perception augmentation in form of the complementary viewpoint can be presented always (i.e., the fixed viewpoint) or dynamically given the robot and task states (i.e., dynamic viewpoint). It can also be augmented with different interface control modes. Here

we present the pilot user study (Pilot Study I) we conducted for iterative design and evaluation.

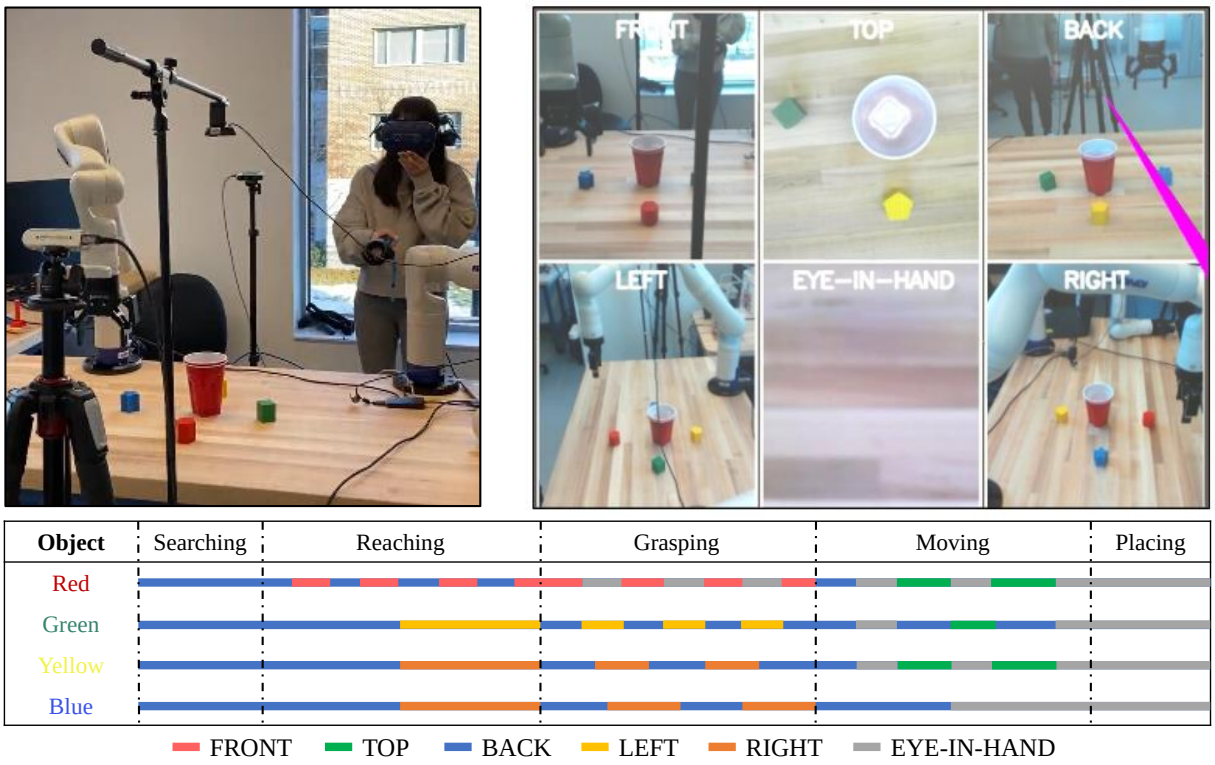



Figure 4.15: Complementary viewpoint filtering.

*Q1 - Do we need multiple viewpoints?* We conducted a pilot study with one expert participant (female, age=33, without visual or motor disability, 100+ hours experience with robot) to determine the preferred viewing angle and distance of the complementary viewpoint. We follow the experiment setup in the literature [394] and set up five workspace cameras (4 RealSense cameras and 1 Webcam) to observe the workspace from different perspectives (the front, back, left, right, and top). Including the viewpoint from the eye-in-hand camera of the robot, we presented six viewpoints to the user and tracked her gaze fixation on each viewpoint during the pick-and-place task. As shown in Figure 4.15, the operator was asked to reach to grasp four blocks of different colors placed around a red cup and place each into the cup. During the task, the participant used an HTC Vive controller to control the robot's motions. For the pilot study, the head-mounted display of the HTC Vive Pro Eye system is used to display graphical user interfaces and track human gaze. In Figure 4.15, the

camera viewpoints that the human looked at are compared between different manipulation actions, and compared between the manipulation of different objects. The operator's gaze fixation mostly switched between the *back view* that looks at the workspace from the operator's standing point and the viewpoint in which she could observe the object to pick up with minimal occlusion. We also found that the participant spent more time looking at the back view than any other viewpoint, which implies that we need to distinguish the primary and complementary viewpoints based on the duration of their fixation.

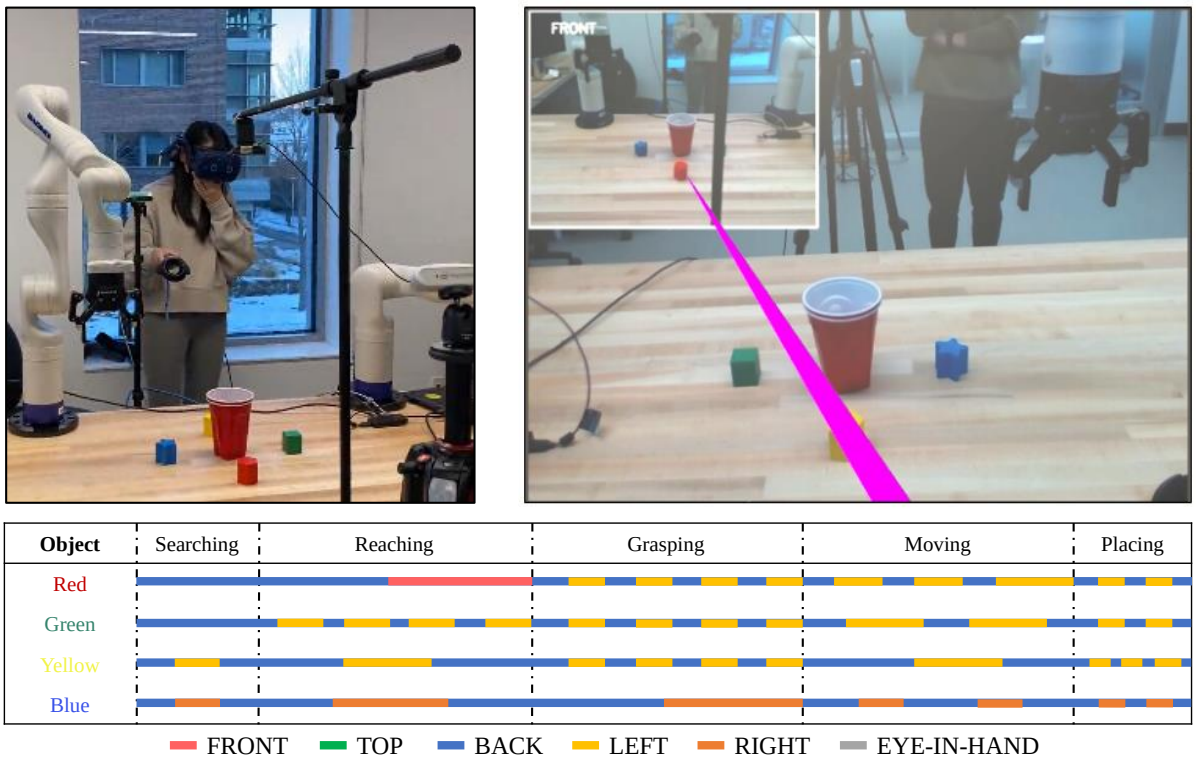



Figure 4.16: Gaze fixation on each camera viewpoint.

*Q2 - Which camera is preferred for the complementary viewpoint?* We conducted another round of the pilot study with the same participant to determine the preferred camera view for a complementary viewpoint. Based on the result from Q1, we implemented a picture-in-picture multi-viewpoint display. By default, we displayed the back view camera to be the primary viewpoint and the front view camera to be the complementary viewpoint. The complementary viewpoint system was se-

lected since the back view was utilized the most with different viewpoints used only when additional information was required. This implied that only one additional viewpoint to the primary viewpoint would be required. As shown in Figure 4.16, the operator could also press the controller's button to switch the complementary viewpoint to be from other workspace cameras. We recorded the robot and task states and tracked the human gaze. We noticed that the operator preferred to only use the left view camera for the complementary viewpoint, because: 1) it shows the additional objects not visible in the back view, and 2) it is less occluded by the robot arm. The participant also mentioned that manually switching the complementary viewpoint increased her cognitive workload and control efforts during a post-study interview. She also mentioned that the complementary view could be improved with a zoom function to provide detailed information about the task and workspace.

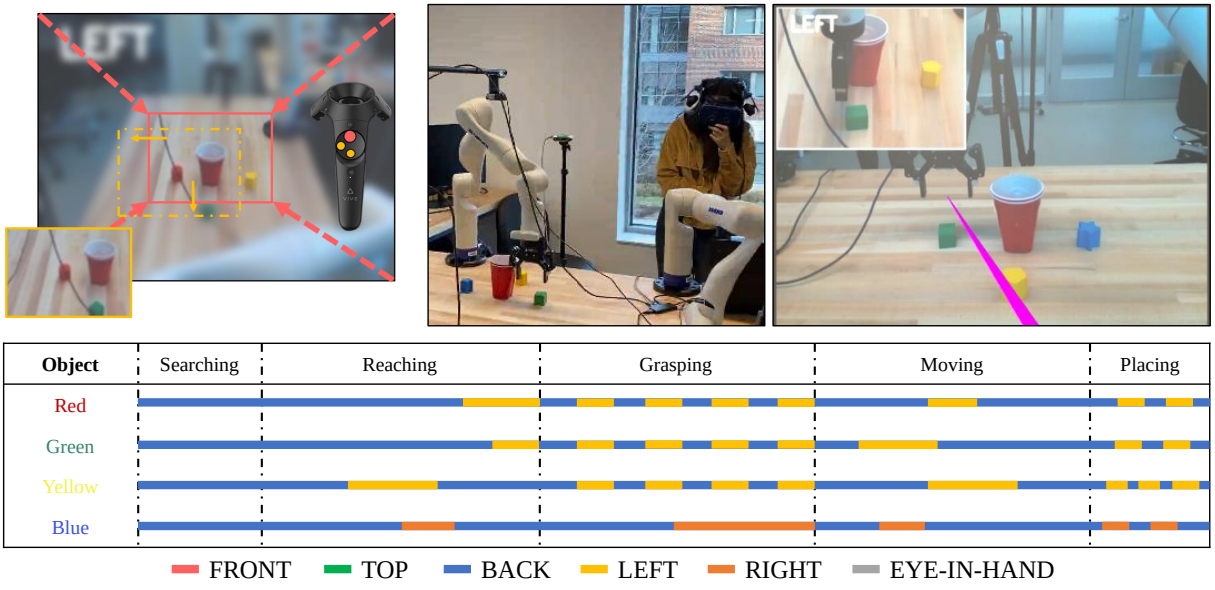



Figure 4.17: Complementary viewpoint with the adaptive field of view.

*Q3 - Do we need to adjust the field of view for the complementary viewpoint?* Based on the feedback from Q2, we enabled operator to use the controller's trackpad to control the complementary viewpoint to shift the center of the field of view (FOV), and to zoom in and out (Figure 4.17). We found that during the same pick-and-place task, the operator still chose the complementary view-

point cameras in a similar way, but preferred to zoom in and shift the FOV to make the target object or container more centered and visible.

**Dynamic Interface Mapping for Action Augmentation —** Dynamic interface mapping, controlled by humans or autonomy, enables humans to use different interface mappings or scaling ratios to effectively control precise and gross manipulation. While manually adjusting the interface mapping could be annoying and tedious, existing autonomy to adjust the interface mapping [338, 395] tend to confuse humans because they do not intuitively inform human about this change, due to which humans may find the interface inconsistent and unpredictable. Our recent work [34] shows that humans can more efficiently control precise manipulation if the interface mapping: 1) allows humans to constrain the motions to be only for the position or orientation control, and 2) autonomously reduce the human-to-robot motion mapping ratio (which will reduce the robot motion speed) close to the objects and environment constraints. We propose to improve efficiency to control precise, directional motions by: 1) allowing humans to use constrained motion control input channels (e.g., the controller's Trackpad) to control the motions of an individual DOF, and 2) reducing the scaling ratio only in the precise motion control direction.

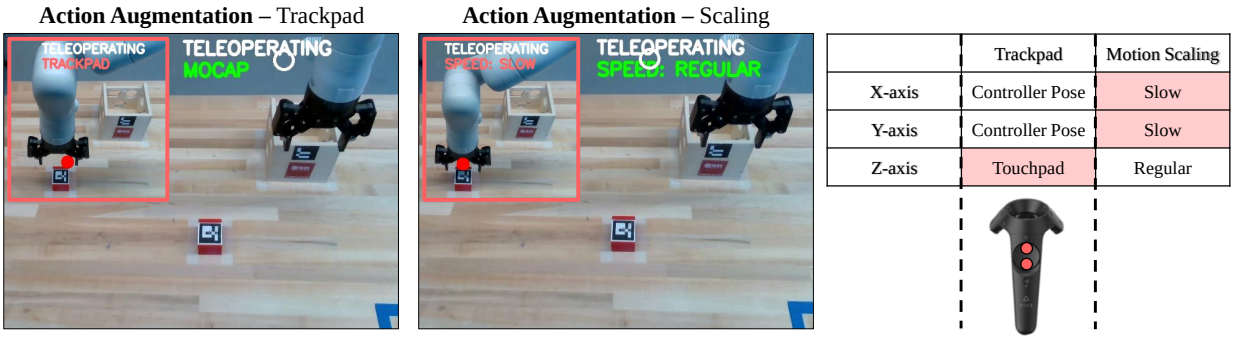



Figure 4.18: Action augmentations using the trackpad and motion scaling.

Figure 4.18 shows the implementation of our two proposed action augmentation approaches. In the "Trackpad" control mode, the human can still control the robot in the X- and Y-axis (horizontal plane motion) using natural hand motions but will control the motion in the Z-axis (vertical motion)

using the controller's trackpad (to avoid collision with the table and object to be grasped). When the human controls the robot to move close to the object or containers, the corresponding AR visual cue will turn from "MOCAP" (in green) to "TRACKPAD" (in red) to inform humans that the interface mapping mode has changed. The trackpad control region for the object (container) is a 0.14 m $\times$ 0.2 m $\times$ 0.6 m (0.2 m $\times$ X 0.26 m $\times$ 0.4 m) bounding box w.r.t. to the center of the object (top of the container). In the "Scaling" mode, the interface will reduce the mapping ratio in the X- and Y-axis to allow humans to precisely adjust the robot to align with the object or container, while maintaining the scaling ratio to be "Regular" in the z-axis. We define the reduced scaling region to be the same as the trackpad control region. The corresponding AR visual cue will turn from "SPEED: REGULAR" in green to "SPEED: SLOW" in red to inform the change in the scaling ratio.

**Integration of Perception and Action Augmentation —** We conducted a pilot user study (Pilot Study II) to understand human's preferred combination of perception and action augmentation. In total, we have 15 different experiment conditions, considering the three interface control modes with different perception and action augmentation. Our pilot study involved 8 participants (4 male and 4 female, 5 novices, and 3 participants who have used the same teleoperation system before). The participants performed a single-object pick-and-place task once under every experimental condition (the order of interfaces was randomized) and reported their preferred combinations of control modes and perception/action augmentation after the experiments.

Table 4.2: Testing conditions with highlighted combinations preferred by the participants.

| | **Baseline** *(Mode 1)* | **AR Visual Cues** *(Mode 2)* | **Assistive Autonomy** *(Mode 3)* |
|---|---|---|---|
| Default | 1a: Single View | 2a: Single View | 3a: Single View |
| PA = Perception Augmentation | 1b: Fixed PIP (PA1) | 2b: Fixed PIP (PA1) | 3b: Fixed PIP (PA1) |
| | 1c: Pop-up PIP (PA2) | 2c: Pop-up PIP (PA2) | 3c: Pop-up PIP (PA2) |
| AA = Action Augmentation | 1d: Trackpad (AA1) | 2d: Trackpad (AA1) | 3d: Trackpad (AA1) |
| | 1e: Motion Scaling (AA2) | 2e: Motion Scaling (AA2) | 3e: Motion Scaling (AA2) |

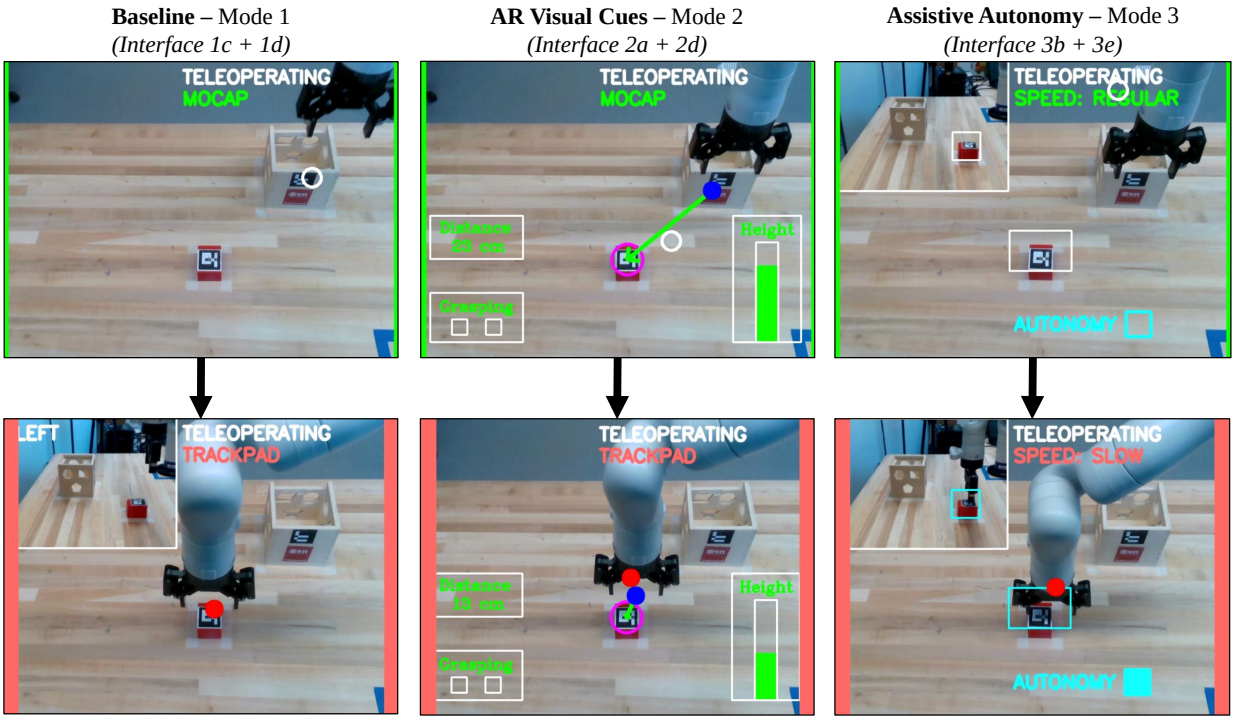



Figure 4.19: Integrated interfaces of perception and action augmentation.

Table 4.2 highlighted the augmentation combination preferred by the majority of the participants for each mode. In Mode 1 (i.e., the baseline control mode), 5 out of 8 participants preferred to have the pop-up picture-in-picture (PIP) display of the complementary viewpoint (interface-1c) and to use trackpad control (interface-1d). Some participants commented: *"...would like to have the pop-up PIP display to provide more workspace information when needed and use a trackpad to control the robot in a single direction for precise motion"*. In Mode 2 when the interface can display AR visual cues, 6 out of 8 participants preferred to use the single camera view display (interface-2a) without any perception augmentation and trackpad control (interface-2d). Participants commented: *"...the PIP display overwhelms the user interface while the AR visual cues are available"*. In mode 3 (assistive autonomy), 7 out of 8 participants preferred to use the fixed PIP display of the complementary viewpoint (interface-3b) and motion scaling (interface-3e). As the participants commented: *"...the fixed PIP display increases the awareness of the region where au-*

*tonomy is triggered"* and *"...motion scaling prevents large movement that moves the robot out of the autonomy zone"*. The preferred combination of interface control modes with perception and action augmentation will be further evaluated in our formal user study.

We further refine the interface display based on the freeform comments from the participants. Shown in Figure 4.19, we use the sidebar in pink and green to prominently indicate the activation of action augmentation. In Mode 3, we also highlight the region to activate the autonomous actions in both the primary and complementary viewpoints. The corresponding AR visual cue (i.e., the square around the object) will be turned from white to light blue color.

**Estimation of Cognitive and Physical Workload —** We estimate the cognitive workload using the operator's gaze and eye movement tracked by a Tobbi Pro Nano eye tracker. We also propose a novel method to estimate the physical workload from human motion tracking.

*Estimation of Cognitive Workload.* Following the methods in the literature [396, 397, 398], we estimate cognitive workload caused by stress $C_{str}$, interface complexity $C_{int}$ and task workload $C_{tsk}$ from the operator's pupil diameter, gaze fixation and movements, and task duration. We track the difference between the operator's pupil diameter and estimate the cognitive workload caused by stress as the difference between average pupil diameter ($D_{tsk}$ during a task and the operator's calibrated pupil diameter $D_{cal}$ before the task start, and normalize with respect to the maximum cognitive workload among all the participants, i.e., $C_{str} = \frac{\overline{D_{tsk}} - D_{cal}}{\max\limits_{p=p_1,\ldots,p_n}(\overline{D_{tsk}} - D_{cal})}$. Prior literature suggests that [396, 397, 398] pupil dilates with the increased workload, thus increasing the difference between the average pupil diameter during a task ($D_{tsk}$) and the operator's calibrated pupil diameter ($D_{cal}$) prior to the start of the task. The cognitive workload caused by interface complexity $C_{int}$ is computed as the ratio between the average distance in pixels of the operator's gaze fixation and the center of visual display and the maximum distance in pixels (from edge to center of visual display i.e., $S_{tsk}$ and $S_{max}$). Thus, the interface complexity can be calculated as, $C_{int} = \overline{S_{tsk}}/S_{max}$). To compute the cognitive workload for each sub-task (e.g., picking-and-place one object), we also es-

timate the cognitive workload caused by task complexity as the ratio between the time to complete a sub-task and total task completion time (namely, $C_tsk = T_{sub}/T_{total}$). Thus, the cognitive workload for a sub-task can be computed as the average of $C_{str}$, $C_{int}$, and $C_{tsk}$. We also contribute the overall workload $C_{task}$ of the entire task caused the stress and interface complexity as the average of $C_{str}$ and $C_{int}$, assuming they have equal contributions.

*Estimation of Physical Workload.* Surface Electromyography (sEMG) signals can provide more accurate measurements of the muscle efforts and physical workload than using subjective feedback (e.g., Rapid Upper Limb Assessment, namely RULA [399, 400, 401]). We used sEMG for the *objective* but *offline* estimation of physical workload in robot teleoperation via whole-body motion mapping [9, 18]. Here we propose to learn predictive models for the online, accurate muscle effort prediction from human motion tracking data. Our prior work [18] shows that: the muscle efforts of the anterior, lateral deltoid and bicep muscle groups, caused by shoulder flexion, abduction, and elbow flexion, contributes most to the physical workload when human controls tele-manipulation using their arm and hand motions.

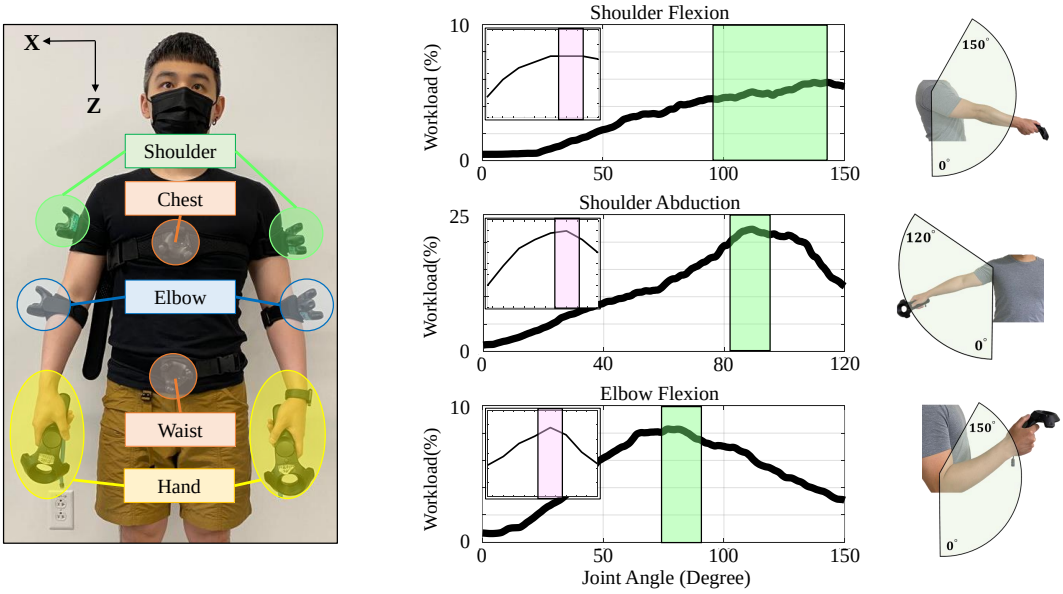



Figure 4.20: Vive trackers attachment and physical workload single joint mapping and validation.

Shown in Figure 4.20, we thus attached 6 body trackers (Vive Tracker 3.0) to the upper arms, forearms, chest, and waist of the human operator, to estimate the shoulder and elbow joint angles. Specifically, the shoulder flexion ($\theta_{SF}$) is estimated on the **sagittal plane** as:

$$\theta_{SF} = \arccos(\frac{\vec{T}_{ua} \cdot \vec{g}}{||\vec{T}_{ua}||\ ||\vec{g}||}) \tag{4.1}$$

which has $\vec{T}_{ua}$ to be the upper arm vector estimated from shoulder and elbow trackers, and the $\vec{g}$ to be the gravity vector, both of which are projected on the sagittal plane (i.e., the X-Y plane).

The shoulder abduction $\theta_{SA}$ is estimated on the **frontal plane** as:

$$\theta_{SA} = \arccos(\frac{\vec{T}_{vertical} \cdot \vec{T}_{ua}}{||\vec{T}_{vertical}||\ ||\vec{T}_{ua}||}) \tag{4.2}$$

which has $\vec{T}_{vertical}$ to be the vector perpendicular to the vector connecting two shoulder trackers, and $\vec{T}_{ua}$ to be the vector of the upper arm formed by shoulder and elbow trackers, both of which are projected on the frontal plane (i.e., the X-Z plane).

The elbow flexion $\theta_{EF}$ is estimated as

$$\theta_{EF} = \arccos(\frac{\vec{T}_{ua} \cdot \vec{T}_{la}}{||\vec{T}_{ua}||\ ||\vec{T}_{la}||}) \tag{4.3}$$

which has the $\vec{T}_{ua}$ to be the upper arm vector, and $\vec{T}_{la}$ is the forearm vector estimated from the elbow tracker and hand-held controller positions. Both these vectors are projected on the sagittal plane (i.e., the Y-Z plane). Note that: $0° < \theta_{SF} < 150°$, $0° < \theta_{SA} < 120°$ and $0° < \theta_{EF} < 150°$.

Shown in Figure 4.21 (Left), before tele-manipulation, we asked the human operator to perform a compound arm exercise that involves the active coordination of the anterior and lateral deltoid and the bicep muscle groups. The participants held one HTC Vive controller in each hand and

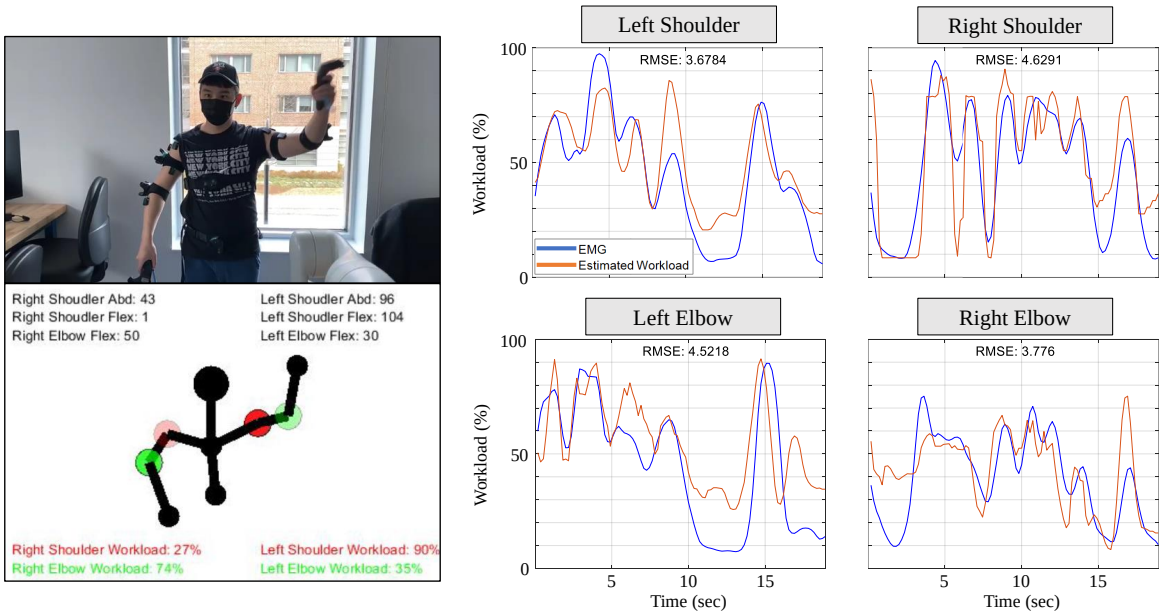



Figure 4.21: Physical workload estimation via Vive trackers.

move their shoulder and elbow from fully extended to fully flexed for 20 seconds at the speeds and angular velocities for typical robot control motions. We computed the joint angles of shoulder flexion, abduction, and elbow flexion from the body and arm motions tracked by the HTC Vive trackers, and used the corresponding sEMG data to estimate the offline muscle efforts [18]. For the offline workload estimation, we used a band pass filter to extract the 40 Hz-700 Hz EMG signals from the wireless sEMG sensors (Delsys Trigno Avanti Sensors) attached to the anterior, lateral deltoid, and bicep muscle groups. We pre-processed the data using a high pass filter (cutoff frequency 10 Hz) to remove the soft tissue artifact and offset the frequency baseline, and use a full-wave rectification then a sixth-order elliptical low pass filter (cutoff frequency 50 Hz) to remove noise and transients and develop a linear envelope of the EMG signals, following the method in the literature [153] but choose tunable parameters for our own task and data. The shoulder muscle efforts were computed using the weighted sum of the anterior and lateral deltoid (at the ratio of 3:4 based on their capabilities of force generation [402]), while the elbow efforts were calculated from the bicep flexion. The muscle efforts were computed by normalizing the processed EMG data with

respect to the person's maximum voluntary contraction following the standard procedure in the literature [154]. We averaged the shoulder and elbow muscle efforts for each arm, and estimate the operator's overall physical workload as the weighted sum of muscle efforts from the dominant and non-dominant arm (at the ratio of 9:1), for the tasks that operators extensively move their dominant arms for robot motion control:

$$P_{overall} = 0.9 \times (\frac{P_{DS} + P_{DE}}{2}) + 0.1 \times (\frac{P_{NDS} + P_{NDE}}{2}) \tag{4.4}$$

where $P_{DS}$ and $P_{NDS}$ are the shoulder muscle efforts of the dominant and non-dominant arms, while the $P_{DE}$ and $P_{NDE}$ are the elbow muscle efforts. A set of injunctive mapping functions were learned to predict the muscle efforts based on the arm joint angles with good accuracy. Figure 4.20 shows that our predictive model can estimate the sEMG-based physical workload based on the joint angles in isolation exercises, comparable to literature results [400, 401]. For compound exercises, Figure 4.21 (Right) shows an example of the prediction accuracy of our simple models for one male (32 years old) and 1 female (33 years old) of functional upper extremities and normal body mass index. The root mean square errors (RMSE) between the proposed method and EMG data are 3.68, 4.52, 4.63, and 3.78 for male and 4.97, 3.81, 4.12, and 4.37 for female participants for the left should, left elbow, right shoulder, and right elbow.

### 4.4.2 User Study

We conduct a user study to investigate: 1) what aspects of dexterous tele-manipulation are improved while using generally preferred improvements upon freeform teleoperation? 2) how do the different types of augmentation impact performance, workload, and preference? 3) when should the teleoperators be provided with the visual and action augmentation to increase task performance and decrease operational workload? 4) who should be provided with what type of augmentation

for freeform teleoperation?

**Experiment Setup —** Figure 4.11 shows the tele-manipulation system used for our experiments. The participants were instructed to control the Kinova Gen 3 manipulator robot to perform a single-object pick-and-place task using an HTC Vive hand-held controller. Two RealSense D435 cameras were set up to provide the primary and complementary viewpoints, to provide the back view and left side view of the workspace, while a desktop monitor was used to display the GUI (with camera viewpoints and AR visual cues) to the operator. HTC Vive body trackers and hand-held controllers were used to tracking their body and arm motions for online physical workload estimation. The Tobii Pro Nano screen-based eye tracker attached to the desktop monitor display was used to track the operator's gaze and eye movements for cognitive workload estimation. Unlike the pilot study, for this user study, a screen-based tracker was used because the visual interface was relayed on a computer screen as opposed to a head-mounted display.

**Participants —** Our experiments include N=23 participants ($28 \pm 10$ years old) diverse in gender, technological and professional experience. The participants were divided into several user groups based on the factors including:

- **Gender:** Based on gender, participants comprised of 14 male and 9 female participants.
- **Background:** The 23 participants comprised of 5 nurses and 18 users who do not have a nursing background. Participants were determined to have a nursing background if they are nursing students or registered nurses. Participants with nursing backgrounds were recruited in order to incorporate our intended future users for a teleoperation platform nursing in the development stage.
- **Proficiency:** The participants could be divided into 9 experienced and 14 inexperienced users based on their experience in having used the teleoperation system. Users were classified as experienced users if they had more than one hour of experience controlling the robot via

teleoperation. They must have also teleoperated the robot within one year to the day of their participation in the user study. The experienced users included participants of the pilot study for this user study, in addition to other experienced participants from prior user studies for different experiments.

- **Gaming:** The participants were divided into 16 infrequent video game players and 7 frequent video game players. Participants who spent less than 5 hours a week playing video games were classified as infrequent video game players.

- **Spatial:** Based on their spatial reasoning skills (via a spatial test from AssessmentDay [403]), the 23 participants were divided into 10 people with low spatial reasoning skills and 13 people with high spatial reasoning skills. The participants' spatial reasoning skill was evaluated using a spatial reasoning test that was part of the pre-user study survey. Participants who scored less than the 60 percentile on the test were evaluated to have low spatial reasoning skills.

- **Mode Order:** The participants were also divided based on the order in which the participants used the interface modes. 8 users used the interface modes in the $1 \rightarrow 2 \rightarrow 3$ order. 7 users used the interface modes in the $2 \rightarrow 3 \rightarrow 1$ order. 8 users used the interface modes in the $3 \rightarrow 1 \rightarrow 2$ order.

**Task —** The participants performed the same single-object pick-and-place task in all the trials. To focus on the comparison of different augmentation approaches, we simplify the task of picking up and dropping a block object into the container, which does not involve the control of robot/object orientation. The robot, object, and container were set to the same positions at the start of each trial (Figure 4.22.a). This task involves general-purpose manipulation actions and requires both gross and precise manipulation. Specifically, each task can be decomposed to four ***action phases*** (Figure 4.22.b and Figure 4.22.c), including: 1) **Reaching** to the object from the robot start position

to within 0.3 m from the target object; 2) adjust the robot pose for **Grasping** the object; 3) **Moving** the grasped object to be 0.15 m from the container; 4) adjust the robot pose for **Placing** the object to the container. Note that the tele-manipulation tasks could be more diverse and difficult than the single-object pick-and-place, our event-based robot autonomy, and augmentation (perception and action) can still adapt to different purpose tasks (e.g., assist the object alignment in the stacking task or interact with multi-target workspace in correct order).

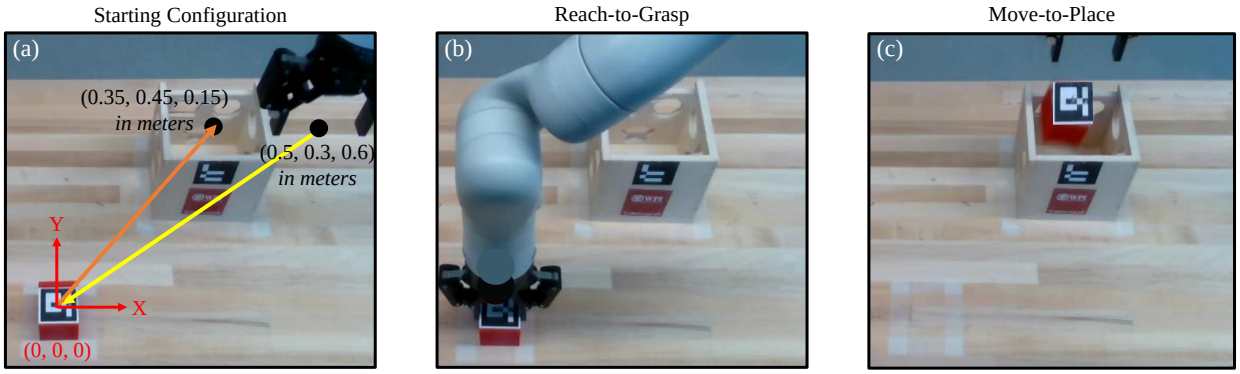



Figure 4.22: Robot starting configuration and task.

**Experiment Conditions —** In each mode, a participant performed the task twice: 1) without any augmentation (Default), 2) with each perception augmentation (PA1 and PA2), 3) with each action augmentation (AA1 and AA2), and 4) using the preferred combination identified in Pilot Study II. The total trials for each participant is 36 = 2 repetitions × 3 modes × (1 default + 2 PAs + 2 AAs + 1 preferred integration). To avoid the learning effects, participants performed a random selection of one of the three mode orders mentioned above.

**Experiment Procedure —** The experiment consists of a ***training section*** and ***performing section***. In the training section, the experimenter explained and demonstrated how to use the default interface of the selected starting mode, to perform the tele-manipulation task without any robot autonomy and interface augmentation for perception and action. The participants could practice the task (for a maximum of 15 min). The participants who felt confident to perform the task after practice would perform the practiced task under the aforementioned experiment conditions. Every participant

stated they felt confident in using the teleoperation interface within the allocated practice time. We recorded the task performance (e.g., task completion time, types, and occurrence of errors) during both the training and performing sections. Before the experiments, participants filled in a survey to report their experiences in video games, virtual reality environments, and spatial reasoning (via a spatial test from AssessmentDay [403]). Before the performing section, we asked the participants to look at the monitor for 30 seconds and recorded their pupil diameters, for the calibration required to estimate their cognitive workload. After completing the trials in each control mode, the participants filled in generic surveys, including NASA-TLX and System Usability Survey (SUS), and report their rating for each of the six interface conditions. After the experiment, the participants also filled in a customized questionnaire to report their preference for the control mode and interface conditions.

**Data Analysis —** Our data analysis considered *interface conditions* and *action phases* to be the independent variables, and considered the ***task performance***, ***workload***, and ***user preference*** to be the dependent variables. We measured the task performance objectively using the completion time, robot trajectory length, and types and occurrence of errors, for the entire task and for each action phase. We consider the physical and cognitive workload estimated using the methods mentioned previously and reported in the NASA-TLX survey. We also consider the user's preference inferred from the gaze fixations and distributions on the interfaces and reported in the SUS and the customized surveys. For all the comparisons, we analyzed data from all dependent variables using one-way repeated-measures analysis variance (ANOVA), including control modes, augmentations, action phases, and user groups, as a within-participants variable. All pairwise comparisons used Holm-Bonferroni correction to control for Type I error in multiple comparisons.

### 4.4.3 Effects of AR Visual Cues and Assistive Autonomy

From the comparison between different control modes (without any perception or action augmentations), we have the following results regarding the task performance, cognitive and physical workload. As shown in fig:results5-1, we found that: 1) *using autonomous actions can significantly reduce the occurrence of errors*; 2) *using AR visual cues can significantly reduce cognitive workload*; 3) *the overall preference of the participants for each mode was Mode 3 > Mode 2 > Mode 1*.

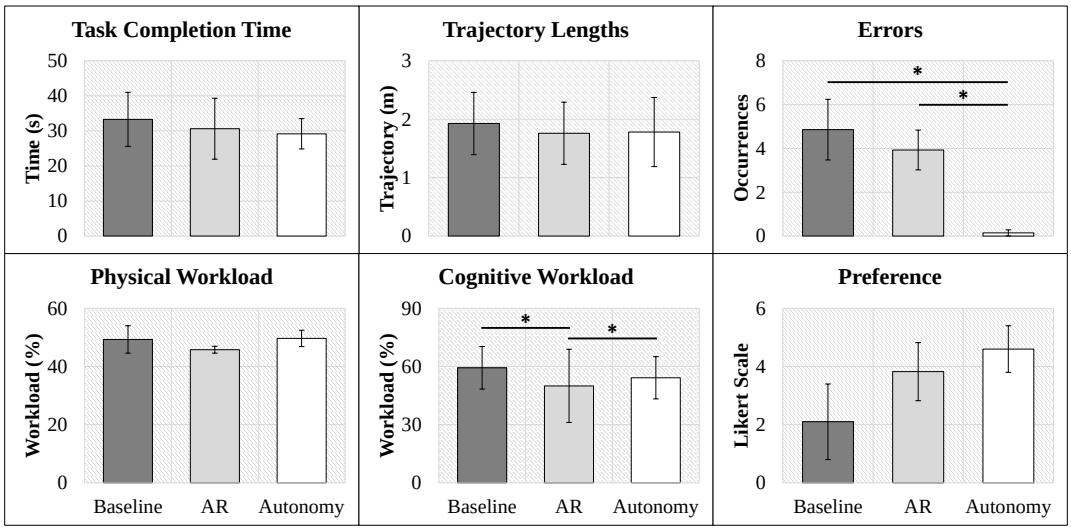



Figure 4.23: Comparison of task performance, workload, and preference between control modes.

**Task Performance —** The ***task completion time*** for Mode 1 (baseline), Mode 2 (with AR visual cues), and Mode 3 (with autonomous actions) were on average 33.3 $\pm$ 7.7, 30.6 $\pm$ 8.7 and 29.2 $\pm$ 4.3 seconds for all the participants. The participants completed the task faster (by 8% and 12%) with the assistive AR visual cues or autonomous actions than in Mode 1. The ***total trajectory lengths*** of the robot during the task for Mode 1, 2, and 3 are 1.99 $\pm$ 0.53, 1.76 $\pm$ 0.53, and 1.78 $\pm$ 0.59 meters, respectively. The trajectory lengths were shorter in Modes 2 and 3 (by 12% and 11%) than in Mode 1. The ***occurrence of errors*** during the task were 4.86 $\pm$ 1.38, 3.93 $\pm$ 0.91, and 0.15 $\pm$ 0.14 occurrences, for Mode 1, 2, and 3, respectively. The ANOVA analysis showed

no significant differences in the task completion time or the total trajectory lengths. However, post hoc comparisons showed that the occurrence of errors using Mode 3 (with autonomous actions) was significantly lower ($p < .01$) than using Mode 1 and 2, by 97% and 96%, respectively.

**Workload —** The ***physical workload*** while using Mode 1, 2, and 3 were on average 49.3 $\pm$ 4.7, 45.8 $\pm$ 1.2, and 49.7 $\pm$ 2.8 percent of the muscle capabilities. Mode 2 (with AR visual cues) led to a lower physical workload than Mode 1 (baseline) and Mode 3 (with autonomous actions) but without significant differences. The ***cognitive workloads*** were on average 59.3 $\pm$ 11, 50 $\pm$ 18.9 and 54.2 $\pm$ 10.9, respectively, when using Mode 1, 2, and 3. Post hoc comparisons showed that Mode 2 (with AR visual cues) led to a significantly lower ($p < .05$) cognitive workload compared to the other Modes.

**Preference —** Our post-experiment survey asked the participants *"Overall, how much do you prefer to use this interface for controlling the robot on a daily basis for your work?"*. The participants rated their preference for each mode using the Likert scale from 1 (the least) to 5 (the most). Mode 3 was the most preferred (with the highest score of 4.6 $\pm$ 0.8), while the scores for Mode 2 and 1 were 3.8 $\pm$ 1 and 2.1 $\pm$ 1.3, respectively.

### 4.4.4 Effects of Various Perception and Action Augmentation

From the comparison between different augmentations (perception, action, and integration), we have the following results regarding task performance and workload. Table 4.3 compares the performance and workload between the interfaces for each mode. The green (red) color indicates the best (worst) case among all the augmentation interfaces for each mode. We found that: 1) *using Fixed PIP (PA 1) can improve the performance of task completion time and total trajectory lengths*; 2) *using Trackpad (AA 1) can reduce the occurrence of errors*; 3) *using the Integrated interface can reduce the cognitive workload*. Table 4.6 compares the subjective feedback from NASA-TLX

and SUS forms between the interfaces for each mode. We found that: 1) *using Trackpad (AA 1) results in higher mental and physical workload*; 2) *using Scaling (AA 2) results in higher overall workload and lower SUS score*. Moreover, the analysis of gaze fixation indicates that: 1) *participants tend to check on the PIP more for all perception augmentations in Mode 1*; 2) *the perception augmentations reduce the usage of the AR visual cues in Mode 2*.

Table 4.3: Task performance and overall workload for all interfaces with mean and standard deviation. The green (red) color indicates the best (worst) case among all the augmentation interfaces for each mode.

| | Default *(Single View)* | Perception Augmentation (PA) *Fixed PIP* | Perception Augmentation (PA) *Pop-up PIP* | Action Augmentation (AA) *Trackpad* | Action Augmentation (AA) *Scaling* | Integrated *(PA + AA)* |
|---|---|---|---|---|---|---|
| *Time (s)* | | | | | | |
| Mode 1 | 33.3 (7.7) | 28.4 (5.8) | 28.8 (8.6) | 36.6 (4.8) | 34.5 (2.3) | 36.2 (7.9) |
| Mode 2 | 30.6 (8.7) | 31.4 (7.8) | 28.9 (6) | 40.6 (9.7) | 36.5 (3.1) | 34.6 (9.3) |
| Mode 3 | 29.2 (4.3) | 27 (4) | 28.8 (4.7) | 36.7 (3.4) | 32.1 (7.1) | 31.7 (3.1) |
| *Trajectory (m)* | | | | | | |
| Mode 1 | 1.99 (0.53) | 1.75 (0.39) | 1.75 (0.49) | 1.74 (0.3) | 1.83 (0.1) | 1.76 (0.31) |
| Mode 2 | 1.76 (0.53) | 1.58 (0.31) | 1.62 (0.48) | 1.73 (0.21) | 1.6 (0.19) | 1.61 (0.37) |
| Mode 3 | 1.78 (0.59) | 1.56 (0.37) | 1.64 (0.52) | 1.59 (0.21) | 1.64 (0.32) | 1.65 (0.37) |
| *Error (num.)* | | | | | | |
| Mode 1 | 4.86 (1.38) | 5.3 (1.77) | 4.61 (2.1) | 3.14 (0.85) | 5.63 (1.39) | 3.7 (1.26) |
| Mode 2 | 3.93 (0.91) | 4.39 (1.42) | 4.29 (2.06) | 1.93 (0.79) | 3.96 (1.44) | 2.01 (0.42) |
| Mode 3 | 0.15 (0.14) | 0.54 (0.42) | 1.27 (1.09) | 1.78 (1.59) | 1.59 (1.23) | 2.33 (1.22) |
| *Physical (%)* | | | | | | |
| Mode 1 | 49.3 (4.7) | 48.2 (2.3) | 48.6 (1.8) | 51 (1.3) | 49 (1.7) | 50.6 (2.9) |
| Mode 2 | 45.8 (1.2) | 48.6 (2.3) | 51 (2.2) | 52.5 (2.4) | 50 (2.3) | 51 (1.8) |
| Mode 3 | 49.7 (2.8) | 50.5 (2.5) | 51 (3.1) | 51.1 (1.4) | 50 (2) | 48 (2.5) |
| *Cognitive (%)* | | | | | | |
| Mode 1 | 59.3 (11) | 50.9 (12.8) | 51.5 (11.7) | 55.5 (10.5) | 55.3 (11.1) | 47.6 (12.4) |
| Mode 2 | 50 (18.9) | 46 (16.8) | 45.1 (18.1) | 47.8 (18.1) | 49.3 (13) | 48.7 (15.1) |
| Mode 3 | 54.2 (10.9) | 48.9 (12.7) | 45.4 (19.3) | 50.2 (16) | 48.2 (15.1) | 43.3 (13.7) |

*Mode 1: baseline | Mode 2: AR visual cues | Mode 3: assistive autonomy

**Task Performance and Workload —** Based on Table 4.3, we conducted multiple post hoc comparisons: 1) between the best and worst case for each mode, and 2) of the best and worst case across

three modes (Figure 4.24). The significant differences we found include:

*Task Completion Time.* In Mode 1, the interface with *Fixed PIP* outperforms the interface with *Trackpad*, with $p < .05$; In Mode 2, the interface with *Pop-up PIP* outperforms the interface with *Trackpad*, with $p < .05$; In Mode 3, the interface with *Fixed PIP* outperforms the interface with *Trackpad*, with $p < .01$; The ANOVA analysis showed no significant difference for the comparisons of the best and worst cases across three modes.

**Total Trajectory Length —** The ANOVA analysis showed no significant difference for the comparisons between the best and worst interface in each mode as well as of the best and worst cases across three modes.

**Occurrence of Errors —** In Mode 1, the interface with *Trackpad* outperforms the interface with *Scaling*, with $p < .01$; In Mode 2, the interface with *Trackpad* outperforms the interface with *Fixed PIP*, with $p < .01$; In Mode 3, the *Default* interface outperforms the interface with Preferred *Integration* of PA and AA, with $p < .01$; For the best interfaces in each mode, the ANOVA analysis showed a significant difference between modes and the post hoc comparisons indicated the *Default* in Mode 3 outperforms the *Trackpad* in both Mode 1 and 2, with $p < .05$. For the worst interfaces in each mode, the ANOVA analysis showed a significant difference between modes and the post hoc comparisons indicated the *Integrated* in Mode 3 outperforms the *Scaling* in Mode 1 and *Fixed PIP* in Mode 2, with both $p < .05$.

**Physical Workload —** In Mode 2, the *Default* interface outperforms the interface with *Trackpad*, with $p < .05$; The ANOVA analysis showed no significant difference for the comparisons of the best and worst cases across three modes.

**Cognitive Workload —** In Mode 1, the interface with preferred *Integration* of PA and AA outperforms the *Default* interface, with $p < .01$; In Mode 3, the interface with preferred *Integration* of PA and AA outperforms the *Default* interface, with $p < .01$; For the worst interfaces in each mode,

the ANOVA analysis showed a significant difference between modes and the post hoc comparisons indicated the *Default* in Mode 2 outperforms the *Default* in both Mode 1 and 3, with both $p < .05$.

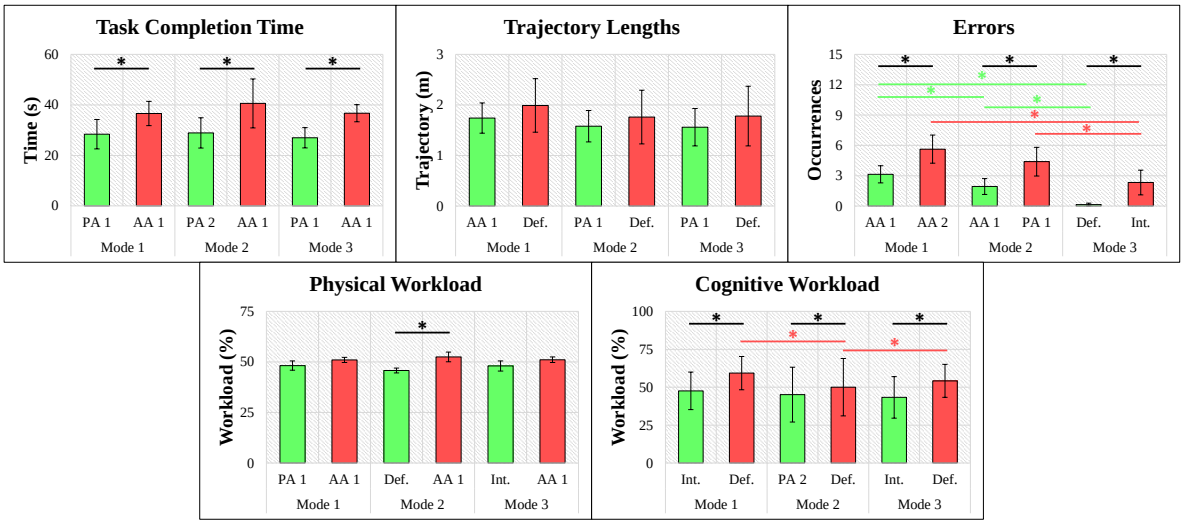



Figure 4.24: Comparison of task performance and workload for augmentations.

**Usage of Complementary View and AR Visual Cues —** Table 4.4 shows the duration of the gaze fixation on the complementary viewpoint (measured by the percentage of task completion time). Note that we only count the gaze fixation longer than 0.1 sec on a particular interface feature. We found that with perception augmentations including PA 1 (Fixed PIP), PA 2 (Pop-up PIP), and Integrated, the time that participants spent looking at the complementary viewpoint is longer (yet not significantly) in Mode 1 than in the other two modes.

Table 4.4: Duration of gaze fixation on the complementary viewpoints (w.r.t task completion time).

| | PA 1 | PA 2 | Integrated |
|---|---|---|---|
| Mode 1 | 29 (16) | 17 (9) | 21 (13) |
| Mode 2 | 16 (13) | 10 (9) | – |
| Mode 3 | 17 (16) | 10 (11) | 12 (15) |

Table 4.5 compares the gaze fixation duration on the different AR visual cues (measured by the percentage of task completion time). It shows that the height bar, the hint boxes, and the distance

box were the least used because the participants only glanced at them to find out the action affordance and confirmation. On the other hand, the object and box AR features were much more used. This is because the participants need to look at them to control the continuously performed reaching and moving motions, and to precisely adjust the robot position for grasping or placements. Table 4.5 also shows that the perception and action augmentations may change the participants' reliance and usage of the AR visual cues. Post hoc comparisons showed that PA 1 (Fixed PIP) significantly reduced the use of the "object" AR visual cue ($p < 0.05$), compared to AA 1 (Trackpad) or the integrated interface. However, the perception and action augmentations have no significant impacts on the use of the "box" cue.

Table 4.5: Duration of gaze fixation on the AR visual cues (w.r.t task completion time).

| | **AR Visual Cues** | | | | |
|---|---|---|---|---|---|
| | Object AR | Box AR | Height Bar | Hint Boxes | Distance |
| Default | 33 (12) | 26 (12) | 2 (4) | 0.8 (1) | 0.3 (1) |
| PA 1 | 30 (12) | 19 (11) | 1 (3) | 0.3 (0.9) | 0.2 (0.6) |
| PA 2 | 32 (13) | 22 (13) | 0.2 (0.5) | 0.2 (0.5) | 0.8 (1) |
| AA 1 | 34 (12) | 25 (13) | 2 (3) | 0.2 (0.6) | 0.2 (0.6) |
| AA 2 | 31 (11) | 28 (13) | 0.7 (1) | 0.5 (2) | 0.3 (2) |
| Integrated | 36 (12) | 25 (13) | 0.3 (1) | 0.4 (2) | 0.6 (2) |

**Subjective Feedback —** Table 4.6 compares the reported mental and physical workload from the NASA-TLX survey between the augmentation interfaces for each mode, and the reported usability from the SUS. We also compare the overall NASA-TLX score, using the coefficients of: 5 for mental demand, 4 for physical demand, 0 for temporal demand, 2 for performance, 3 for effort, and 1 for frustration. The weighting coefficients were generated by choosing from a series of pairs of rating scale factors that were deemed to be important based on the official instructions. Similar to Table 4.3, the green (red) color indicates the best (worst) case for each mode. We conducted multiple post hoc comparisons between the best and worst case for each mode. Here are the significant differences we found from the comparisons:

Table 4.6: NASA-TLX and SUS subjective feedback. The green (red) color indicates the best (worst) case among all the augmentation interfaces for each mode.

| | **Default** *(Single View)* | **Perception Augmentation (PA)** *Fixed PIP* | *Pop-up PIP* | **Action Augmentation (AA)** *Trackpad* | *Scaling* | **Integrated** *(PA + AA)* |
|---|---|---|---|---|---|---|
| *Mental Demand (NASA-TLX)* | | | | | | |
| Mode 1 | 41 (14) | 34 (2) | 33 (3) | 55 (11) | 54 (8) | 53 (17) |
| Mode 2 | 31 (2) | 33 (3) | 32 (10) | 49 (17) | 52 (13) | 48 (12) |
| Mode 3 | 24 (2) | 25 (3) | 27 (5) | 58 (11) | 47 (5) | 42 (10) |
| *Physical Demand (NASA-TLX)* | | | | | | |
| Mode 1 | 39 (4) | 33 (11) | 32 (11) | 48 (10) | 41 (8) | 42 (11) |
| Mode 2 | 28 (11) | 30 (11) | 30 (12) | 51 (7) | 42 (15) | 42 (15) |
| Mode 3 | 40 (9) | 25 (4) | 27 (4) | 45 (11) | 40 (11) | 23 (5) |
| *Overall Workload (NASA-TLX)* | | | | | | |
| Mode 1 | 39 (9) | 33 (3) | 31 (4) | 50 (11) | 54 (7) | 48 (7) |
| Mode 2 | 30 (3) | 32 (2) | 32 (7) | 48 (6) | 53 (10) | 45 (7) |
| Mode 3 | 42 (12) | 26 (3) | 28 (4) | 50 (13) | 46 (8) | 24 (4) |
| *SUS* | | | | | | |
| Mode 1 | 73 (2) | 81 (3) | 82 (2) | 51 (13) | 49 (7) | 55 (12) |
| Mode 2 | 81 (1) | 76 (4) | 76 (3) | 57 (14) | 56 (10) | 61 (15) |
| Mode 3 | 84 (3) | 81 (2) | 77 (5) | 50 (10) | 61 (6) | 65 (10) |

*Mode 1: baseline | Mode 2: AR visual cues | Mode 3: assistive autonomy

*Mental Demand.* In Mode 1, the interface with PA 1 (*Fixed PIP*) significantly outperforms the interface with AA 1 (*Trackpad*), with $p < .01$; In Mode 2, the *Default* interface significantly outperforms the interface with AA 2 (*Scaling*), with $p < .05$; In Mode 3, the *Default* interface significantly outperforms the interface with AA 1 (Trackpad), with $p < .01$.

*Physical Demand.* In Mode 2, the *Default* interface significantly outperforms the interface with AA 1 (*Trackpad*), with $p < .05$; In Mode 3, the *Integrated* Interface significantly outperforms the interface with AA 1 (Trackpad), with $p < .01$;

*Overall Workload.* In Mode 1, the interface with PA 2 (*Pop-up PIP*) significantly outperforms the interface with AA 2 (*Scaling*), with $p < .01$; In Mode 2, the *Default* interface significantly outperform the interface with AA 2 (*Scaling*), with $p < .01$; In Mode 3, the *integrated* interface

significantly outperforms the interface with AA 1 (*Trackpad*), with $< .05$;

*SUS.* In Mode 1, the interface with PA 2 (*Pop-up PIP*) significantly outperforms the interface with AA 2 (*Scaling*), with $p < .01$; In Mode 2, the *Default* interface significantly outperforms the interface with AA 2 (*Scaling*), with $p < .05$; In Mode 3, the *Default* interface significantly outperforms the interface with AA 1 (*Trackpad*), with $p < .01$;

### 4.4.5 Effects on Different Action Phases

We further analyze the interface modes and perception/action augmentation on the performance and workload for the different action phases. From the comparison between different action phases in each mode, we have the following results regarding the task performance, cognitive and physical workload. As shown in Figure 4.25, we averaged the data from all augmentation interfaces (default, PAs, AAs, integrated) for each action phase (i.e., reaching, grasping, moving, and placing) in each mode (i.e., baseline, AR visual cues, and assistive autonomy). The ANOVA analysis and multiple post hoc comparisons showed that: 1) *the action phase of grasping takes a significantly ($p < .01$) longer time than the reaching phase in all modes*; 2) *the action phase of placing results in a significantly ($p < .01$) higher physical workload in all modes*; 3) *the action phase of grasping and placing results in a significantly ($p < .05$) higher cognitive workload than the reaching and moving phases respectively in all modes*.

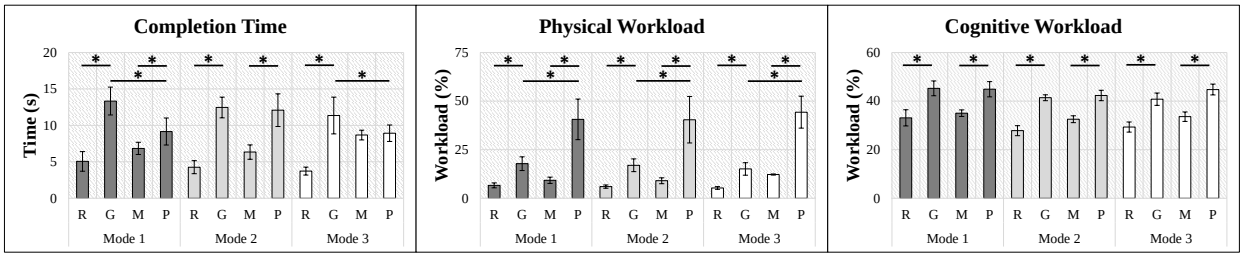



Figure 4.25: Comparison of completion time and workload between action phases for each mode.

*Task Completion Time.* Table 4.7 shows the task completion time of each action phase for all the interfaces of all the 3 modes. Post hoc comparisons showed that: The *grasping* action takes significantly longer time than the *reaching* action, for all interfaces and all the modes, with $p < .01$. The *placing* action also takes significantly less time than the *moving* action for all the interfaces of Mode 1 and 2, but not for Mode 3, with $p < .01$. We noticed that the interface that takes the least time for the **grasping** action is different for each mode. Specifically, it is the interface with PA 2 (*Pop-up PIP*) in Mode 1, the *Default* interface in Mode 2, and the interface with PA 1 (*Fixed PIP*) in Mode 3. The interface with PA 1 (*Fixed PIP*) in Mode 3 takes the least time to grasp, across all the modes and interfaces. We also noticed that the interface that takes the least time for the **placing** action is different for each mode. Specifically, it is the interface with PA 1 (*Fixed PIP*) in Mode 1, the interface with PA 2 (*scaling*) for Modes 2 and 3. The interface with PA 2 in Mode 2 takes the least time to place, across all the modes and interfaces. Note that the interfaces that take the least time for the grasping and placing action are highlighted in green in Table 4.7.

Table 4.7: Time of each action phase for all interfaces.

| | Default | PA1 | PA2 | AA1 | AA2 | Integrated |
|---|---|---|---|---|---|---|
| **Mode 1: Baseline** | | | | | | |
| Reaching | 7.2 (3.4) | 5.5 (1.3) | 6.1 (1.8) | 4.4 (1.1) | 3.4 (0.5) | 3.7 (0.6) |
| Grasping | 14.2 (3.4) | 11.5 (2.5) | 10.6 (3.8) | 16.2 (1.1) | 12.8 (2.6) | 14.7 (2.3) |
| Moving | 6.6 (2.8) | 6.4 (2.7) | 5.7 (1.8) | 7.5 (3) | 6.5 (1.4) | 8.3 (3.6) |
| Placing | 7.8 (0.6) | 7.2 (1.6) | 7.8 (2.7) | 9.2 (2.1) | 12.6 (4.3) | 10.3 (4.5) |
| **Mode 2: AR Visual Cues** | | | | | | |
| Reaching | 5.5 (1.4) | 4.8 (0.7) | 3.9 (0.9) | 4.9 (0.2) | 2.9 (0.4) | 3.5 (0.9) |
| Grasping | 9.7 (2.1) | 12.6 (3.5) | 11.9 (3.5) | 14.3 (2.8) | 13.2 (1.3) | 13 (3.2) |
| Moving | 6.1 (2.5) | 5 (0.5) | 5.2 (1) | 7.6 (2.9) | 7.3 (1.8) | 6.8 (2.4) |
| Placing | 10.7 (5.6) | 10.4 (5.9) | 9.1 (2.9) | 15.5 (7.3) | 14.2 (2.3) | 12.6 (6.3) |
| **Mode 3: Assistive Autonomy** | | | | | | |
| Reaching | 4.9 (1.1) | 3.4 (0.3) | 3.6 (0.8) | 3.6 (0.2) | 3.6 (0.5) | 3.2 (0.4) |
| Grasping | 9.5 (2.4) | 8.6 (1.6) | 9.6 (1.7) | 16.1 (2.2) | 11.7 (4) | 12.6 (2.5) |
| Moving | 7.6 (1.1) | 8.7 (2.2) | 9.3 (3.3) | 8.2 (1.6) | 9.6 (4.1) | 8.6 (1.6) |
| Placing | 8.6 (1.2) | 8 (1.9) | 7.1 (1.3) | 10.4 (2.1) | 9.7 (2.2) | 9.7 (1.1) |

*PA: perception augmentation | AA: action augmentation

*Physical Workload.* Table 4.8 shows the physical workload of each action phase for all the interfaces of all the 3 modes. Post hoc comparisons showed that: The *grasping* action has a significantly higher physical workload than the *reaching* action, for all interfaces and all the modes, with $p < .05$. The *placing* action also has a significantly higher physical workload than the *moving* action for all the interfaces of all the modes, with $p < .01$. We noticed that the interface that has the least physical workload for the **grasping** action is different for each mode. Specifically, it is the interface with PA 2 (*Pop-up PIP*) in Mode 1, the *Default* interface in Mode 2, and the interface with PA 1 (*Fixed PIP*) in Mode 3. The interface with PA 1 (*Fixed PIP*) in Mode 3 takes the least time to grasp, across all the modes and interfaces. However, the interface that has the least physical workload for the **placing** action is the *Default* interface for all three modes. The *Default* interface in Mode 2 has the least physical workload to place, across all the modes and interfaces. Note that the interfaces that have the least physical workload for the grasping and placing action are highlighted in green in Table 4.8.

Table 4.8: Physical workload of each action for all interfaces.

| | Default | PA1 | PA2 | AA1 | AA2 | Integrated |
|---|---|---|---|---|---|---|
| **Mode 1: Baseline** | | | | | | |
| Reaching | 8.5 (0.6) | 7 (0.6) | 7.9 (0.6) | 5.9 (0.6) | 4.9 (0.5) | 5.5 (0.1) |
| Grasping | 16.7 (0.9) | 15 (1.4) | 13.3 (1.4) | 23.9 (3.6) | 17.2 (3.7) | 20.6 (3.1) |
| Moving | 8.4 (1.2) | 7.7 (0.8) | 7.4 (1) | 10.8 (2.3) | 9 (0.6) | 11.9 (2.6) |
| Placing | 29.4 (4.9) | 33 (5.2) | 36.7 (2.4) | 34.6 (2.8) | 57.1 (5.1) | 52.7 (2.8) |
| **Mode 2: AR Visual Cues** | | | | | | |
| Reaching | 6.4 (0.3) | 6.5 (0.6) | 5.8 (0.9) | 7.3 (1.2) | 4.8 (1.3) | 5 (1) |
| Grasping | 11.9 (1.3) | 15.7 (0.2) | 13.9 (1.2) | 20.2 (1.3) | 20 (2.1) | 20 (3.7) |
| Moving | 7.3 (1.1) | 7.4 (0.7) | 8 (0.7) | 11.3 (2.5) | 10.1 (0.4) | 9.8 (1.3) |
| Placing | 17.4 (1.4) | 35.4 (5.3) | 42 (4.8) | 42.6 (2.9) | 53.7 (1.4) | 51.3 (2.8) |
| **Mode 3: Assistive Autonomy** | | | | | | |
| Reaching | 6.3 (0.6) | 5.1 (0.4) | 4.6 (0.6) | 6.1 (1.6) | 5.2 (0.9) | **4.3 (0.7)** |
| Grasping | 12.8 (1.1) | 11.9 (2) | 12.8 (1.3) | 21.4 (1.5) | 15.6 (2.4) | 15.9 (1.4) |
| Moving | 12 (0.4) | 11.9 (0.4) | 12.5 (1.3) | 11.9 (0.1) | 12.5 (2) | 12.4 (1.1) |
| Placing | 32.6 (2) | 38.9 (6.1) | 40.8 (4.6) | 43.6 (3.5) | 54.8 (5.7) | 55 (5.1) |

*PA: perception augmentation | AA: action augmentation

*Cognitive Workload.* Table 4.9 shows the cognitive workload of each action phase for all the interfaces of all the 3 modes. Post hoc comparisons showed that: The *grasping* action has a significantly higher cognitive workload than the *reaching* action, for all interfaces and all the modes, with $p < .01$. The *placing* action also has a significantly higher cognitive workload than the *moving* action for all the interfaces of all the modes, with $p < .01$. We noticed that the interface that has the least cognitive workload for the **grasping** action is different for each mode. Specifically, it is the integrated interface in Mode 1, the AA 2 (scaling) interface in Mode 2, and the integrated interface in Mode 3. The integrated interface in Mode 3 takes the least cognitive workload to grasp, across all the modes and interfaces. In terms of *placing*, PA 2 (pop-up PIP) caused the least cognitive workload for Mode 1 and 2, and integrated interface caused the least cognitive workload for Mode 3. The PA 2 (pop-up PIP) interface in Mode 2 caused the least cognitive workload to place, across all the modes and interfaces. Note that the interfaces that have the least cognitive workload for the grasping and placing action are highlighted in green in Table 4.9.

Table 4.9: Cognitive workload of each action for all interfaces.

| | Default | PA1 | PA2 | AA1 | AA2 | Integrated |
|---|---|---|---|---|---|---|
| **Mode 1: Baseline** | | | | | | |
| Reaching | 37 (8.7) | 34.7 (9.6) | 37 (8.3) | 31.3 (8.4) | 29.7 (6.5) | 28.7 (8.5) |
| Grasping | 49.7 (7) | 43.3 (7.6) | 42.5 (9) | 49 (7.9) | 44.8 (8.4) | 42.1 (8.4) |
| Moving | 37.6 (10.6) | 34.3 (10.4) | 35.7 (7.9) | 35 (7.4) | 34.1 (9.2) | 33.4 (10.7) |
| Placing | 48 (9.8) | 42.9 (9.9) | 41.7 (8.2) | 45.1 (8.7) | 49.8 (9.8) | 41.7 (8.6) |
| **Mode 2: AR Visual Cues** | | | | | | |
| Reaching | 31.3 (12.2) | 29.2 (11.4) | 25.4 (13.3) | 28.1 (10.7) | 25.5 (10) | 27.4 (8.1) |
| Grasping | 42.2 (12.5) | 41.6 (10.3) | 39.9 (10.4) | 41.3 (11.3) | 39.9 (9.3) | 43.3 (10.7) |
| Moving | 33.3 (11.1) | 30.5 (11.7) | 31.8 (11.2) | 32.4 (12.3) | 35 (8.1) | 32.1 (10) |
| Placing | 43.2 (11.2) | 39.9 (11.4) | 39.6 (11.3) | 42.1 (12.2) | 45.9 (9.2) | 43 (10.7) |
| **Mode 3: Assistive Autonomy** | | | | | | |
| Reaching | 33.3 (8) | 29 (8.8) | 28.6 (11.8) | 30.2 (9.7) | 28 (8.4) | 26.6 (8.2) |
| Grasping | 44 (7.3) | 40.3 (8.6) | 38 (10.6) | 44.1 (9.5) | 40.5 (8.3) | 37.7 (8.4) |
| Moving | 37.2 (7.6) | 34.4 (9.1) | 33.6 (12.6) | 31.7 (10.1) | 33.1 (10.6) | 31.4 (10.1) |
| Placing | 47.7 (8.3) | 45.7 (9.1) | 42.4 (12.5) | 46 (10.7) | 45.4 (11.3) | 41.2 (10.3) |

*PA: perception augmentation | AA: action augmentation

### 4.4.6 Effects of Other Human Factors

We further analyze the effects of several human factors, including gender, background, and experience with technology, on the performance and usage of the interfaces. The participants were divided into groups based on conditions defined previously. the comparison between different user groups, we have the following results regarding task performance in terms of completion time. As shown in Figure 4.26, the ANOVA analysis and multiple post hoc comparisons showed that: 1) *users' background impacts the task performance with more augmentation interfaces and control modes*; 2) *using assistive autonomy can mitigate the gap between different user groups in task performance*.

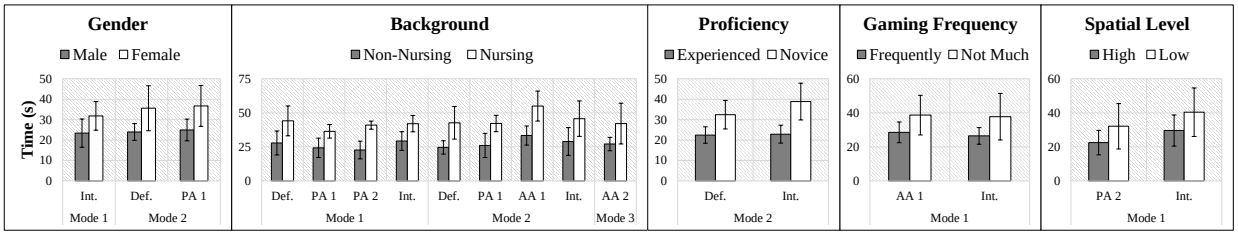



Figure 4.26: Indication of the significant differences in the comparison of completion time between user groups for each mode.

Table 4.10 compares the task completion time between user groups of different gender, background, proficiency with the tele-manipulation interface, gaming frequency, spatial reasoning level, and mode order. We highlighted the identified significant differences, which indicate that the male, non-nursing, experienced users, frequent video game players, and high-level spatial reasoning users took less time to complete the task.

Table 4.11 compares the use of the complementary viewpoint between user groups, using the duration of gaze fixation on the complementary viewpoint with respect to the total task completion time. We found significant differences between the male and female user groups when using the interface with PA 1 (*Fixed PIP*) and using the *integrated* interface in Mode 1. For both interfaces,

Table 4.10: Comparison between groups: completion time.

| | *Mode 1* | | | | | | *Mode 2* | | | | | | *Mode 3* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Def. | PA1 | PA2 | AA1 | AA2 | Int. | Def. | PA1 | PA2 | AA1 | AA2 | Int. | Def. | PA1 | PA2 | AA1 | AA2 | Int. |
| G | .53 | .30 | .09 | .40 | .35 | < .05 | < .05 | < .05 | .31 | .25 | .50 | .12 | .45 | .33 | .37 | .31 | .28 | .87 |
| B | < .05 | < .05 | < .01 | .14 | .50 | < .01 | < .01 | < .05 | .10 | < .05 | .32 | < .05 | .06 | .07 | .09 | .15 | < .05 | .19 |
| P | .50 | .19 | .09 | .08 | .56 | .08 | < .05 | .06 | .06 | .07 | .91 | < .01 | .09 | .06 | .07 | .11 | .23 | .55 |
| F | .63 | .17 | .10 | < .05 | .85 | < .05 | .11 | .15 | .61 | .25 | .14 | .15 | .24 | .08 | .27 | .18 | .28 | .17 |
| S | .35 | .36 | < .05 | .17 | .52 | < .05 | .13 | .10 | .36 | .06 | .47 | .38 | .14 | .32 | .16 | .92 | .21 | .87 |
| M | .97 | .37 | .28 | .57 | .83 | .73 | .64 | .89 | .75 | .38 | .79 | .62 | .67 | .52 | .95 | .98 | .48 | .69 |

*PA: perception augmentation | AA: action augmentation
*G: gender | B: background | P: proficiency | F: gaming frequency | S: spatial reasoning level | M: mode order

the male participants used the complementary viewpoints more than the female participants by 13% and 8.7% on average, respectively. We also noticed that: in Mode 1, both PA 1 and PA 2 led to a significant difference in the use of complementary viewpoints between the users with and without nursing professional experience, with $p < .05$ and $p < .01$, respectively. Participants without nursing experience or training used the complementary viewpoint 13.4% more on average for the interface with PA 1 (*Fixed PIP*), and 9.8% more on average for the interface with PA 2 (*Pop-up PIP*). Moreover, we found a significant difference between user groups of different spatial reasoning skills in the use of complementary viewpoints in Mode 1 and Mode 3 when using the integrated interface, with $p < .01$ and $p < .05$. In Mode 1 and Mode 3, when using the integrated interface, participants with lower spatial reasoning skills use the complementary viewpoint more by 10.3% and 9.2% on average.

Table 4.11: Comparison between user groups: use of complementary view.

| | **Mode 1** | | **Mode 2** | | **Mode 3** | | **Integrated** (Mode 1) | **Integrated** (Mode 3) |
|---|---|---|---|---|---|---|---|---|
| | Fixed | Pop-up | Fixed | Pop-up | Fixed | Pop-up | | |
| Gender | < .05 | .08 | .21 | .14 | .12 | .13 | < .05 | .47 |
| Background | < .05 | < .01 | .46 | .90 | .68 | .10 | .06 | .21 |
| Proficiency | .58 | .51 | .36 | .96 | .89 | .65 | .43 | .78 |
| Gaming | .71 | .63 | .80 | .45 | .60 | .64 | .80 | .22 |
| Spatial | .17 | .21 | .21 | .14 | .41 | .10 | < .01 | < .05 |
| Mode Order | .35 | .22 | .12 | .09 | .33 | .60 | .82 | .08 |

Table 4.12 highlights the significant difference between user groups by the use of the “object”

and "box" AR visual cues. In Mode 2, we found significant differences ($p < .05$) between users of different **background** in the use of the "object" cue when using the interface with AA 1 (*Trackpad*), and in the use of the "box" cue when using the *Default* interface. Specifically, the users without nursing experience used the "object" cue more by 8% when using the interface with AA 1, while the users with nursing experience use the "box" cue more by 10.3% when using the Default interface. Regarding the factor of **proficiency**, we found that: in Mode 2, the participants with prior experience of robot teleoperation used the "object" cue significantly more (with $p < .05$) than the other group by 7.9%, when using the interface with PA 1 (*Fixed PIP*). We also found that the high-proficiency group used the "object" cue significantly more (with $p < .05$ and $p < .01$) than the other group by 8.8% and 9.0% when using the interface with AA 1 (*trackpad*) and AA 2 (*scaling*). Regarding the effects of **gaming experience**, we found that frequent video game players used the "object" cue significantly more ($p < .01$) than the other user group when using the interface with AA 1 (*Trackpad*), by 10.3% on average. We also found that frequent video game players used the "object" cue significantly more ($p < .05$) than the other user group when using the *integrated* interface, by 7.3% on average. Regarding the factor of **spatial reasoning skills**, we found that the participants with better spatial reasoning skills used the "object" cue significantly more ($p < .05$) than the other users when using the interface with AA 1 (*Trackpad*), by 7.0% on average. We also found that the participants with better spatial reasoning skills used the "box" cue significantly less ($p < .05$) than the other users when using the interface with AA 1 (*Trackpad*, by 9.1% on average). Regarding the effects of **mode orders**, we found that the users performed the mode in $3 \rightarrow 1 \rightarrow 2$ order used the "box" cue significantly more ($p < .05$) than in $2 \rightarrow 3 \rightarrow 1$ order when using the *Default* interface, by 12.3% on average. We also found that the participants used the mode in $1 \rightarrow 2 \rightarrow 3$ order used the "box" cue significantly more ($p < .05$) than in $2 \rightarrow 3 \rightarrow 1$ order when using the interface with PA 1 (*Fixed PIP*), by 10.8% on average.

Table 4.13 shows the trend in the use of the height bar, hint box, and distance between user

Table 4.12: Comparison between user groups: use of object and box AR visual cues.

| | Def. | PA1 | PA2 | AA1 | AA2 | Int. | Def. | PA1 | PA2 | AA1 | AA2 | Int. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ***Object AR*** | | | | | | ***Box AR*** | | | | | |
| G | .18 | .09 | .56 | .55 | .17 | .15 | .16 | .29 | .21 | .66 | .99 | .56 |
| B | .42 | .64 | .98 | < .05 | .05 | .23 | < .05 | .45 | .42 | .19 | .95 | .44 |
| P | .08 | < .05 | .15 | < .05 | < .01 | .40 | .59 | .99 | .23 | .59 | .73 | .35 |
| F | .74 | .17 | .86 | < .01 | .22 | < .05 | .67 | .92 | .91 | .69 | .12 | .85 |
| S | .82 | .68 | .37 | < .05 | .07 | .33 | .50 | .73 | .77 | < .05 | .75 | .76 |
| M | .63 | .85 | .13 | .64 | .62 | .34 | < .05 | < .05 | .09 | .53 | .96 | .34 |

*G: gender | B: background | P: proficiency | F: frequency of gaming | S: spatial | M: mode order

groups. The table presents a comparison of the total number of frames for each AR cue used while performing the task across all the participants in the user group. We highlighted the user group that used these features more, measured by the total number of frames.

Table 4.13: Comparison between user groups: use of height bar, hint boxes and distance AR visual cues.

| | Def. | PA1 | PA2 | AA1 | AA2 | Int. | Def. | PA1 | PA2 | AA1 | AA2 | Int. | Def. | PA1 | PA2 | AA1 | AA2 | Int. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ***Height Bar*** | | | | | | ***Hint Boxes*** | | | | | | ***Distance*** | | | | | |
| G | | | | | | | | | | | | | | | | | | |
| B | | | | | | | | | | | | | | | | | | |
| P | | | | | | | | | | | | | | | | | | |
| F | | | | | | | | | | | | | | | | | | |
| S | | | | | | | | | | | | | | | | | | |
| M | | | | | | | | | | | | | | | | | | |

*Highlighted: male (G); nurse (B); novice (P); low gaming (F); bad spatial (S); Order 1 ■ and Order 2 □ (M)

## 4.5 Assistive Autonomy Levels and AR Preferences

To enhance remote perception, AR visual cues are overlaid on the video stream from remote cameras to indicate the robot and task states, and the spatial relationship in a 3D environment, which may be difficult for human operators to estimate precisely. They are also used to indicate the mo-

tion, action, path, and task that robot autonomy plans to perform, in order to enhance human's understanding of the robot's intent, behavior, and capabilities. Thus far, effective AR visual cues are mostly developed case-by-case for remote robot manipulation under direct to supervisory control. It is still not clear what AR visual cues humans need or prefer to use to control a remote manipulator robot with various levels of autonomy.

To this end, we proposed systematic AR visual cues for remote robot manipulation assistance. These AR cues can guide human's control of robot motion toward the target and around the obstacles in a 3D workspace and indicate whether the robot autonomy is activated, and its planned motion or action.

### 4.5.1 Control Modes for Remote Manipulation

Table 4.14: Human (**H**) and Robot (**R**) task division in each control mode.

| **Control Mode** | **Gross Manipulation** | **Obstacle Avoidance** | **Precise Manipulation** | **Autonomy Activation** |
|---|---|---|---|---|
| Direct | H | H | H | N/A |
| Assisted | H/R | H/R | H/R | Auto |
| Shared | H/R | R | R | Auto |
| Supervisory | R | R | R | Auto |

Table 4.14 illustrates the human-robot task division in each control mode. In the **Direct Control** mode, a human manually controls the robot through the entire task. In the **Assisted Control** mode, robot autonomy uses a virtual fixture to constrain human-controlled robot motion in directions that avoid collision with environmental constraints and obstacles as well as move to grasp/place an object. In the **Shared Control** mode, the robot assists human-controlled gross manipulation motions (e.g., approaching or moving an object) with automated motions for collision avoidance, and for precise manipulation (e.g., grasping or placing) when the robot end-effector is close enough to the target object or location. In the **Supervisory Control** mode, robot autonomy plans the motion

for the entire pick-and-place task, while humans supervise its execution. The robot's autonomous perception using AruCo markers [404] are available for all the control modes, in order to locate the target object, container, obstacle, and track the robot end-effector.

### 4.5.2 AR Features

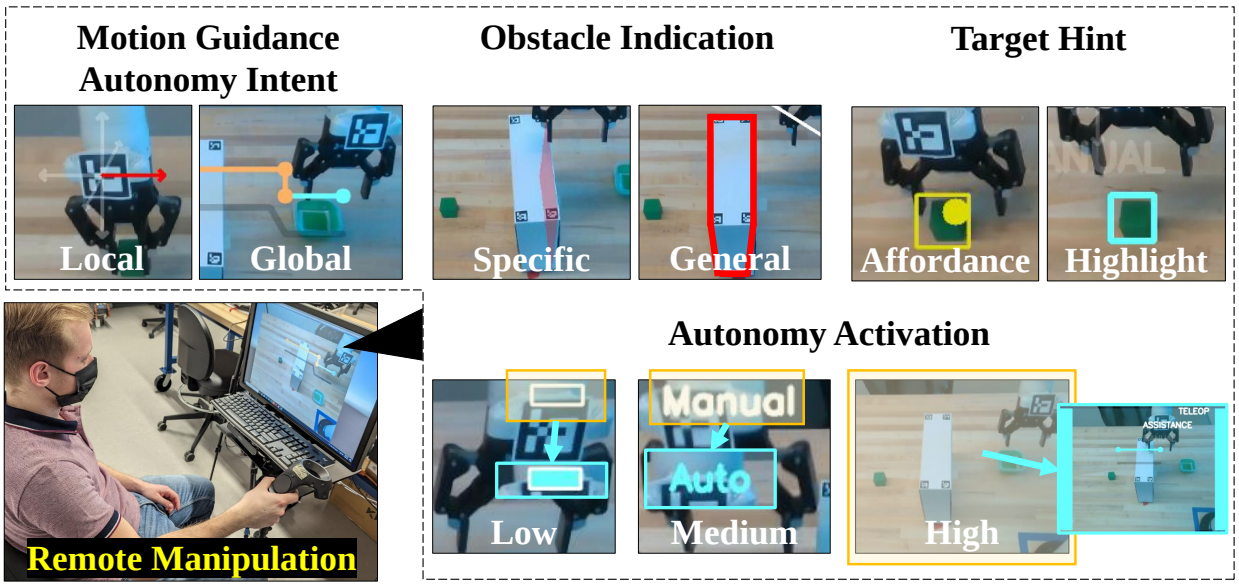



Figure 4.27: Proposed AR visual cues to assist humans to control or supervise remote robot manipulation, and to communicate the robot autonomy's activation, capabilities, and intents.

Shown in Figure 4.27, we propose five types of AR visual cues and representation options:

- **Motion Guidance** indicates the robot's instantaneous motion direction using a *3-axis arrows* overlaid on the robot end-effector or displays the robot's suggested path [405].

- **Obstacle Indication** has the options to highlight close-to-collision features on the obstacle (e.g., plane, edge, or vertex, as in [406]), and to display the obstacle's 3D bounding box (as in [407]).

- **Target Hint** provides the grasp/place *affordance* [380] by changing the color of the square around the target and the dot overlaid on the robot end-effector. Alternatively, the target can be *highlighted* [282] to provide an intent inference to the user.

- **Autonomy Activation** indicates if the robot's autonomy has been activated. We provide three options for the representation of different visual salience, including a blue light (low salience, as in [408]), text with "AUTO" (medium salience, as in [409]), and blue bars on both sides of the camera view (high salience, as in [406]).

- **Autonomy Intention** has the option to indicate the robot's motion, action, and path plan (as in [408, 410]). The autonomy intention is displayed similarly to the visual cues used for motion guidance.

Table 4.15: AR visual cue choices recommended by experienced users.

| **Control Mode** | **Motion Guidance** | **Obstacle Indication** | **Target Hint** | **Autonomy Activation** | **Autonomy Intention** |
|---|---|---|---|---|---|
| Direct | Arrow | Planes | Affordance | – | – |
| Assisted | Arrow | Box | None | Low | Path |
| Shared | Path | None | Highlight | High | Path |
| Supervisory | – | None | None | Medium | Path |

Our user study allows participants to choose the types and options of AR visual cues for each control mode. Table 4.15 shows the choices recommended by experienced users.

### 4.5.3 User Study

We conduct a user study to investigate: 1) what AR visual cues do humans prefer when controlling a robot with various levels of autonomy? 2) whether this preference can be influenced by the way humans learn to use the interface.

**Participants and Task —** We recruited 18 participants (13 male and 5 female, $25.4 \pm 6.9$ years) from the WPI campus to perform a single object pick-and-place task with an obstacle between the object and the box. A post hoc power analysis with (1-$\beta$) = 0.8 and $\alpha$ = 0.05 found an observed power of 0.85 (d = -1.05) with this sample size. Shown in Figure 4.28, participants controlled the robot to pick up an object on the other side of an obstacle (box of size 200 mm $\times$ 80 mm $\times$ 200 mm), and bring it back and place it in a small container. Participants performed the task under three different initial task states: the object location remained the same, while the container and robot were placed in three different ways such that the robot planned path would move around different sides of the obstacle to reach the object and container.

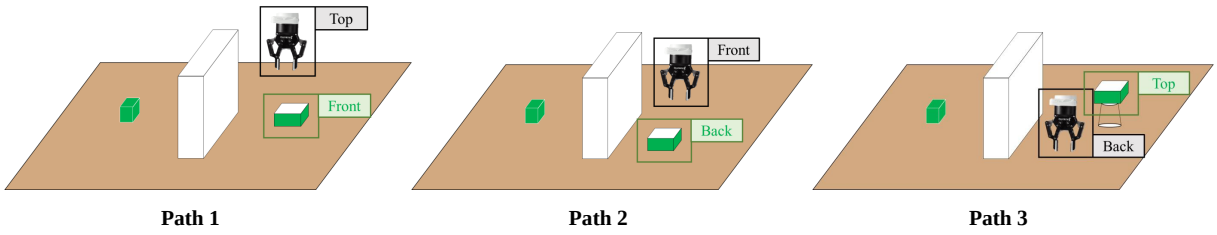



Figure 4.28: Workspace configurations for the pick and place experimental task.

**Experiment Procedure —** Participants were trained on the baseline control interface and three levels of assistive autonomy before the experiment. The experimenter then provided video instructions to demonstrate each AR feature and gather their preferences (Preference 1) for each level of autonomy. Participants performed a single trial of the task in Path 1 using their selected AR visual cues for all control modes. Following the completion of the task in Path 1, participants were given the opportunity to switch their preferences (Preference 2) based on their hands-on experience and then perform the same task again with Paths 2 and 3 for one trial each. Lastly, participants were required to perform the same task using the AR feature proposed by the experienced user for Paths 2 and 3 for one trial each and made their final selection of AR preferences (Preference 3). Note that a trial was skipped if the preferred AR combination was the same between Preference 1, 2, and 3.

**Data Collection and Analysis —** To assess control efficiency, we recorded the complete length of the trajectory covered by the handheld controller. Previous research has shown that the size of the pupil increases as the level of stress rises [398]. We utilized pupil diameter as a measure for estimating subject-specific cognitive workload. To establish a baseline for each participant, we instructed them to look at a blank screen for 30 seconds, assuming a stress-free state. The cognitive workload was calculated by finding the average difference between real-time pupil diameter and baseline value, which was then normalized by the maximum average distance between real-time pupil diameter and baseline value across all trials for each participant. For the comparisons in control efficiency and cognitive workload, we first used F-test to compare the variances of the two sample groups and then used Student's t-test (equal variance) and Welch's t-test (unequal variance) with $p < .05$.

### 4.5.4 AR Preferences for Each Level of Autonomy

Note that preferences 1, 2, and 3 are referred to as P1, P2, and P3. Figure 4.29.a presents participants' final selection (after using the recommendation of experienced users, refer to as P3) of each AR visual cue for different levels of robot autonomy.

*Direct Control.* Most of the participants (13 out of 18) selected local information (arrow) as a preferred way to continuously guide their motion to perform the task and avoid an obstacle. All the participants (18 out of 18) preferred having detailed (planes of the obstacle) information indicating possible collisions and specific (target affordance) methods to indicate if the position of the robot end-effector is good to perform the precise manipulation.

*Assisted Control.* Most of the participants (15 out of 18) still selected local information (arrow) as a preferred way to guide their motion, especially in the aspect of explicitly showing the required control direction while the assisted autonomy is activated. Half of the participants preferred no-

tification of the activation of the autonomy with a low salience (light-up a small square) method while using global information (path) to be informed of the autonomy intention had been selected by most of the participants (14 out of 18). Most of the participants (11 out of 18) preferred having a general obstacle indication (highlight with a red boundary) that potentially provides a reason for the activation of assisted autonomy to avoid an obstacle. Most of the participants (11 out of 18), however, preferred not to have AR visual cues for precise manipulation with grasping and placing an object because they feel assisted autonomy will handle it and like to have minimal visual clutter on the screen.

*Shared Control.* In contrast to the direct and assisted control, most of the participants (12 out of 18) selected global information (path) as a preferred AR visual cue to guide their motion to help approach the autonomy zone around the targets and the obstacle. Almost all the participants (17 out of 18) preferred the most salience (highlight the entire user interface with color and two thick bars) method to indicate the activation of autonomy and global information (path) to indicate the intent of autonomy. This way the participants could get a better sense of the timing of resuming control when the autonomy was completed. Most of the participants (13 out of 18) preferred to have no AR visual cue for obstacle indication as they feel the autonomy will handle it automatically, and use general information (highlight the current target) to ensure the autonomy is being applied to the correct target.

*Supervisory Control.* Most of the participants (13 and 16 out of 18 participants respectively) preferred having a medium salience (pop-up text) to indicate autonomy activation and global information (path) to demonstrate the autonomy intention. No AR visual cues for obstacle indication and target hint were selected as they were deemed not necessary by most of the participants (15 and 14 out of 18) as these will be handled by robot autonomy.

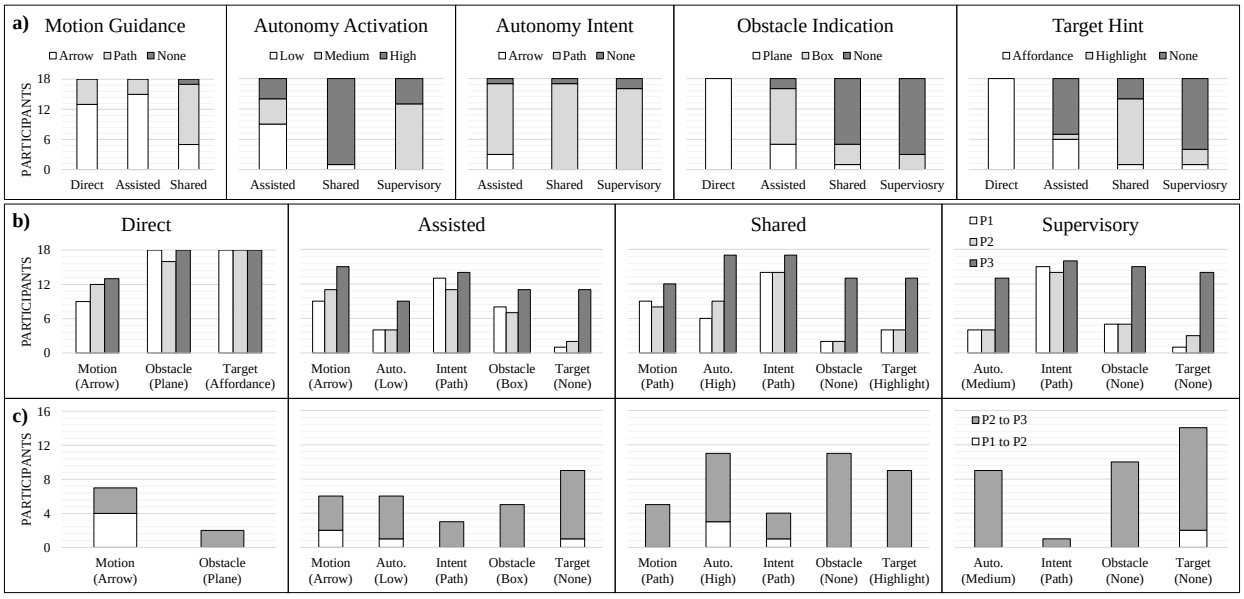



Figure 4.29: a) The final selections of the participants for the different AR features for all the control modes; b) The number of participants who selected the recommended AR feature for all the control modes; c) The number of participants who changed their preferences from other AR features to the recommended AR feature when moving from P1-2 and P2-3.

### 4.5.5 Influence of Interface Learning Method

Figure 4.29.b shows the users' preference changes for each AR feature suggested by the experienced user from P1 to P3.

*Video Instruction-Based (Initial Selection).* In **direct control** mode, half the participants selected local and half the participants selected global information as their preferred motion guidance indication. All the participants already preferred having detailed information (planes of possible collision with the obstacle) for avoiding the obstacle and grasp/place affordance to assist precise manipulation. In **assisted control** mode, half the participants selected local and half the participants selected global information as their preferred motion guidance indication. Few participants selected the low salience (light-up small square) method to be informed of the autonomy activation while most of the participants already chose the global information (planned path) to show the autonomy intention. Similar to direct control, most of the participants still preferred having detailed

information (planes) over the general method (highlight obstacle boundary) to avoid the obstacle. Most of the participants chose grasp/place affordance to assist precise manipulation even if assisted autonomy was provided. In **shared control** mode, half the participants selected local and half the participants selected global information as their preferred motion guidance indication. Few of the participants selected the high salience (highlight the entire screen with bars) method to show the autonomy activation while most of the participants already chose the global information (planned path) to show the autonomy intention. Most of the participants selected the general information for obstacle avoidance, over having no AR visual cues, and grasp/place affordance to assist precise manipulation. In **supervisory control** mode, most of the participants preferred having the most salience method to indicate the activation of the autonomy while most of the participants already chose the global information (planned path) to show the autonomy intention. Similar to direct control, most of the participants selected the general information for obstacle avoidance over having no AR visual cues and grasp/place affordance to assist precise manipulation even the autonomy will handle both obstacle avoidance and precise manipulation.

*Hands-On Engagement-Based (Intermediate Selection).* In **direct control** mode, a few participants changed their preferences for motion guidance from global information to local information in the form of arrows. The preferences for obstacle collision, and target grasping and placing largely remained the same. For **assisted control** and **shared control** mode, the preferences for the AR features for motion guidance, obstacle collision formation, target grasping/placing, and autonomy indication and intent generally remained the same. This trend continued for **supervisory control** mode as well with minimal changes in the selections for the AR features for obstacle collision information and autonomy indication and intent from the previous selections.

*Expert Recommendation-Based (Final Selection)* — For **direct control** mode, minimal changes happened across all the AR features between the intermediate selection and the final selection with most of the participants selecting the local information for motion guidance and all the participants

wanting detailed information for obstacle collision (planes) and target grasp/place (affordance). For the **assisted control** mode, most of the participants selected local information in the form of arrows for motion guidance with some participants changing their preferences from global information in the form of path AR. Half the participants selected the low salience light notification for autonomy indication with a few participants changing their choices from high salience to low salience when moving from intermediate to final selection. Most participants continued to select the path AR to provide global information about the intent of autonomy. For information about a collision with the obstacle, most participants selected general information in the form of box AR. Finally, for the target hint, most participants chose to have no hint to help with target grasp/place with nearly half the participants changing their intermediate selections. Most of the participants selected global information for motion guidance while making their final selection for **shared control** mode, with some participants changing their selections from local information for motion guidance. Nearly all the participants selected the high salience autonomy indication with nearly half the participants changing their selections from the intermediate selection. Similar to assisted control, the selections for autonomy intent remained largely unchanged, with nearly all the participants preferring to learn about robot autonomy intent through the global information provided by path AR. More than half the participants changed their preference for obstacle information, with most of the participants now preferring no information about the obstacle collision. Most of the participants prefer to highlight the target while grasping or placing with half the participants changing their preference from the previous intermediate selection. For **supervisory control** mode, most of the participants selected the medium salience text indication for autonomy indication with half the participants changing their selection to medium salience. Similar to both assisted and shared control, most participants selected the global information via path AR for autonomy intent with minimal participants changing their preferences. Most of the participants preferred no information about collisions with the obstacle with nearly half the participants changing their preferences from other AR features. Finally, most of the participants now preferred no information regarding the target with most of the

participants changing their preferences from the intermediate selection.

## 4.6 Summary and Outlook

In this chapter, we first proposed a generalizable sensory feedback design to assist dexterous telemanipulation tasks. The designs well-supported the findings in the human factor experiment (Chapter 3) that investigated the visual and haptic sensory integration in the usage of active telepresence cameras for general-purpose manipulation. We also conducted a systematic evaluation to compare the sensory augmentation using haptic and AR visual cues and investigate how the users' preference of sensory modality may be affected by the secondary tasks that introduced different types of cognitive workload. Throughout the user studies, we found that there have been advantages in the use of haptic and AR visual cues for sensory integration. However, despite good feasibility and transparency, the workload arising from precise manipulation still poses a challenge. On the other hand, although shared autonomy has been proposed as a solution to reduce operational effort (Chapter 2), it falls short in terms of communication between humans and robots. We then proposed a high-level terminology for these types of assistance, namely perception and action augmentation that should be considered simultaneously for better results. The objective is to investigate the best manner in which augmentation can be used or combined that improves transparency and reduces the effort for remote manipulation via the motion mapping interface. Based on the user study, it was found that most participants preferred a combination of perception and action augmentation. Specifically, participants favored a shared autonomy control approach that incorporated AR visual cues to indicate when autonomy was activated and the usable area. Additionally, they appreciated the scaled motion feature which slowed down the robot's actions within the usable area, providing a damping effect to avoid moving out of the usable area. To this end, the integration of the perception and action assistance augmented the feasible interface with increased transparency and

reduced control effort. Lastly in this chapter, we investigate the generalizability of the proposed combination of robot autonomy and AR visual cues by providing various options for AR features for different levels of assistive autonomy.

**Benefits of Multi-sensory Feedback —** In terms of objective task performance and subjective workload evaluation, our results demonstrate the potential benefit of haptic and AR visual cues to represent the information critical to the tele-manipulation task. Participants using haptic feedback significantly outperformed the baseline interface when the robot started from the far-away condition, showing a 27% decrease in completion time. While the AR visual cues also significantly outperformed the baseline interface when the robot started from the out-of-view condition, showing a 32% decrease in time taken. This finding implies the desired autonomy in switching the mechanism of sensory feedback based on the visibility of the robot platform in the remote user interface. Whereas for the robot starting from the close condition and the rest of the comparison not mentioned before, our work does not support a significant difference between the haptic, AR visual cues, and baseline without any sensory feedback. It is possible that, for the close condition, the distance between the robot end-effector and the target object is not enough to exaggerate the significant difference but the haptic cues still trended in the lowest completion time. As the task performance has been improved by introducing the haptic and AR visual cues, our subjective evaluation results show a comparable reduction in operational workload and an addition to system usability and transparency.

We implemented the haptic and AR cues upon the baseline motion tracking teleoperation interface to communicate four critical information of tele-manipulation tasks. We categorized the information into two groups, the information needs the user to take immediate action (e.g., environment constraints, grasp confirmation), and the information to serve as the status monitoring (e.g., direction and distance to target, the affordance of grasping). Participants using haptic feedback demonstrated the lowest accumulated occurrences of table collision across all the robot starting

locations which was also supported by their subjective preference. Participants also reported that using haptic feedback can increase the awareness of the sense that they grasp the object. While the participants using AR visual cues significantly reduced the length of the robot's trajectory and the occurrences of hitting the object. Many participants reported that the AR visual cues transfer the alignment of the grasping to a game-like task which simplifies the precise manipulation. The findings helped us to conclude a philosophy for sensory augmentation design: the haptic/vibrotactile feedback may encode the information that requires prompt attention and AR visual feedback may present the continuous system status.

In practical healthcare settings, nurses often need to handle multiple requests and tasks simultaneously, which increases their cognitive workload. To explore how the preference for sensory feedback is affected by different types of workloads, we conducted a study in which we introduced three secondary tasks that simulate the additional workload that nursing applications may entail: critical thinking (decision-making), haptic monitoring (other medical devices or emergency on-call), and visual monitoring (patient vital signs). Our post-study preference survey revealed that participants' sensory feedback preferences shifted toward options that required less cognitive effort to integrate with the secondary task. For instance, more participants preferred augmented reality (AR) visual cues over haptic feedback when performing the haptic monitoring secondary task, and vice versa. Many participants explicitly mentioned that using the same sensory feedback for both the primary assistance and secondary task could lead to confusion and a higher mental workload. These results highlight the importance of implementing multi-sensory feedback systems that can adapt to the various types of secondary tasks that nursing workers may encounter. By tailoring the sensory feedback to minimize cognitive load, we can reduce stress and fatigue on nurses, ultimately improving their performance and the quality of patient care.

**The Effective Integration of Perception and Action Augmentation —** Related work in the literature has developed various augmented reality interfaces and shared autonomy for action supports,

in order to assist remote perception and robot motion control [411, 412, 413], yet there are still no comprehensive comparisons to evaluate their individual and integrated impacts on task performance, workload, and user preference. To fill this gap, we first conducted a user study to compare interfaces that only have the perception assistance (of AR visual cues) *or* the action assistance (of autonomous actions), to the baseline interface without any assistance. Our results show that both perception and action assistance can effectively reduce cognitive workload. Moreover, assistive action can effectively reduce the occurrence of errors.

We further evaluated to what extent the additional perception and augmentations proposed in this chapter can further assist the tele-manipulation. Our results show that the effectiveness of perception and action assistance (and their integration) depends on the task performance objective. Specifically, we found that: for the tasks that need to be completed as fast as possible, the integration of Fixed PIP display and autonomous actions is the most effective, because it led to the least task completion time and motion efficiency (measured by the total robot trajectory length). As the participants commented: "... *the complementary viewpoint helped me to clearly understand the relationship between robot and target so that can move faster in the right direction to active the autonomy feature*". For the task that emphasizes the reliability and precision of motion control, the interface that only provides autonomous actions for assistance turns out to be most effective, because the operators can focus more on the use of autonomous actions and will not be distracted by other interface augmentation designed to enhance task efficiency or to reduce the workload. However, when the human workload and comfort are prioritized in the tele-manipulation, it is more effective to provide only the AR visual cues and assistive autonomy integrated with Fixed PIP interfaces, as they can effectively reduce human's physical and cognitive workload, respectively. These findings suggest the extension of the task- or goal-dependent perception and augment assistance design, which will be intelligent to not only provide suitable augmentation but activate the assistance based on the online estimation of human comfort. From the user preference and SUS, we

also found that humans strongly prefer to use autonomy to assist their remote perception and action, but *only if they are reliable*. Most of the participants commented that: "*if the robot autonomy is reliable, I would like to use it on daily bases*". Particularly, the nursing participants "*... would like the robot to be as autonomous as possible because we do not have more bandwidth to control the robot during nursing duty*". However, the robot autonomy may not be consistently reliable due to the perception and action uncertainty of the robots, as well as the complexity of the manipulation tasks. It is still unclear how to adjust the level and type of robot assistance if the reliability of the robot autonomy may vary. Our future work therefore will further investigate the impacts of unreliable autonomy in human-robot collaboration for robot remote manipulation.

**AR Features for Different Levels of Robot Autonomy —** As the level of autonomy transitions from direct to supervisory control, we found that: (1) humans' priority for AR visual cues shifts ***from guiding robot control to communicating autonomy activation and intention*** based on their subjective feedback in Table 4.16; (2) humans' preference for AR visual cues changes ***from providing local information to offering global guidance in robot control*** as indicated by a large portion of participants (13 and 12 out of 18) selecting arrow for direct control and the path for shared control. However, the use of ***global information to display the autonomy's intention*** remains consistent across all interfaces with autonomy, as more than 14 participants preferred having a planned path to display it; (3) the ***efficacy of the AR features that share the same purpose as the robot autonomy is decreased*** as observed by most of the participants preferring not to have any AR visual cues for both obstacle indication and target hint in supervisory control mode.

Table 4.16: Usefulness of AR features for each control mode.

| **Interface** | **Priority** |
|---|---|
| Direct | Motion Guidance>Obstacle>Target |
| Assisted | Motion Guidance>Autonomy>Intent>Target>Obstacle |
| Shared | Autonomy>Motion Guidance>Intent>Target>Obstacle |
| Supervisory | Intent>Autonomy>Target>Obstacle |

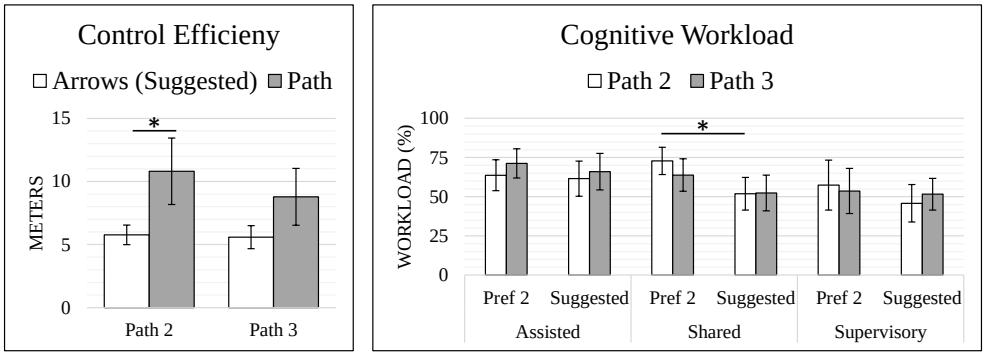



Figure 4.30: Control efficiency evaluated by the handheld controller's trajectory in direct control and cognitive workload.

Regarding the influences of how participants learn to use the AR visual cues, we found that ***participants' preference for AR visual cues converged to the recommendation of experienced users***, which is observed by a large change after hands-on robot control using suggested AR features (Figure 4.29.c) and most of the participants selecting the suggested AR features as their final choice across all control modes. Additionally, the participants commented: *"...would like to have clear guidance on how I should move the robot when not much robot autonomy is available."* which was supported by the total trajectory lengths of the handheld controller being significantly shorter ($p < .05$) while using the suggested AR visual cue (arrows) in direct control mode (Figure 4.30). The participants also commented: *"...the most obvious way to inform the activation of the autonomy will be preferred for shared control so that I do not need to put too much effort when the autonomy is on."* and this was supported by the significantly lower cognitive workload ($p < .05$) while using the suggested AR visual cues in shared control (Figure 4.30). We also found that ***video instruction and hands-on practice tend to provide sufficient information for the selection of AR visual cues in direct control without autonomy*** while experience and proficiency play a role in selecting suitable AR features when various autonomy is available. This is supported by the observation that participants' final preference for AR visual cues remained consistent with their initial selection when using direct control, but there were notable differences when using autonomy.

# Chapter 5

# Reliability of Robot Autonomy

## 5.1 Motivation

Teleoperation via human motion tracking interfaces (e.g., motion capture systems, exoskeletons, hand-held controllers) enables humans to efficiently and intuitively control remote manipulator robots to perform dexterous, freeform manipulations. To further reduce the operator's workload, robot autonomy has been leveraged to provide perception and action assistance, including the autonomy to detect and recognize objects, infer human intents, and motion planning and control. The assistive autonomy, which may vary in the levels of assistance (e.g., from shared to supervisory control) and types of assistance (e.g., for perception, decision, and action), is expected to enhance human-robot collaboration in remote robot control. However, the robot autonomy may not be consistently reliable due to the perception and action uncertainty of the robots, as well as the complexity of the manipulation tasks. It is still unclear *how to adjust the level and type of robot assistance if the reliability of the robot autonomy may vary*.

This chapter aims to investigate how to adjust human-robot collaboration when robot autonomy is unreliable to different extents (with respect to the task). We focus on the dexterous manipulation tasks that rely on the general-purpose gross and precise manipulation actions for approaching, moving, grasping, and placing objects. The tasks are unstructured and may involve the manipulation of unknown objects. They require humans to plan and control at least at action levels, e.g., by moving the end-effector close enough to the object to be manipulated during the task or selecting the

sequence of objects and manipulation actions to be executed later under supervisory control. They may also require humans to detect errors (e.g., due to the incorrect choice of object or action, low precision of autonomous action execution) using the remote camera visual feedback, and correct them using the control at the action or motion level.

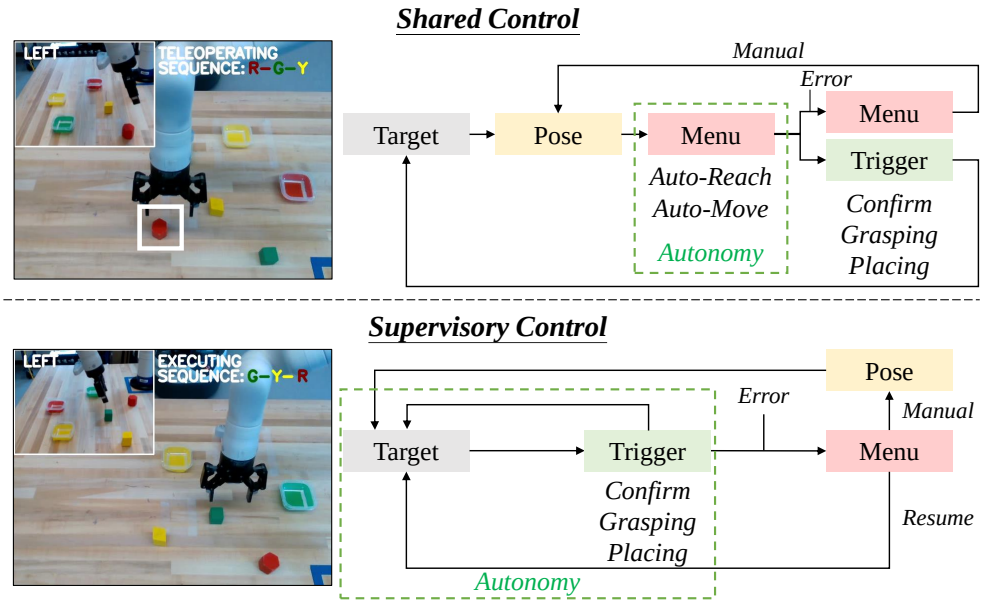



Figure 5.1: Shared and supervisory control paradigms for assisted remote robot manipulation, and how the errors were introduced.

We developed two human-robot collaboration (HRC) paradigms to effectively assist humans to control robots to perform these dexterous manipulation tasks using a motion tracking interface. (1) The **shared control** paradigm allows the robot autonomy to control the precise manipulation actions but relies on humans for the gross manipulation control and error correction; (2) The **supervisory control** paradigm allows the robot autonomy to control both the gross and precise manipulation actions and only relies on humans for error correction. The shared control paradigm integrates robot autonomy, including: (1) human goal *intent inference* based on human gaze, robot status, and task states; (2) *autonomous actions* for the precise manipulation which tends to cause high cognitive and physical workload for human control (e.g., object grasping and placing actions [9]). Our proposed HRC paradigms were developed to be *generalizable* to incorporate other task-dependent assistive

autonomy with intent inference, autonomous actions, and the methods for the estimation of human cognitive workload and muscle efforts (e.g., based on gaze or motion tracking). We conducted a user study to evaluate the effectiveness of the proposed HRC paradigms when the autonomy is reliable (User Study I). We further investigated how the unreliable assistive autonomy (User Study II) that results in errors of different types and frequencies during the tasks may influence the performance, workload, human preference, and usage of autonomy between the two HRC paradigms. Shown in the control flow in Figure 5.1, the errors may happen to the autonomous actions triggered by the operator, at the action level (e.g., picking up the wrong object) or the motion level (e.g., missing to grasp an object). The operator thus needs to switch to manual control to correct the error.

## 5.2 Literature Review

**Autonomy to Assist Motion Tracking Teleoperation —** Human motion tracking interfaces enable human operators to control the dexterous manipulation of remote robots (e.g., manipulators [414], mobile manipulators [415] and humanoid robots [1]) using the natural motion coordination of their body, arms, and hands. These motion tracking interfaces include various motion capture systems (e.g., vision-based [416] vs IMU-based [417]), stylus/joysticks (e.g., stylus/joysticks [1]), exoskeletons (e.g., soft [418] and rigid [419], passive [420] vs actuated [421]), virtual reality gaming systems (e.g., with hand-held controller and body trackers [34, 380]), and the customized integration of multimodal control interfaces (e.g., [422]). Although these interfaces are more efficient and intuitive for dexterous manipulation control than the alternatives, human operators still need assistive autonomy to enhance their control precision and reduce their control workload. Among all human-robot collaboration paradigms (see the review on the levels of autonomy in [63]), shared autonomy [9] and supervisory control [423] are both proven to be effective for handling the more complex, un-

structured and error-prone manipulation tasks. However, during a comprehensive manipulation, the operator's preference for the level of assistive autonomy may vary with the reliability of the autonomy with respect to "the current sub-task" of the dexterous manipulation task. Investigating how this preference varies with the reliability of the autonomy will enable the design to adjust the human-robot collaboration paradigms (more shared or supervisory control), and the types and levels of assistive autonomy, in order to provide the best possible assistance to the operator.

**Causes and Effects of Unreliable Autonomy —** Related work in the literature has analyzed the causes and effects of the failures in human-robot interactions and how to mitigate possible negative impacts (see the review in [29]). In general, robot failures may differ in their functional severity, social severity, relevance to general or specific robot systems, frequency, condition (when the failure happens), and symptoms (that indicate the failure). The failures of robots may affect task performance (e.g., task completion time), human workload and comfort, and human perception of robot intelligence, transparency, safety, and influence human's trust, satisfaction, impression, and attitude toward robots. When failures happen, robots are preferred to communicate the errors to help humans to better perceive and comprehend the failures, and to leverage human help to resolve the failures. In this chapter, we focus on the failures common to general-purpose robot manipulation tasks, and common to both shared or supervisory control. At high-level, the robot may apply the wrong action to the wrong object, due to errors in the prediction of goal or action intent [424], or detection of object-action affordance [25]. At low-level, the robot may not successfully perform the manipulation motions (e.g., missing to grasp or place an object) due to errors in perception, motion planning, and execution. Thus far, it is still unclear how the types and frequency of errors may affect human performance, workload, perception, and preference for the level of autonomy for assistance. This chapter will fill the gap and inform the adaptive shared autonomy for dexterous manipulation.

## 5.3 Human-Robot Collaboration Paradigms

**Remote Manipulation System —** Figure 5.2 shows the tele-manipulation system we integrated to perform the pick-and-place task. We used the hand-held controller of the HTC virtual reality system to track human hand motion to control a 7-DOF (Degrees of freedom) Kinova Gen 3 robotic manipulator with a two-fingered Robotiq gripper. The scaling of human-to-robot motion mapping is 5:3:3 for the linear velocity in x-, y- and z-axis. We constrained the robot's rotational motions because this chapter focuses on investigating the impact of unreliable robot autonomy instead of the controllability of teleoperation. A desktop monitor displayed a graphical user interface (GUI) of Unity 3D window (1440 $\times$ 1080 pixel) to stream the video from the workspace cameras (back and side views, using picture-in-picture display to trivialize the impact of the loss of depth information) at 30 Hz frame rate. The GUI also used overlay text to indicate the robot status ("TELEOPERATING", "EXECUTING", "PAUSED") and sequence of objects to operate.

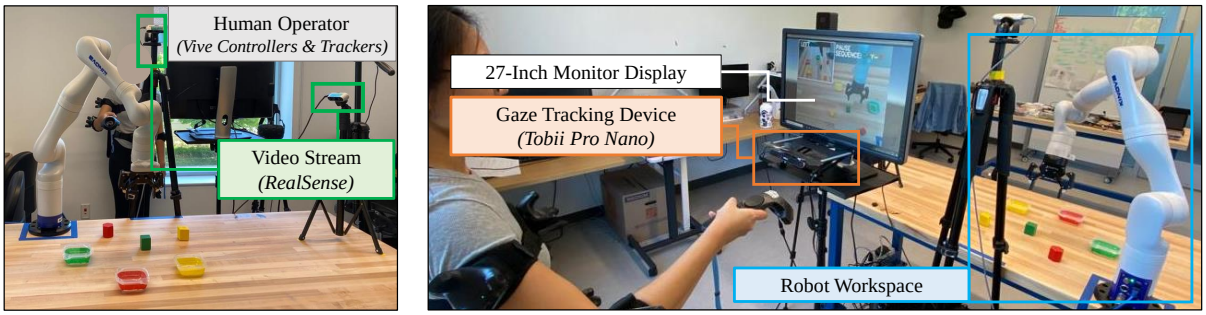



Figure 5.2: Assisted Tele-manipulation System.

**Human-Robot Collaboration Paradigms —** Table 5.1 shows the three HRC paradigms we developed for remote robot control based on task allocation between human and robot for sensing the environment, making action decisions, and executing the planned motion [63, 425].

*Shared Control.* In ***shared control*** mode, humans can use hand motions to control the robot to reach the target object and to move it close to the desired location and can trigger a button to control

Table 5.1: Human-Robot Collaboration Paradigms.

| **Autonomy** *(Level)* | **Perception** *(Option)* | **Decision** *(Selection)* | **Motion Planning** *(Action)* |
|---|---|---|---|
| Manual | Human | Human | Human |
| Shared | Human/Robot | Human | Robot/Human |
| Supervisory | Robot | Robot/Human | Robot |

the robot's actions to precisely grasp and place an object. This HRC paradigm, developed and evaluated in our prior work in [9], enables humans and robots to complement each other's skills and strength to perform unstructured remote manipulation tasks: while humans can intuitively and efficiently control the robot's freeform gross manipulation motions to move the robot freely across the cluttered workspace and approach the targets. The robot autonomy will perform precise manipulation actions that cause significant human workload, based on inferred human goals and action intent. We infer human goals and action intents, by tracking human gaze fixation (using Tobii Pro Nano eye tracking device) and robot states (i.e., distances to each object and container in the workspace, the opening and closing of the gripper). When the end-effector is close enough to the target object to pick or location to place (within 50 mm for our task), the robot will determine the appropriate autonomous action and execute it after the operator triggers a button on the hand-held controller to approve autonomous action. We trivialize object detection by pre-defining the object and container locations. To correct the errors, humans can press the menu button on the controller to undo a completed action or switch to ***manual control*** mode to control the robot's motions directly.

*Supervisory Control.* In the ***supervisory control*** mode, the robot performs all aspects of the task which autonomously picks and places the object following a pre-planned sequence based on the general procedure to perform this type of task. The human operator who supervised the robot can confirm the robot's selected actions if they are appropriate, or control the robot's actions and motions to correct any errors.

## 5.4 Unreliability of Robot Autonomy

We manipulated the reliability of the robot autonomy by introducing the errors at the action- and motion-level that are common to manipulation tasks (see Figure 5.3 (Bottom)). When a ***sequence error*** happens, robot autonomy may pick up a wrong object (given the pre-defined object manipulation sequence) or place it into a container that does not match its color. When a ***grasp/place error*** happens, robot autonomy may miss to grasp an object or place the object with an offset from the desired location. We also manipulated the frequency of errors that occur at least once and up to three in six actions (three actions for both grasping and placing).

## 5.5 User Study

We conduct a user study to evaluate: 1) When the autonomy is reliable, the supervisory control interface will have the best task performance, the lowest workload, and the highest user preference, even with a loss of engagement; 2) When the robot autonomy is unreliable, a higher error frequency will lead to worse performance and higher workload, and increases the operator's preference and usage of a lower level autonomy; 3) When the robot autonomy is unreliable, users will not lose the preference for using the robot autonomy if the effort to correct the error is lower.

**Participants and Tasks —** We conducted two user studies (with the same 13 participants, 10 males, 3 females, age = 26$\pm$4) to investigate the effectiveness of the two proposed HRC paradigms and evaluate the influence of different types and frequencies of the error on the tasks and human operators. The experimental protocol was approved by WPI's Institutional Review Board (IRB-21-0004). As shown in Figure 5.3 (Top), participants were required to perform a general-purpose multi-object manipulation task to collect objects and place them in the correspondingly colored box in the sequence of green-yellow-red (User Study I) and red-green-yellow (User Study II). Note that

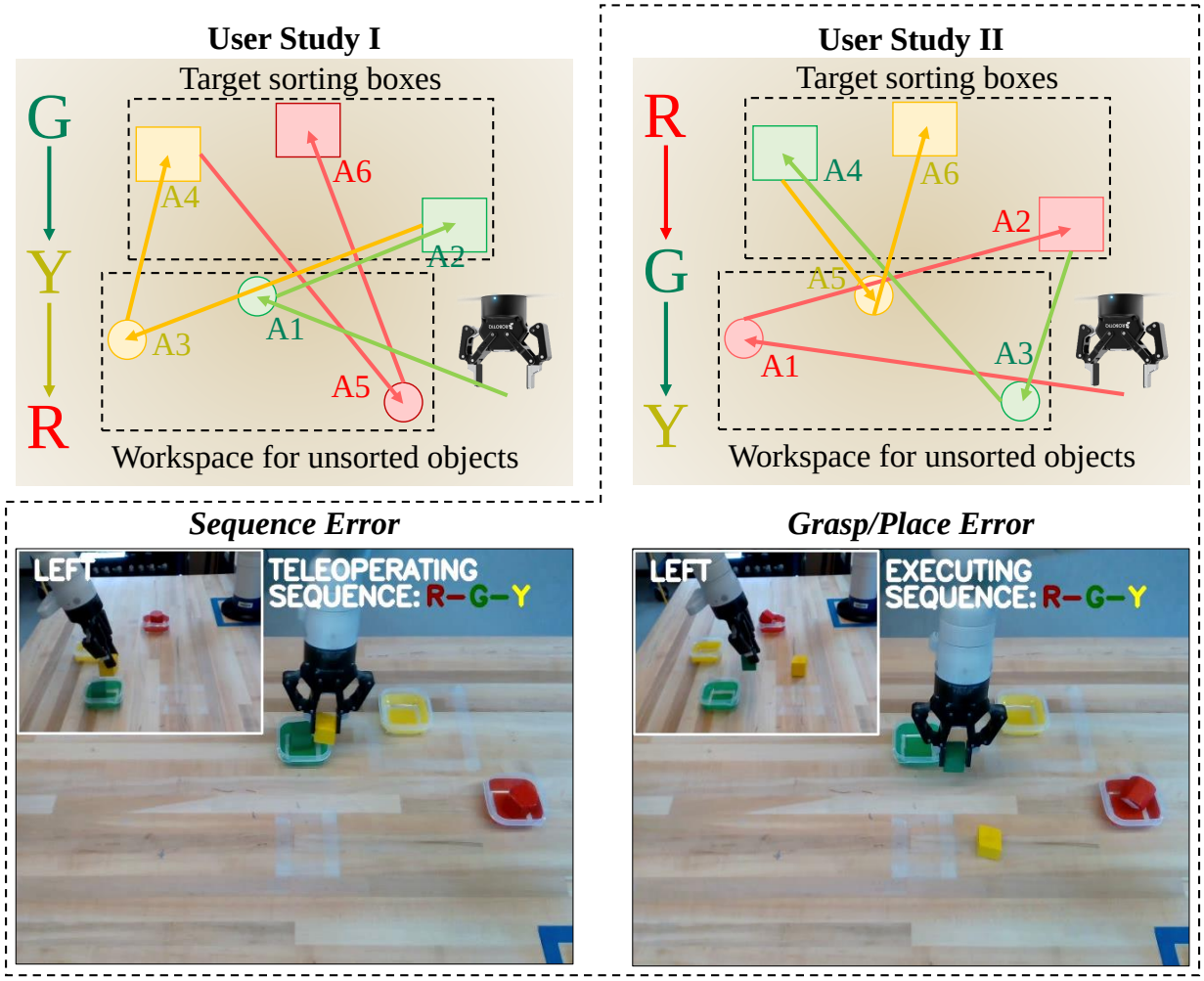



Figure 5.3: (Top) Task and action sequence in each user studies; (Bottom) Sequence and grasp/place errors.

each object in the workspace requires 2 actions: reach-and-grasp (i.e., A1, A3, A5) and move-to-place (i.e., A2, A4, A6).

**Experimental Procedure —** Before the user studies, the experimenter explained and demonstrated how to control a robot manipulator using the motion tracking controller and how to trigger robot autonomy. The participants practiced single object pick-and-place with manual, shared, and supervisory control for maximally 10 minutes. The participants were then asked to look at a blank screen for 30 seconds and had their pupil diameters recorded for the calibration required to esti-

mate their cognitive workload. After User Study I and II, the participants answered generic surveys (the NASA-TLX and System Usability Scale) and reported their preferred control interfaces via a customized questionnaire.

*User Study I.* In the first user study, participants performed a multi-object sorting task, in which they controlled the remote robot manipulator to grasp and place the objects in the box following the green-yellow-red color sequence. The order of the interfaces was randomized of manual, shared, and supervisory control. Note that the autonomy in this user study is reliable without errors. The participants performed a total of 6 trials (3 interfaces $\times$ 2 repetitions). This user study aims to investigate the effectiveness of the two HRC paradigms (shared and supervisory control) and build up a sense of how robot autonomy influences the task.

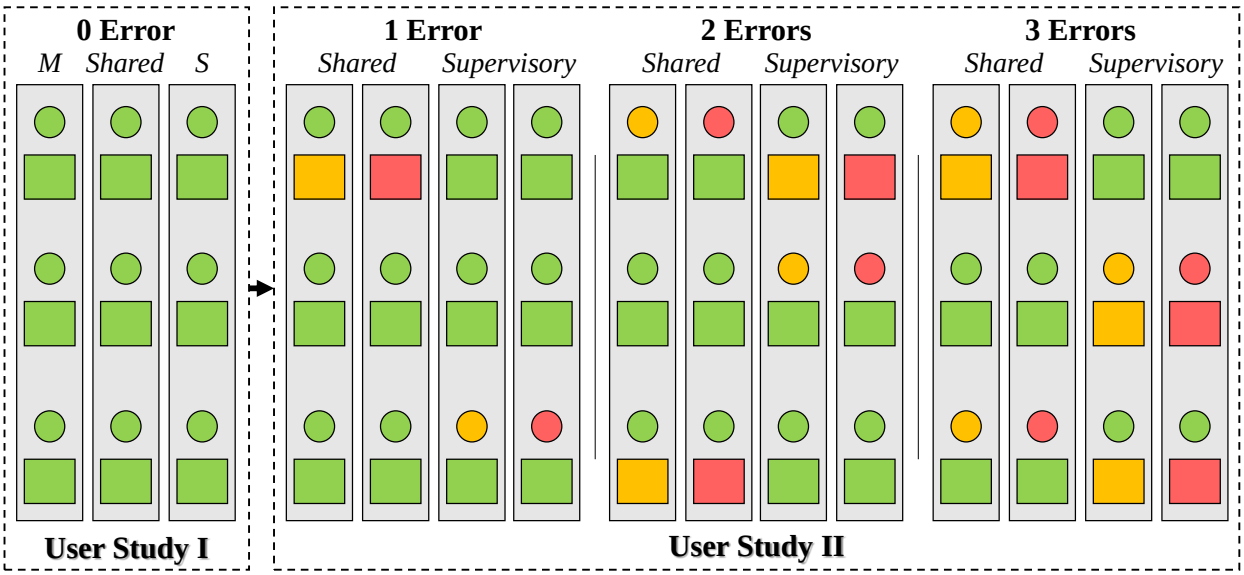



Figure 5.4: Experimental conditions for User Study I and II. The circles and squares represent the reach-to-grasp actions and the move-to-place actions respectively. Yellow (Red) highlights denote the sequence (grasp/place) errors, which may occur 1 to 3 times per trial.

*User Study II.* We further manipulated the reliability of robot autonomy by varying the type and frequency of the errors and evaluated their influence on the tasks and human operators. Participants performed the same multi-object sorting task in a red-green-yellow color sequence by using

the randomized order of the shared and supervisory control. Shown in Figure 5.4, we implemented at least one and up to three errors in 6 actions for both the sequence (yellow marks) and grasp/place type errors (red marks). The participants are allowed to correct the errors only before the confirmation of the grasping and placing action. If the participant grasped the wrong object and failed to correct a sequence error, the subsequent placing box should be the same color as this object, and the new color sequence would be reassigned based on the priority of the red-green-yellow sequence after placing the object. The participants performed a total of 12 trials (2 interfaces $\times$ 3 frequencies of the error $\times$ 2 types of the error).

**Evaluation Metrics —** In both user studies, we measured the ***task performance*** using task completion time and the length of the trajectory traveled by the robot end-effector (which indicates the motion efficiency). We measured the users' ***utilization*** in the robot autonomy objectively using how many of the tasks the participants completed with autonomy and how many times the participants switched from autonomy to manual control, assuming fewer human interventions indicated more trust.

We also analyzed how participants' behaviors change due to unreliable robot autonomy to estimate their levels of engagement in visual perception and actions to control robots. Shown in Figure 5.5 (Left), we tracked human eye movements using Tobii Pro Nano and calculated the percentage of the task for which the gaze fixation was in the area of interest (i.e., targets and picture-in-picture) to estimate ***visual engagement***. To estimate the ***action engagement*** (level of activity), we tracked the positions of the two handheld controllers and 6 body trackers (Vive Tracker 3.0) attached to the operator's upper arms, forearms, chest and waist to measure the shoulder and elbow joint angles (namely, shoulder abduction $\theta_{SA}$ on the frontal plane, shoulder flexion $\theta_{SF}$ on the sagittal plane and elbow flexion $\theta_{EF}$) by inner product formula. Our prior work [18] shows that: the muscle efforts of the anterior, lateral deltoid, and bicep muscle groups, caused by shoulder flexion, abduction, and elbow flexion, contributes most to the physical workload when human

controls tele-manipulation using their arm and hand motions. Figure 5.5 (Right) shows the gesture demonstrations and the threshold we defined for the low level of activity given the motion range of each joint angle ($0° < \theta_{SA} < 120°$, $0° < \theta_{SF} < 150°$, $0° < \theta_{EF} < 150°$). The feedback from the users indicated that humans tend to have more relaxed arm postures and are less ready for robot control in such arm postures.

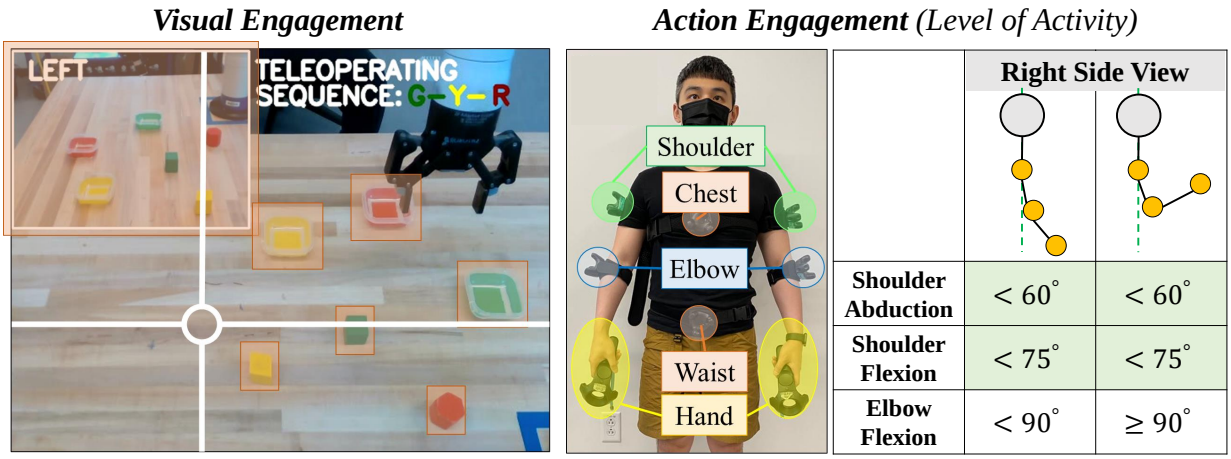



Figure 5.5: Visual engagement and level of activity estimation.

We estimate the ***physical workload*** (i.e., EMG-based muscle efforts) based on human motions mentioned in Chapter 4. Similarly, we also estimated the ***cognitive workload*** caused by stress ($C_{str}$) and error complexity ($C_{err}$) from the operator's pupil diameter, gaze fixation, and movements. We tracked variation in the operator's pupil diameter and estimated the cognitive workload caused by stress as the difference between average pupil diameter during a task ($D_{tsk}$) and the operator's calibrated pupil diameter ($D_{cal}$) calculated before the task's start. Pupil diameter is expected to increase with the increase in stress [396]. This result was normalized with respect to the maximum cognitive workload across all the trials for that subject, such that $C_{str} = (\overline{D_{tsk}} - D_{cal}) / \max_{p=p_1,\ldots,p_n} (\overline{D_{tsk}} - D_{cal})$. The cognitive workload caused by error complexity ($C_{err}$) was computed as the ratio between the average distance in pixels of the operator's gaze fixation motion ($S_{tsk}$) and the maximum distance of fixation motion across all the trials for the participant

($S_{max}$). Thus, the error complexity can be calculated as $C_{err} = \overline{S_{tsk}}/S_{max}$. Complex errors are expected to result in greater gaze motion distances as they are assumed to use other visual cues to compensate. We also calculate the overall workload ($C_{task}$) of the entire task as the average of the workloads caused by stress ($C_{str}$) and error complexity ($C_{err}$), assuming their equal contributions.

For each user study, we collected the ***subjective feedback*** from the participants using a NASA-TLX questionnaire on a scale from 1 to 20 and the System Usability Scale (SUS) survey on a scale of 0 to 100. The NASA-TLX score is calculated as the overall workload by weighting six sub-scales (mental demand=5, physical demand=4, temporal demand=0, performance=2, effort=3, frustration=1). The weighting coefficients were generated by choosing from a series of pairs of rating scale factors that were deemed to be important based on the official instructions. Participants also answered our customized questionnaire on the preferred interfaces considering different factors (i.e., reliability of robot autonomy, frequency, and type of error) at the end of the user study.

For all the comparisons, we analyzed data from all evaluation metrics using one-way repeated-measures analysis variance (ANOVA), including HRC paradigms for user study I, and error frequency and type for user study II, as a within-participants variable. All pairwise comparisons used Holm-Bonferroni correction to control for Type I error in multiple comparisons.

## 5.6 Impacts of Unreliable Autonomy

**Effectiveness of HRC Paradigms —** Figure 5.6 compares the performance (completion time and trajectory), workload (physical and cognitive), engagement (visual and action), and subjective feedback (NASA-TLX and SUS) of the manual, reliable shared, and supervisory control. The supervisory control outperformed both shared and manual control with a significantly ($p < .05$) faster task completion time, shorter traveled trajectory, lower physical, cognitive, and subjective overall workload, and higher system usability score. The supervisory control has significantly ($p < .05$)

lowest percentage of engagement than shared and manual control which implied the participants prefer the high level of robot autonomy if guaranteed reliability. The preference results (Figure 5.9) also supported this observation that 12 out of 13 participants rated the supervisory control as the most preferred interface.

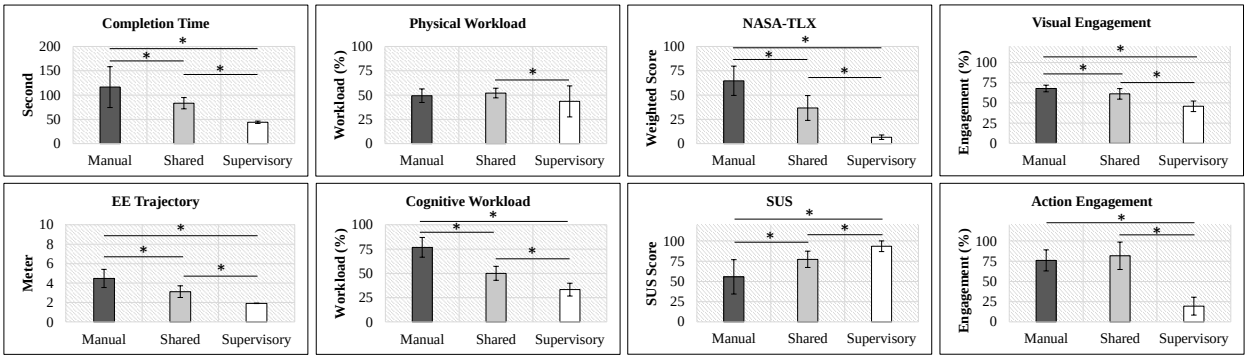



Figure 5.6: User study I: comparison of manual, shared and supervisory control with reliable autonomy.

**Impacts of Error Frequency —** Figure 5.7 indicates significantly ($p < .05$) longer completion time, traveled trajectory, and higher physical workload with the 3 sequence errors condition for both shared and supervisory control and 3 grasp/place errors condition for supervisory control. The black and green lines indicate significant differences between different error frequencies and types of errors, respectively. We found no significant difference in cognitive workload across the different frequency of errors for both shared and supervisory control. In Figure 5.9, 12 out of 13 participants reported they preferred the supervisory control if errors occurred occasionally. However, when errors happened frequently, 5/13 participants prefer manual control and 5/13 participants prefer supervisory control, which indicates that some people still try to use a high level of autonomy even with frequent errors. We further analyzed the correlation between the usage of robot autonomy and manual intervention (Figure 5.8). We noticed that participants tend to let robot autonomy perform most of the actions (5-6 out of 6 actions) and only switch to manual control (1-3 depending on the frequency of the error) to correct the error if necessary while using supervisory control. The correlation for shared control in fewer errors condition is similar to supervisory con-

trol. This observation confirms that the participants might still utilize the high levels of autonomy for reducing operational effort even with frequent errors. However, certain participants tended to give up on the usage of autonomy (use autonomy for 1-2 out of 6 actions) and manually completed the task if errors were frequent. To sum up, we found that the lower error frequency leads to better performance, lower workload, higher preference and usage in high levels of autonomy; however, the user's preference and usage in the lower level of autonomy does not necessarily increase with the error frequency. This might be because operators tend to be more relaxed (minimal workload and engagement) if the tasks are not time sensitive or not extremely unreliable. As the participants commented: *"...would like to use supervisory control to mitigate the control effort unless the success rate is less than 20 percent"* and *"...the supervisory control or higher levels of autonomy interface will be preferred so that I can focus on other duties no matter how many mistakes the robot might cause."*.

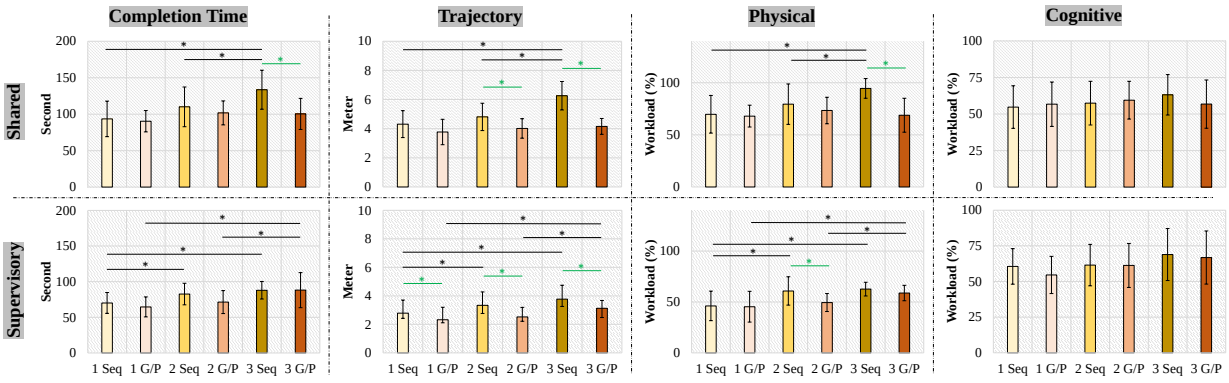



Figure 5.7: User study II: comparison of task completion time, end-effector trajectory, the physical and cognitive workload for shared and supervisory control with different error frequencies and types.

**Impacts of Error Types —** We also compared the impacts of error types across all the error frequencies for both shared and supervisory control. Shown in Figure 5.7, the sequence error results in a significantly ($p < .05$) longer traveled trajectory and higher physical workload for both shared and supervisory control in most of the error frequencies which aligned with our assumption that the effort to correct sequence errors is higher than grasp/place errors. The subjective feedback (Fig-

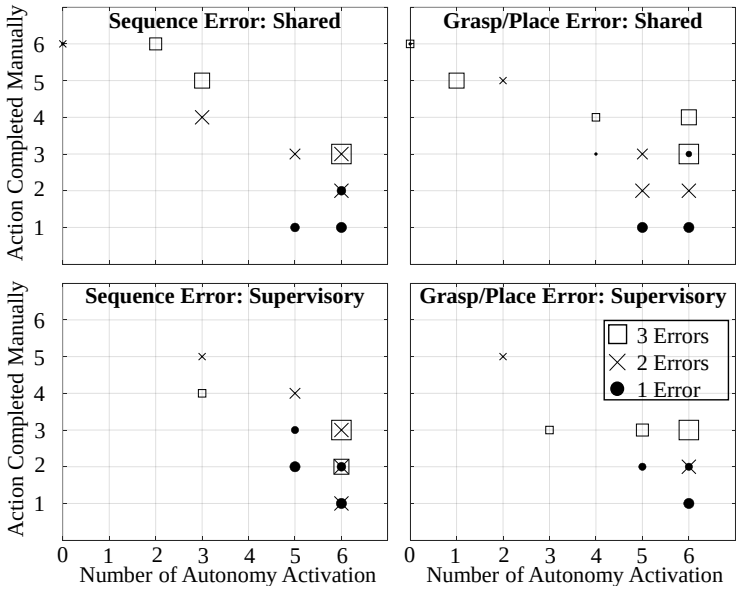



Figure 5.8: User study II: robot autonomy and manual intervention. Marker size indicates the group size.

ure 5.9) also indicated that 12/13 participants have a higher preference in the usage of robot autonomy (shared and supervisory control) if the error is the grasp/place error instead of the sequence error because it takes less effort to correct. This imply users' preference for usage of autonomy could be better retained if the effort for error correction is lower.

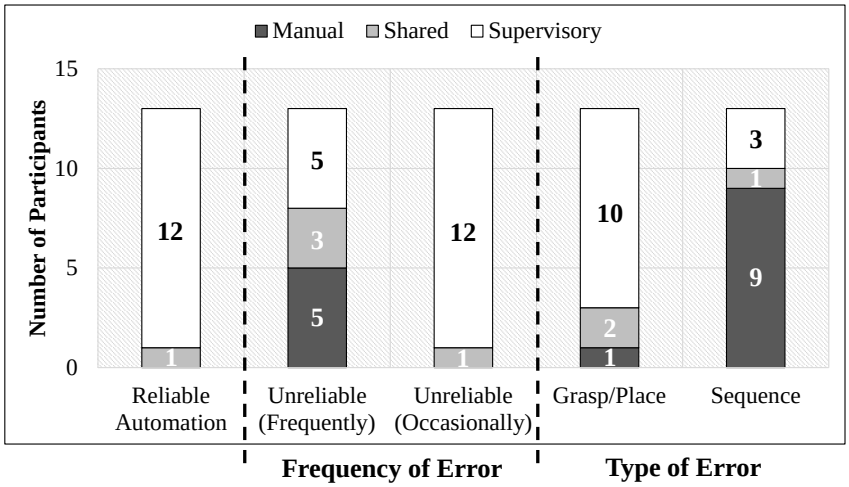



Figure 5.9: The subjective feedback.

## 5.7 Summary and Outlook

**Effective HRC for Unstructured Remote Manipulation —** The work in this chapter compared two HRC paradigms (i.e., shared vs supervisory control) that provide different levels of autonomy to assist humans to control unstructured remote robotic manipulation tasks. The ***supervisory control*** paradigm provides a higher level of assistive autonomy to plan and execute the entire task, but allows humans to use freeform reaching and moving motions to re-select the target location and objects during the task, and the action to apply. The ***shared control paradigm*** provides a lower level of assistive autonomy to perform the human-intended precise manipulation actions which tend to be difficult for human control, while the human operators use their hand pose to control the freeform gross manipulation. Our user studies discovered new knowledge about the effectiveness of HRC paradigms for unstructured remote manipulation: while the supervisory control paradigm with a higher level of assistive autonomy outperformed the shared for all the measures, some participants still had a similar rating for the supervisory and shared control, because they felt more engaged when using the shared control making it easier to detect and respond to robot errors. We also found the intent inference based multi-modal inputs (human gaze, task and robot states) are simple and effective, and can be applied to various remote robot control and supervision.

While our user studies compared two generalizable shared and supervisory control paradigms, we are aware of the various levels of robot autonomy and human-robot collaboration paradigms (see the review in [63]) to be examined in future work. The levels of assistive autonomy may be adjustable or even adaptive, which implies the dynamic role arbitration in the human-robot collaboration [285]. Our user studies are also limited because most participants are experienced video game players. A further study with users of diverse backgrounds, ages, gaming, and VR interfaces is necessary for a holistic evaluation of the impacts of unreliable autonomy.

**Influences of Unreliable Assistive Autonomys —** In this chapter, we manipulated the error type

and frequency and evaluate the impacts of unreliable autonomy on performance, workload, and user preference. We found that in general, more frequent and severe errors have more significant impacts on task performance, human workload, and the operator's preference to use autonomy (which may imply the user's trust). Particularly, we noticed that humans still prefer to use unreliable autonomy if the errors are easy to correct, which implies that an effective mechanism for error recovery may significantly improve the resilience and usability of robot autonomy. We also noticed that compared to the shared control, supervisory control has significantly higher ($p < .05$) task performance and physical workload but comparable cognitive workload. Because the robot control in our task heavily relied on remote camera visual feedback, it is likely that a more unreliable robot autonomy has more adverse impacts on the cognitive workload (to track the task more carefully and detect the errors) than the physical workload (to correct the errors).

Besides error types and frequency, we noticed other factors that may influence the users' perception of robot autonomy, and preference to use a higher level of autonomy. Our post-study survey shows that the participants always choose to use the supervisory control (with a higher level of assistive autonomy) when it is reliable, but only occasionally if the autonomy is unreliable. However, we also found most participants (9 out of 13) would still prefer using a higher level of robot autonomy if the errors happened at a later stage of the task, which implies the perception of the robot autonomy depend not only on the frequency and types of errors but also when the errors are likely to happen. Moreover, some participants (3 out of 13) indicated they preferred not using a higher level of robot autonomy since it is tedious to switch to a lower level of autonomy when they have to correct the error. Our future work will address the limitation of this chapter by examining the influence of these additional factors on human perception and trust in robot autonomy. Because trust is a complex construct that needs iterative establishment over time, we will choose objective measurements to assess trust. Moreover, we will refer to the related work that summarizes the attributes of robot autonomy (see the review in [426]), consider a more systematic approach to

manipulate the errors, and consider errors caused by unreliable robot perception, decision-making, and action.

**Estimation of Human Workload and Engagement —** This chapter presented a novel approach to achieve the **objective** and **accurate** estimation of physical workload, cognitive workload, and engagement based on human gaze and motion tracking. Our proposed approach can be used to assess the usability and comfort of a wide range of physical human-robot interactions and interfaces and has the potential to be converted from offline to online estimation to enable shared autonomy to adapt to the human workload.

In this chapter, our proposed approach was used to estimate physical and cognitive workload in the post-study offline evaluation. Thus far, we have developed the predictive model for the online estimation of physical workload. Our future work will develop the predictive model for the online estimation of cognitive workload. We also need to validate the effectiveness and accuracy of these models in future user studies.

# Chapter 6

# Conclusion

This dissertation delves into a series of interrelated challenges in the realm of remote manipulation for tele-nursing robots. Specifically, it focuses on (1) developing teleoperation interfaces that adapt to workload demands, enabling efficient, intuitive, and effortless control, (2) designing remote manipulation assistance in the form of perception and action augmentation based on natural human perception-action coordination, and (3) investigating the impact of reliability of teleoperation assistance in workload and user preference.

## 6.1 Summary of Findings and Main Contributions

In this dissertation, we have presented findings on effective and user-friendly approaches for remotely controlling tele-nursing robots. Our research focuses on achieving key design objectives in terms of efficiency, intuitiveness, and ergonomic usability, which are essential for the seamless integration of these robots into the workflow of nursing professionals, as well as enhancing the experience of future end-users. Building upon the insights gained from our research on effective and effortless control of tele-nursing robots, we have developed novel methods that address the issue of limited sensory feedback in remote manipulation. Our proposed approach is based on natural human behaviors and systematic perception and action augmentation. By integrating these methods, we aim to enhance user adaptivity and establish human-preferred remote manipulation assistance. Additionally, this dissertation takes a preliminary step toward evaluating the impact

of unreliable robot execution. The developed human-robot interfaces ensure effectiveness and are generalizable to various remote manipulation systems that may differ in robot platform, control input, and feedback mechanism.

In **Chapter 2**, our general evaluation found that: whole-body motion mapping interfaces have the best task performance and learning efforts among freeform teleoperation interfaces for nursing robots. However, it may cause non-negligible physical fatigue and prevent users from teleoperating with the robots for a long time. Our integrated participant interview and survey also identified and ranked the fatigue-causing factors, including maintaining steady postures for wrist camera controls and adjusting arm posture for stable object grasping. Our robot autonomy design and specialized evaluation focused on reducing the physical workload of the interface and improving its ergonomics of the interface. For the specialized evaluation, we proposed a novel EMG-based muscle effort index, to provide a more detailed, objective, and accurate physical workload assessment. The outcomes of the specialized evaluation validated the efficacy of our robot autonomy design, as well as our proposed framework for interface evaluation and evolution. We believe the proposed framework, including most of the evaluation metrics, applies to human-robot teaming interfaces beyond teleoperation. The main contributions of this chapter are:

**An Evaluation Framework and Methods —** A proposed hierarchical framework that evaluates a human-robot interface from general to specific characteristics; The specific evaluation focuses on addressing the limitation of the characteristics of the interface, rather than augmenting the task-specific robot autonomous functions.

**Integrated Design Evaluation and Evolution —** Apply the proposed evaluation framework to the interface design evolution of general-purpose tele-nursing robots. We focus on the skill sets necessary for freeform teleoperation instead of evaluating a large variety of specific nursing tasks; We also consider the needs of the primary user population, the nurses.

In **Chapter 3**, we aim to transform the design philosophy for tele-robotic interfaces, based on a deep understanding of the perception-action coupling of cyber-human systems. Among the many aspects of motion control, the coordination between perception and action is most critical to tele-nursing task performance. Knowledge about perception-action coupling has been leveraged in human-robot interaction to a limited extent and has already yielded effective models and approaches for predicting human intent [30], optimizing camera motions and viewpoints [31], interactive perception [32], and sensory augmentation of human-robot interfaces for motor skill training and rehabilitation [33]. While remote robots limit human perception and motion capabilities, they also provide opportunities for the human motor system to explore. Novel perception-action coupling skills do not exist in the repertoire of human motor control, yet are critical for robot teleoperation. Through the robot teleoperation interface, the human and the robot are closely coupled as an integrated cyber-human system, and novel perception-action coordination needs to be developed to adapt this system's new perception and action capabilities. To facilitate this adaptation, both robot teleoperation interface and assistive autonomy need to be designed based on the human behavior and preference of perception-action coupling, which has been studied extensively in human movement science [216, 318], but not at all for cyber-human systems. The research efforts in this chapter aim to bridge this gap, by proposing a **novel experimental paradigm** that can simulate human natural behavior and preference in the usage of active telepresence. We further conduct comprehensive user studies with this experimental paradigm to 1) discover the *perception-action coupling* of a coupled human-robot system, and 2) reveal its implications to the *design of robot teleoperation interface and assistive autonomy*. The key findings of this chapter suggest that the head should control the camera selected for telepresence, and the shared autonomy camera's movements should be simple and predictable for the users, such as translations or rotations. Additionally, this chapter highlights the importance of integrating vision and haptic feedback in robot teleoperation interfaces. These insights can be used to improve the design of telepresence systems and enhance the user experience. The main contribution of this chapter is:

**Novel Experimental Paradigm —** The novel experimental paradigm we proposed was designed to study the perception-action coordination, human adaptation, and preference in the usage of active telepresence cameras. To eliminate the effort of controlling the robot, the experimental paradigm provided a simulated telepresence setting with video streams from the cameras attached to the user's head, torso, dominant and non-dominant hands as well as a standalone workspace camera while retaining the humans' ability to perform object manipulation. These video streams were used by the participants to stack lightweight plastic cups into a pyramid.

In **Chapter 4**, we first proposed a generalizable sensory augmentation design to assist dexterous tele-manipulation tasks. The designs well supported the findings in our human factor experiment that investigated the visual and haptic sensory integration in the usage of active telepresence cameras for general-purpose manipulation. We found out that transparency, performance, and workload were improved with the haptic and AR visual cues over the baseline without sensory feedback. Also, the preference for sensory integration tended to avoid feedback overlap when involving the different types of secondary tasks. Secondly, we proposed perception and action augmentation on a representative interface. We used the HTC Vive hand-held controller to control robot motion and used a desktop monitor to display the remote camera viewpoints and AR visual cues. The robot could provide autonomous actions (e.g., grasping and placing actions) or switch to constrained motion control using a trackpad when humans operated the robot end-effector near the target object or location to place. The implementation can be generalized to the teleoperation system using various contemporary tele-manipulation control devices (e.g., hand-held, touch-based, wearable), display (e.g., screen-based and head-mounted visual display). Our comprehensive user study shows that *the effectiveness of and preference for the perception and action augmentation depend on the task performance objective, the user's need for assistance, and the types of users*. Lastly, we presented multiple control modes with varying levels of autonomy for a pick-and-place remote robot manipulation task. We also provided several AR features that participants could select to create their ideal

visual interface for each control mode. Our user study investigated how participants select AR cues based on the level of robot autonomy. The participants' priority for AR visual cues shifted from guiding human motion for robot control to communicating autonomy activation and intention as the level of autonomy transitioned from direct to supervisory control. With the increasing levels of autonomy, their preference for AR cues shifted from providing local information to global guidance for robot control. Additionally, AR cues that served the same purpose as the robot autonomy had reduced efficacy and was generally not preferred by the participants. Our user study also identified that the participant's preference for AR visual cues tended to converge to the recommendation of experienced users with hands-on experience using the visual interfaces regardless of their initial selections based on video instructions. The main contributions of this chapter are:

**Assist-as-Needed Shared Autonomy —** A novel shared autonomy to leverage human capabilities of freeform control and an assist-as-needed robot autonomy for effective, intuitive, and ergonomic human-robot collaboration to perform tele-manipulation tasks.

**Generalizable Sensory Feedback —** A generalizable design of haptic and AR visual cues to provide the information critical to the precision and performance of the remote manipulation.

**Systematic Perception and Action Augmentation —** The integration and comparison of different types of perception and action augmentation to discover new knowledge on optimal human-robot collaboration for freeform tele-manipulation.

**Objective Physical and Cognitive Workload Estimation —** A novel approach for objective physical and cognitive workload estimation based on human motion and eye tracking devices.

In **Chapter 5**, we compared the human-robot shared and supervisory control of the remote robots for dexterous manipulation and investigated how the reliability of the robot autonomy can influence the human operator's performance, workload, and preference for robot assistance. Specifically, we proposed two human-robot collaboration (HRC) paradigms for remote manipulation: (1)

the shared control paradigm allows humans to control gross manipulation and the robot autonomy to control the precise manipulation actions, and (2) the supervisory control paradigm allows the robot autonomy to control both the gross and precise manipulation actions but relies on humans to detect and correct the manipulation errors. We conducted two user studies: one to compare the two HRC paradigms and characterize their effectiveness when the assistive autonomy is reliable; and the other to investigate how the type and frequency of the errors affect the tasks and human operators in the two HRC paradigms when assistive autonomy is not reliable. Our results show that: (1) the interface with a higher level of reliable autonomy yields significantly better performance, lower workload, and higher user preference but lower engagement, and (2) the frequency and type of the error have significant impacts on the task performance and human workload but only partially affects the operator's preference and usage of autonomy. The main contributions of this chapter are:

**Human Goal Intent Inference —** We infer human goals and action intents, by tracking human gaze fixation (using Tobii Pro Nano eye tracking device) and robot states (i.e., distances to each object and container in the workspace, the opening and closing of the gripper).

**Objective Visual and Action Engagement Monitoring —** We tracked human eye movements and calculated the percentage of the task for which the gaze fixation was in the area of interest (i.e., targets and picture-in-picture) to estimate *visual engagement*. To estimate the *action engagement* (level of activity), we tracked the positions of the two handheld controllers and 6 body trackers (Vive Tracker 3.0) attached to the operator's upper arms, forearms, chest, and waist to measure the shoulder and elbow joint angles by inner product formula.

## 6.2 Proceeding Work

In this section, we first present a summary of several ongoing research efforts that have resulted from the collaborative works of the author. The examples mentioned herein are potential areas for further exploration, building upon the topics covered in this dissertation.

**Online Workload Estimation —** Chapters 4 and 5 centered on our proposed approaches for estimating physical and cognitive workload in the post-study offline evaluation. Our implementation could be extended to online prediction through the use of learned predictive models and novel definitions of gaze behaviors for real-time physical and cognitive workload estimation. Figure 6.1 depicts the preliminary implementation of online instantaneous and cumulative physical workload estimation. The approach employs a predictive model of joint angle-EMG and an algorithm that employs control speed as the threshold to initiate cumulative instant physical workload measurement. This framework captures physical workload during both dynamic and stationary arm motions. Figure 6.2 highlights three independent aspects of cognitive workload that capture stress, operational concentration, and user interface complexity.

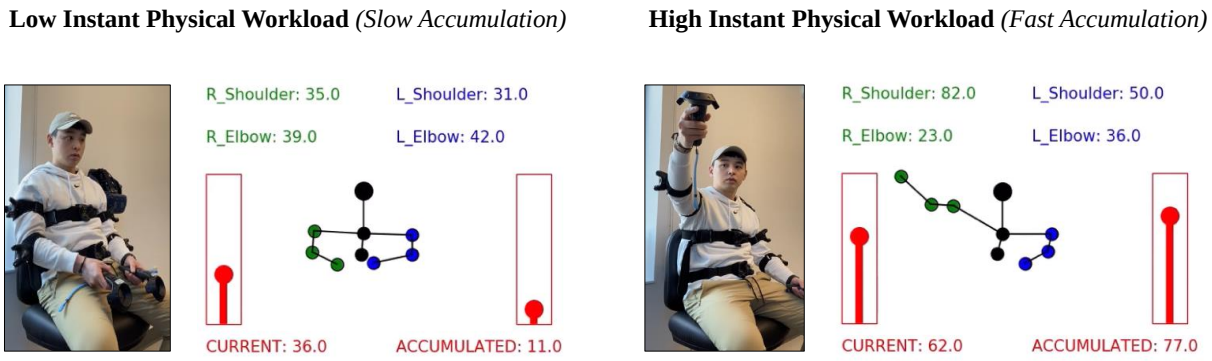



Figure 6.1: Real-time instantaneous and cumulative physical workload based on human motion tracking and control speed.

Integrating online workload estimation into the workload-adaptive control interface would enhance its capabilities to include workload-based action and perception assistance. This would in-

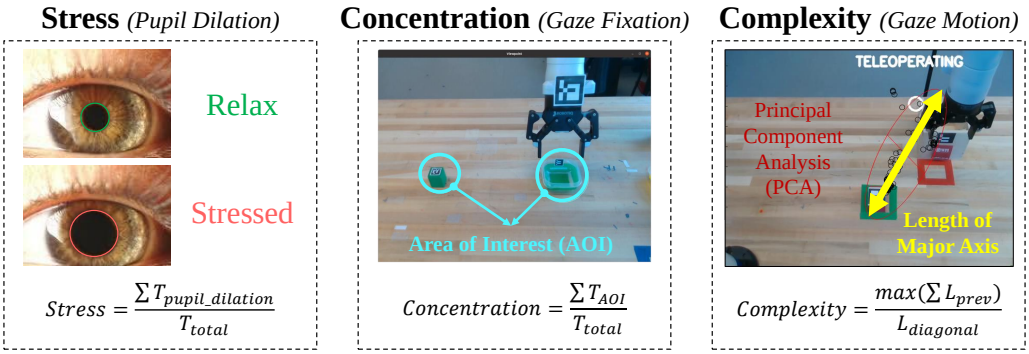



Figure 6.2: Real-time cognitive physical workload based on pupil size, gaze fixation, and distribution. $T$ refer to as time, $L$ as length, and $L_{prev}$ as all the length from the pre-defined window.

volve utilizing real-time physical workload data and stress levels to activate robot autonomy or action augmentation, and adjusting the type and degree of perception augmentation in response to online cognitive workload estimation, such as gaze motion and fixation. This would enable an assist-as-needed mechanism for advanced remote manipulation assistance.

**Intuitive Viewpoints Control —** Chapter 3 presented our investigation into human-preferred perception and action coordination, revealing that intuitive control should be applied to the selected viewpoint. Figure 6.3 showcases the initial implementation of the head-controlled primary viewpoint when using a head-mounted display as the visual device and the gaze-controlled primary viewpoint when using a monitor display as the visual device.

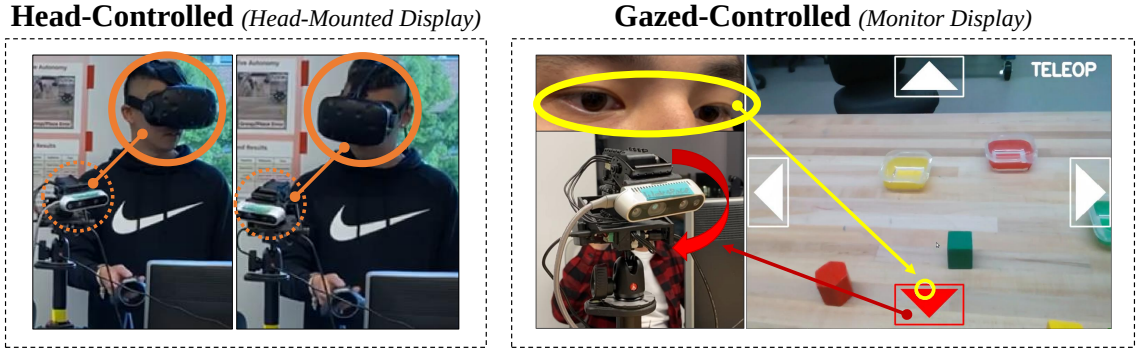



Figure 6.3: Head- and gaze-controlled primary viewpoint for head-mounted and monitor display.

To enhance the usability of screen-based remote manipulation and incorporate a complementary perspective into the visual interface, various levels of camera autonomy can be applied to both primary and complementary viewpoints to enable intuitive camera control. The framework of different levels of camera autonomy, ranging from direct control to full automation for primary viewpoint control, is illustrated in Figure 6.4. Apart from direct control using a trackpad and gaze, the shared autonomy paradigm can detect the robot's status (end-effector location) and map it to camera tilt and pan control. Furthermore, the full automation paradigm can adjust the camera pose to focus on the target object based on the task state. This approach offers users greater flexibility and control, allowing them to manipulate the camera with ease and precision.

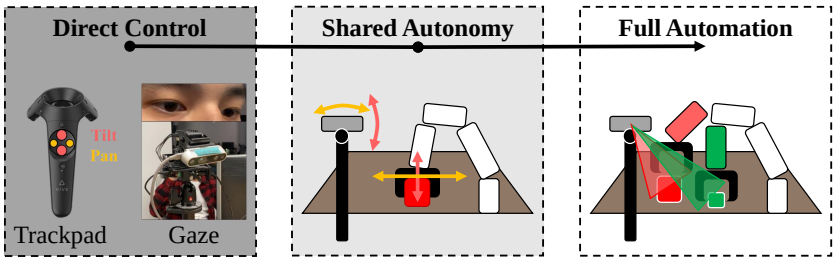



Figure 6.4: Primary viewpoint control spectrum.

## 6.3 Limitations and Future Directions

Previously, we discussed the research directions that we are currently exploring. In the following context, we outline additional potential avenues for future research that we have yet to pursue but we believe are highly pertinent to the results presented in this dissertation.

**Motion Tracking Interface —** In this dissertation, we employed two different approaches to control remote manipulation. Firstly, we used the Vicon system to generate a complete human skeleton, which served as the baseline for manipulation, perception, and navigation through the motion mapping algorithm discussed in Chapter 2. Secondly, we leveraged the portable HTC Vive system to control the KINOVE Gen3 robot through a handheld controller in Chapter 4 and 5. This alterna-

tive approach was implemented to address the relative complexity associated with the preparation requirements of the Vicon system. However, the use of the handheld controller came at the cost of sacrificing the advantages of tracking natural human posture. Tracking natural human posture has the potential to enable joint space control for the robotic arm, generating human-like postures that can avoid unacceptable motion and even robot singularity. An intriguing avenue for future research would be to explore joint space control by integrating motion trackers, such as IMUs, to create a human motion model as an extension of the motion tracking interface via the HTC Vive handheld controller. This could potentially enhance the naturalness of the robot's movements and improve its ability to mimic human-like postures.

**Tele-nursing Robot Control —** Controlling a tele-nursing robot demands the proper coordination of manipulation, perception, and navigation. In Chapter 2, we introduced a whole-body motion mapping interface for teleoperating the tele-nursing robot. While we subsequently advanced the motion tracking interface to a handheld controller in Chapter 4, we have yet to test this new interface for controlling the tele-nursing robot in its entirety. Instead, our focus has been primarily on the robot manipulator. An intuitive extension of the HTC Vive handheld controller interface would be to leverage the pose and buttons of two controllers for coordinating bimanual manipulation, navigation, active telepresence, and assistive autonomy activation. This could potentially enhance the robot's capability to perform complex tasks and improve the telepresence experience.

**Haptic Feedback —** In this dissertation, we extensively investigated and utilized perception augmentation, such as augmented reality visual cues. This technology provides rich information and enables transparent communication between human users and robot platforms, ultimately enhancing the telepresence experience. Chapter 4 showcased the benefits of using haptic feedback to aid in remote manipulation. However, the implementation of this technology was limited by hardware constraints. The device we used was only capable of conveying a single-channel vibration that could vary in magnitude and frequency, thus restricting the range of feedback that could be pro-

vided to the operator. Integrating more sophisticated haptic devices (e.g., multiple tactile sensors or detailed haptic gloves) could provide a potential solution to the limitations of our previous implementation. By comparing the effectiveness of these devices with the AR interfaces proposed in this dissertation, we could potentially identify suitable methods for integrating multi-sensory feedback.

**General-purpose Manipulation Task —** Remote manipulation encompasses a wide range of tasks with varying levels of difficulty, which can be affected by multiple factors such as the workspace (e.g. cluttered or simple, small or large), robot capability (e.g. gripper setup), object properties (e.g. solid, deformable, fragile), and objectives (e.g. efficiency, precision). Throughout the dissertation, we conducted several user studies that utilized different types of remote manipulation tasks. These included nursing-related tasks in Chapter 2, high-complexity dexterous tasks in Chapter 3, widely used single object pick-and-place tasks in Chapter 4, and multi-object sorting tasks in Chapter 5. However, there are many more manipulation tasks that could be considered, and it is impractical to test them all to ensure generalizability. One possible approach is to identify a set of sub-tasks that could generate the majority of manipulation tasks by combining them in a certain sequence.

**User Groups —** Involving target or end-users in design leads to wider acceptance, better satisfaction, and improved usability. User input helps identify issues and suggest unconsidered improvements. To this end, we recruited nursing professionals as participants in Chapters 2 and 4 to gather their feedback and conducted comprehensive post-study discussions and interviews with them. Throughout our user studies, we discovered that there are many human factors that can impact performance, preference, and usability. For example, participants who had prior gaming experience demonstrated better situational awareness when controlling the remote perception, resulting in outstanding performance. To ensure a comprehensive evaluation of the remote manipulation system, it is necessary to conduct further studies with users of diverse backgrounds, ages, gaming experiences, and VR experiences. This can help identify any potential biases or limitations in the current system and ensure that it is accessible and usable for a wider range of users.

# Bibliography


[1] Z. Li, P. Moran, Q. Dong, R. J. Shaw, and K. Hauser, "Development of a tele-nursing mobile manipulator for remote care-giving in quarantine areas," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3581–3586, IEEE, 2017.

[2] Y. Zhu, A. Smith, and K. Hauser, "Automated heart and lung auscultation in robotic physical examinations," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4204–4211, 2022.

[3] E. Ackerman, "Moxi prototype from diligent robotics starts helping out in hospitals," *IEEE Spectrum. https://spectrum. ieee. org/automaton/robotics/industrial-robots/moxi-prototype-fro m-diligent-robotics-starts-helping-out-in-hospitals*, 2018.

[4] D. Tolani and N. I. Badler, "Real-time inverse kinematics of the human arm," *Presence: Teleoperators & Virtual Environments*, vol. 5, no. 4, pp. 393–401, 1996.

[5] A. M. Zanchettin, N. M. Ceriani, P. Rocco, H. Ding, and B. Matthias, "Safety in human-robot collaborative manufacturing environments: Metrics and control," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 882–893, 2015.

[6] T. L. Gibo, W. Mugge, and D. A. Abbink, "Trust in haptic assistance: weighting visual and haptic cues based on error history," *Experimental brain research*, vol. 235, no. 8, pp. 2533–2546, 2017.

[7] T. Williams, M. Bussing, S. Cabrol, E. Boyle, and N. Tran, "Mixed reality deictic gesture for multi-modal robot communication," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 191–201, IEEE, 2019.

[8] L. Peternel, N. Tsagarakis, D. Caldwell, and A. Ajoudani, "Robot adaptation to human physical fatigue in human–robot co-manipulation," *Autonomous Robots*, vol. 42, no. 5, pp. 1011–1021, 2018.

[9] T.-C. Lin, A. U. Krishnan, and Z. Li, "Intuitive, efficient and ergonomic tele-nursing robot interfaces: Design evaluation and evolution," *ACM Transactions on Human-Robot Interaction*, 2022.

[10] R. Ebad and K. Jazan, "Telemedicine: Current and future perspectives telemedicine: Current and future perspectives," 2013.

[11] V. J. Munster, M. Koopmans, N. van Doremalen, D. van Riel, and E. de Wit, "A novel coronavirus emerging in china—key questions for impact assessment," *New England Journal of Medicine*, 2020.

[12] A. J. Kucharski, A. Camacho, S. Flasche, R. E. Glover, W. J. Edmunds, and S. Funk, "Measuring the impact of ebola control measures in sierra leone," *Proceedings of the National Academy of Sciences*, vol. 112, no. 46, pp. 14366–14371, 2015.

[13] E. Ackerman, "irobot and intouch health announce rp-vita telemedecine robot," *IEEE Spectrum*, 2012.

[14] E. Ackerman, "Ava robotics introduces autonomous telepresence robot," *IEEE Spectrum*, 2018.

[15] E. Ackerman, "Suitable tech introduces beampro 2 telepresence platform," *IEEE Spectrum*, 2018.

[16] K. Tsui, A. Norton, D. Brooks, H. Yanco, and D. Kontak, "Designing telepresence robot systems for use by people with special needs," in *Int. Symposium on Quality of Life Technologies: Intelligent Systems for Better Living*, 2011.

[17] J. L. Wright, J. Y. Chen, and S. G. Lakhmani, "Agent transparency and reliability in human–robot interaction: the influence on user confidence and perceived reliability," *IEEE Transactions on Human-Machine Systems*, 2019.

[18] T.-C. Lin, A. U. Krishnan, and Z. 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)*, pp. 2240–2245, IEEE, 2019.

[19] R. Elbasiony and W. Gomaa, "Humanoids skill learning based on real-time human motion imitation using kinect," *Intelligent Service Robotics*, vol. 11, no. 2, pp. 149–169, 2018.

[20] L. Penco, B. Clément, V. Modugno, E. M. Hoffman, G. Nava, D. Pucci, N. G. Tsagarakis, J.-B. Mouret, and S. Ivaldi, "Robust real-time whole-body motion retargeting from human to humanoid," in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pp. 425–432, IEEE, 2018.

[21] J. Oh, I. Lee, H. Jeong, and J.-H. Oh, "Real-time humanoid whole-body remote control framework for imitating human motion based on kinematic mapping and motion constraints," *Advanced Robotics*, vol. 33, no. 6, pp. 293–305, 2019.

[22] D. Kent, C. Saldanha, and S. Chernova, "A comparison of remote robot teleoperation interfaces for general object manipulation," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 371–379, 2017.

[23] D. Rakita, B. Mutlu, M. Gleicher, and L. M. Hiatt, "Shared control–based bimanual robot manipulation," *Science Robotics*, vol. 4, no. 30, p. eaaw0955, 2019.

[24] T. Dardona, S. Eslamian, L. A. Reisner, and A. Pandya, "Remote presence: Development and usability evaluation of a head-mounted display for camera control on the da vinci surgical system," *Robotics*, vol. 8, no. 2, p. 31, 2019.

[25] D. Rakita, B. Mutlu, and M. Gleicher, "An autonomous dynamic camera method for effective remote teleoperation," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 325–333, 2018.

[26] R. Rahal, G. Matarese, M. Gabiccini, A. Artoni, D. Prattichizzo, P. R. Giordano, and C. Pacchierotti, "Caring about the human operator: haptic shared control for enhanced user comfort in robotic telemanipulation," *IEEE transactions on haptics*, vol. 13, no. 1, pp. 197–203, 2020.

[27] R. T. Chadalavada, H. Andreasson, M. Schindler, R. Palm, and A. J. Lilienthal, "Bi-directional navigation intent communication using spatial augmented reality and eye-tracking glasses for improved safety in human–robot interaction," *Robotics and Computer-Integrated Manufacturing*, vol. 61, p. 101830, 2020.

[28] S. Javdani, H. Admoni, S. Pellegrinelli, S. S. Srinivasa, and J. A. Bagnell, "Shared autonomy via hindsight optimization for teleoperation and teaming," *The International Journal of Robotics Research*, vol. 37, no. 7, pp. 717–742, 2018.

[29] S. Honig and T. Oron-Gilad, "Understanding and resolving failures in human-robot interaction: Literature review and model development," *Frontiers in psychology*, vol. 9, p. 861, 2018.

[30] R. M. Aronson and H. Admoni, "Eye gaze for assistive manipulation," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 552–554, 2020.

[31] D. Rakita, B. Mutlu, and M. Gleicher, "Remote telemanipulation with adapting viewpoints in visually complex environments," *Robotics: Science and Systems XV*, 2019.

[32] J. Bohg, K. Hausman, B. Sankaran, O. Brock, D. Kragic, S. Schaal, and G. S. Sukhatme,

"Interactive perception: Leveraging action in perception and perception in action," *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1273–1291, 2017.

[33] D. C. Irimia, W. Cho, R. Ortner, B. Z. Allison, B. E. Ignat, G. Edlinger, and C. Guger, "Brain-computer interfaces with multi-sensory feedback for stroke rehabilitation: a case study," *Artificial organs*, vol. 41, no. 11, pp. E178–E184, 2017.

[34] A. Unni Krishnan, T.-C. Lin, and Z. Li, "Design interface mapping for efficient free-form tele-manipulation," in *to appear in the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2022.

[35] M. Ferre, R. Aracil, J. M. Bogado, and R. J. Saltarén, "Improving force feedback perception using low bandwidth teleoperation devices," in *Proceedings of EuroHaptics Conference EH'2004*, 2004.

[36] P. Schleer, P. Kaiser, S. Drobinsky, and K. Radermacher, "Augmentation of haptic feedback for teleoperated robotic surgery," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 3, pp. 515–529, 2020.

[37] B. S. Peters, P. R. Armijo, C. Krause, S. A. Choudhury, and D. Oleynikov, "Review of emerging surgical robotic technology," *Surgical endoscopy*, vol. 32, no. 4, pp. 1636–1655, 2018.

[38] J. Bodner, H. Wykypiel, G. Wetscher, and T. Schmid, "First experiences with the da vinci™ operating robot in thoracic surgery," *European Journal of Cardio-thoracic surgery*, vol. 25, no. 5, pp. 844–851, 2004.

[39] M. A. Goodrich, J. W. Crandall, and E. Barakova, "Teleoperation and beyond for assistive humanoid robots," *Reviews of Human factors and ergonomics*, vol. 9, no. 1, pp. 175–226, 2013.

[40] K. Yokoi, K. Nakashima, M. Kobayashi, H. Mihune, H. Hasunuma, Y. Yanagihara, T. Ueno, T. Gokyuu, and K. Endou, "A tele-operated humanoid operator," *The International Journal of Robotics Research*, vol. 25, no. 5-6, pp. 593–602, 2006.

[41] L. Fritsche, F. Unverzag, J. Peters, and R. Calandra, "First-person tele-operation of a humanoid robot," in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pp. 997–1002, IEEE, 2015.

[42] G. Du, P. Zhang, and X. Liu, "Markerless human-manipulator interface using leap motion with interval kalman filter and improved particle filter," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 694–704, 2016.

[43] O. Porges, M. Connan, B. Henze, A. Gigli, C. Castellini, and M. A. Roa, "A wearable, ultralight interface for bimanual teleoperation of a compliant, whole-body-controlled humanoid robot," in *Proceedings of ICRA-International Conference on Robotics and Automation*, 2019.

[44] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays," *The International Journal of Robotics Research*, vol. 38, no. 12-13, pp. 1513–1526, 2019.

[45] A. Cela, J. J. Yebes, R. Arroyo, L. M. Bergasa, R. Barea, and E. López, "Complete low-cost implementation of a teleoperated control system for a humanoid robot," *Sensors*, vol. 13, no. 2, pp. 1385–1401, 2013.

[46] J. Liu and Y. Zhang, "Mapping human hand motion to dexterous robotic hand," in *2007 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 829–834, IEEE, 2007.

[47] K. Fischer, F. Kirstein, L. C. Jensen, N. Krüger, K. Kukliński, M. V. aus der Wieschen, and T. R. Savarimuthu, "A comparison of types of robot control for programming by demonstration," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 213–220, IEEE, 2016.

[48] N. Miller, O. C. Jenkins, M. Kallmann, and M. J. Mataric, "Motion capture from inertial sensing for untethered humanoid teleoperation," in *4th IEEE/RAS International Conference on Humanoid Robots, 2004.*, vol. 2, pp. 547–565, IEEE, 2004.

[49] K. Yamane and J. Hodgins, "Simultaneous tracking and balancing of humanoid robots for imitating human motion capture data," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2510–2517, IEEE, 2009.

[50] B. Dariush, M. Gienger, A. Arumbakkam, Y. Zhu, B. Jian, K. Fujimura, and C. Goerick, "Online transfer of human motion to humanoids," *International Journal of Humanoid Robotics*, vol. 6, no. 02, pp. 265–289, 2009.

[51] J. Koenemann and M. Bennewitz, "Whole-body imitation of human motions with a nao humanoid," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pp. 425–426, ACM, 2012.

[52] G. Du, P. Zhang, J. Mai, and Z. Li, "Markerless kinect-based hand tracking for robot teleoperation," *International Journal of Advanced Robotic Systems*, vol. 9, no. 2, p. 36, 2012.

[53] J. Koenemann, F. Burget, and M. Bennewitz, "Real-time imitation of human whole-body motions by humanoids," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2806–2812, IEEE, 2014.

[54] C.-L. Hwang and G.-H. Liao, "Real-time pose imitation by mid-size humanoid robot with

servo-cradle-head rgb-d vision system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 181–191, 2018.

[55] B. Fang, F. Sun, H. Liu, D. Guo, W. Chen, and G. Yao, "Robotic teleoperation systems using a wearable multimodal fusion device," *International Journal of advanced robotic systems*, vol. 14, no. 4, p. 1729881417717057, 2017.

[56] M. K. Kim, K. Ryu, Y. Oh, S.-R. Oh, and K. Kim, "Implementation of real-time motion and force capturing system for tele-manipulation based on semg signals and imu motion data," pp. 5658–5664, IEEE, 2014.

[57] S. Tortora, M. Moro, and E. Menegatti, "Dual-myo real-time control of a humanoid arm for teleoperation," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 624–625, IEEE, 2019.

[58] Y. Chae, J. Jeong, and S. Jo, "Toward brain-actuated humanoid robots: asynchronous direct control using an eeg-based bci," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1131–1144, 2012.

[59] K. Muelling, A. Venkatraman, J.-S. Valois, J. E. Downey, J. Weiss, S. Javdani, M. Hebert, A. B. Schwartz, J. L. Collinger, and J. A. Bagnell, "Autonomy infused teleoperation with application to brain computer interface controlled manipulation," *Autonomous Robots*, vol. 41, no. 6, pp. 1401–1422, 2017.

[60] N. K. N. Aznan, J. D. Connolly, N. Al Moubayed, and T. P. Breckon, "Using variable natural environment brain-computer interface stimuli for real-time humanoid robot navigation," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 4889–4895, IEEE, 2019.

[61] Y. Ou, J. Hu, Z. Wang, Y. Fu, X. Wu, and X. Li, "A real-time human imitation system using kinect," *International Journal of Social Robotics*, vol. 7, no. 5, pp. 587–600, 2015.

[62] A. K. Sinha, S. K. Sahu, R. K. Bijarniya, and K. Patra, "An effective and affordable technique for human motion capturing and teleoperation of a humanoid robot using an exoskeleton," in *2017 2nd International Conference on Man and Machine Interfacing (MAMI)*, pp. 1–6, IEEE, 2017.

[63] J. M. Beer, A. D. Fisk, and W. A. Rogers, "Toward a framework for levels of robot autonomy in human-robot interaction," *Journal of human-robot interaction*, vol. 3, no. 2, pp. 74–99, 2014.

[64] C. Yang, J. Luo, Y. Pan, Z. Liu, and C.-Y. Su, "Personalized variable gain control with tremor attenuation for robot teleoperation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 10, pp. 1759–1770, 2017.

[65] J. Storms, K. Chen, and D. Tilbury, "A shared control method for obstacle avoidance with mobile robots and its interaction with communication delay," *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 820–839, 2017.

[66] K. H. Khokar, R. Alqasemi, S. Sarkar, and R. V. Dubey, "Human motion intention based scaled teleoperation for orientation assistance in preshaping for grasping," in *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, pp. 1–6, IEEE, 2013.

[67] A. D. Dragan and S. S. Srinivasa, "A policy-blending formalism for shared control," *The International Journal of Robotics Research*, vol. 32, no. 7, pp. 790–805, 2013.

[68] D. Rakita, B. Mutlu, M. Gleicher, and L. M. Hiatt, "Shared dynamic curves: A shared-control telemanipulation method for motor task training," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 23–31, 2018.

[69] M. Laghi, M. Maimeri, M. Marchand, C. Leparoux, M. Catalano, A. Ajoudani, and A. Bicchi, "Shared-autonomy control for intuitive bimanual tele-manipulation," in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pp. 1–9, IEEE, 2018.

[70] Y. Ishiguro, K. Kojima, F. Sugai, S. Nozawa, Y. Kakiuchi, K. Okada, and M. Inaba, "High speed whole body dynamic motion experiment with real time master-slave humanoid robot system," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–7, IEEE, 2018.

[71] J. Oh, O. Sim, H. Jeong, and J.-H. Oh, "Humanoid whole-body remote-control framework with delayed reference generator for imitating human motion," *Mechatronics*, vol. 62, p. 102253, 2019.

[72] E.-J. Rolley-Parnell, D. Kanoulas, A. Laurenzi, B. Delhaisse, L. Rozo, D. G. Caldwell, and N. G. Tsagarakis, "Bi-manual articulated robot teleoperation using an external rgb-d range sensor," in *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 298–304, IEEE, 2018.

[73] A. F. Salazar-Gomez, J. DelPreto, S. Gil, F. H. Guenther, and D. Rus, "Correcting robot mistakes in real time using eeg signals," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6570–6577, IEEE, 2017.

[74] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common metrics for human-robot interaction," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pp. 33–40, 2006.

[75] J. Burke, M. Lineberry, K. S. Pratt, M. Taing, R. Murphy, and B. Day, "Toward developing hri metrics for teams: Pilot testing in the field," *Metrics for Human-Robot Interaction 2008*, p. 21, 2008.

[76] P. Pina, M. Cummings, J. Crandall, and M. Della Penna, "Identifying generalizable metric classes to evaluate human-robot teams," in *Workshop on Metrics for Human-Robot Interaction, 3rd Ann. Conf. Human-Robot Interaction*, pp. 13–20, 2008.

[77] J. A. Saleh and F. Karray, "Towards generalized performance metrics for human-robot interaction," in *2010 International Conference on Autonomous and Intelligent Systems, AIS 2010*, pp. 1–6, IEEE, 2010.

[78] A. Weiss, R. Bernhaupt, and M. Tscheligi, "The usus evaluation framework for user-centered hri," *New Frontiers in Human–Robot Interaction*, vol. 2, pp. 89–110, 2011.

[79] S. Singer and D. Akin, "A survey of quantitative team performance metrics for human-robot collaboration," in *41st International Conference on Environmental Systems*, p. 5248, 2011.

[80] J. Abou Saleh, "A novel approach for performance assessment of human-robotic interaction," 2012.

[81] R. R. Murphy and D. Schreckenghost, "Survey of metrics for human-robot interaction," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 197–198, IEEE, 2013.

[82] C. E. Harriott, *Workload and task performance in human-robot peer-based teams*. PhD thesis, Vanderbilt University, 2015.

[83] Y. Gatsoulis, G. S. Virk, and A. A. Dehghani-Sanij, "On the measurement of situation awareness for effective human-robot interaction in teleoperated systems," *Journal of cognitive engineering and decision making*, vol. 4, no. 1, pp. 69–98, 2010.

[84] J. Y. Chen, E. C. Haas, and M. J. Barnes, "Human performance issues and user interface design for teleoperated robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1231–1245, 2007.

[85] J. Y. Chen, M. J. Barnes, and M. Harper-Sciarini, "Supervisory control of multiple robots: Human-performance issues and user-interface design," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 4, pp. 435–454, 2010.

[86] J. Y. Chen, E. C. Haas, K. Pillalamarri, and C. N. Jacobson, "Human-robot interface: Issues in operator performance, interface design, and technologies," tech. rep., ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD, 2006.

[87] J. Y. Chen, "Robotics operator performance in a multi-tasking environment," *Human-Robot Interactions in Future Military Operations*, pp. 294–314, 2010.

[88] J. Y. Chen and M. J. Barnes, "Human–agent teaming for multirobot control: A review of human factors issues," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 1, pp. 13–29, 2014.

[89] M. Gombolay, A. Bair, C. Huang, and J. Shah, "Computational design of mixed-initiative human–robot teaming that considers human factors: situational awareness, workload, and workflow preferences," *The International journal of robotics research*, vol. 36, no. 5-7, pp. 597–617, 2017.

[90] M. Lohse, M. Hanheide, K. J. Rohlfing, and G. Sagerer, "Systemic interaction analysis (sina) in hri," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 93–100, 2009.

[91] J. Peltason, N. Riether, B. Wrede, and I. Lütkebohle, "Talking with robots about objects: a system-level evaluation in hri," in *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 479–486, IEEE, 2012.

[92] J.-H. Hwang, K. Lee, and D.-S. Kwon, "A formal method of measuring interactivity in hri,"

in *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 738–743, IEEE, 2007.

[93] G. Hoffman, "Evaluating fluency in human–robot collaboration," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 3, pp. 209–218, 2019.

[94] R. H. Wortham, A. Theodorou, and J. J. Bryson, "Robot transparency: Improving understanding of intelligent behaviour for designers and users," in *Annual Conference Towards Autonomous Robotic Systems*, pp. 274–289, Springer, 2017.

[95] D. Labonte, P. Boissy, and F. Michaud, "Comparative analysis of 3-d robot teleoperation interfaces with novice users," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 5, pp. 1331–1342, 2010.

[96] H. A. Yanco, J. L. Drury, and J. Scholtz, "Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition," *Human–Computer Interaction*, vol. 19, no. 1-2, pp. 117–149, 2004.

[97] A. M. Howard, "A methodology to assess performance of human-robotic systems in achievement of collective tasks," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 90–95, IEEE, 2005.

[98] E. Tunstel, "Operational performance metrics for mars exploration rovers," *Journal of Field Robotics*, vol. 24, no. 8-9, pp. 651–670, 2007.

[99] C. M. Humphrey, C. Henk, G. Sewell, B. W. Williams, and J. A. Adams, "Assessing the scalability of a multiple robot interface," in *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 239–246, IEEE, 2007.

[100] J. Wang and M. Lewis, "Assessing cooperation in human control of heterogeneous robots,"

in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 9–16, 2008.

[101] D. Schreckenghost, T. Milam, and T. Fong, "Measuring performance in real time during remote human-robot operations with adjustable autonomy," *IEEE Intelligent Systems*, no. 5, pp. 36–45, 2010.

[102] J. A. Saleh and F. Karray, "Towards unified performance metrics for multi-robot human interaction systems," in *International Conference on Autonomous and Intelligent Systems*, pp. 311–320, Springer, 2011.

[103] I.-H. Kuo, C. Jayawardena, E. Broadbent, R. Q. Stafford, and B. A. MacDonald, "Hri evaluation of a healthcare service robot," in *International Conference on Social Robotics*, pp. 178–187, Springer, 2012.

[104] M. Lewis, "Human interaction with multiple remote robots," *Reviews of Human Factors and Ergonomics*, vol. 9, no. 1, pp. 131–174, 2013.

[105] M. Begum, R. W. Serna, D. Kontak, J. Allspaw, J. Kuczynski, H. A. Yanco, and J. Suarez, "Measuring the efficacy of robots in autism therapy: How informative are standard hri metrics'," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pp. 335–342, 2015.

[106] D. Y. Y. Sim and C. K. Loo, "Extensive assessment and evaluation methodologies on assistive social robots for modelling human–robot interaction–a review," *Information Sciences*, vol. 301, pp. 305–344, 2015.

[107] J. Velez, H. Ka, and D. Ding, "Toward developing a framework for standardizing the functional assessment and performance evaluation of assistive robotic manipulators (arms),"

in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 59, pp. 986–990, SAGE Publications Sage CA: Los Angeles, CA, 2015.

[108] A. Aly, S. Griffiths, and F. Stramandinoli, "Metrics and benchmarks in human-robot interaction: Recent advances in cognitive robotics," *Cognitive Systems Research*, vol. 43, pp. 313–323, 2017.

[109] A. Norton, B. Flynn, and H. Yanco, "Implementing human-robot interaction evaluation using standard test methods for response robots," in *Homeland Security and Public Safety: Research, Applications and Standards*, ASTM International, 2019.

[110] L. Bechade, G. Dubuisson-Duplessis, G. Pittaro, M. Garcia, and L. Devillers, "Towards metrics of evaluation of pepper robot as a social companion for the elderly," in *Advanced Social Interaction with Agents*, pp. 89–101, Springer, 2019.

[111] C. L. Bethel and R. R. Murphy, "Review of human studies methods in hri and recommendations," *International Journal of Social Robotics*, vol. 2, no. 4, pp. 347–359, 2010.

[112] N. Maalouf, A. Sidaoui, I. H. Elhajj, and D. Asmar, "Robotics in nursing: A scoping review," *Journal of Nursing Scholarship*, vol. 50, no. 6, pp. 590–600, 2018.

[113] L. Hagan, D. Morin, and R. Lépine, "Evaluation of telenursing outcomes: satisfaction, self-care practices, and cost savings," *Public Health Nursing*, vol. 17, no. 4, pp. 305–313, 2000.

[114] T. L. Chen and C. C. Kemp, "Lead me by the hand: Evaluation of a direct physical interface for nursing assistant robots," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 367–374, IEEE, 2010.

[115] T. Mukai, S. Hirano, H. Nakashima, Y. Kato, Y. Sakaida, S. Guo, and S. Hosoe, "Development of a nursing-care assistant robot riba that can lift a human in its arms," in *2010*

*IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5996–6001, IEEE, 2010.

[116] J. Hu, A. Edsinger, Y.-J. Lim, N. Donaldson, M. Solano, A. Solochek, and R. Marchessault, "An advanced medical robotic system augmenting healthcare capabilities-robotic nursing assistant," in *2011 IEEE international conference on robotics and automation*, pp. 6264–6269, IEEE, 2011.

[117] A. Richert, M. Schiffmann, and C. Yuan, "A nursing robot for social interactions and health assessment," in *International Conference on Applied Human Factors and Ergonomics*, pp. 83–91, Springer, 2019.

[118] J. Li, Z. Li, and K. Hauser, "A study of bidirectionally telepresent tele-action during robot-mediated handover," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2890–2896, IEEE, 2017.

[119] D. Kent, C. Saldanha, and S. Chernova, "Leveraging depth data in remote robot teleoperation interfaces for general object manipulation," *The International Journal of Robotics Research*, vol. 39, no. 1, pp. 39–53, 2020.

[120] A. Hacinecipoglu, E. I. Konukseven, and A. B. Koku, "Evaluation of haptic feedback cues on vehicle teleoperation performance in an obstacle avoidance scenario," in *2013 World Haptics Conference (WHC)*, pp. 689–694, IEEE, 2013.

[121] R. Bhat, V. Pandey, A. K. Rao, and S. Chandra, "An evaluation of cognitive and neural correlates for indirect vision driving and rover teleoperation," in *2017 2nd International Conference on Man and Machine Interfacing (MAMI)*, pp. 1–5, IEEE, 2017.

[122] Y. Wu, P. Balatti, M. Lorenzini, F. Zhao, W. Kim, and A. Ajoudani, "A teleoperation interface

for loco-manipulation control of mobile collaborative robotic assistant," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3593–3600, 2019.

[123] T. Shibata and K. Tanie, "Influence of a priori knowledge in subjective interpretation and evaluation by short-term interaction with mental commit robot," in *Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000)(Cat. No. 00CH37113)*, vol. 1, pp. 169–174, IEEE, 2000.

[124] M. K. Lee, K. P. Tang, J. Forlizzi, and S. Kiesler, "Understanding users! perception of privacy in human-robot interaction," in *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 181–182, IEEE, 2011.

[125] A. H. Memar and E. T. Esfahani, "Physiological measures for human performance analysis in human-robot teamwork: Case of tele-exploration," *IEEE Access*, vol. 6, pp. 3694–3705, 2018.

[126] C. S. Crawford, M. Andujar, and J. E. Gilbert, "Neurophysiological heat maps for human-robot interaction evaluation," in *2017 AAAI Fall Symposium Series*, 2017.

[127] G. Marquart, C. Cabrall, and J. de Winter, "Review of eye-related measures of drivers' mental workload," *Procedia Manufacturing*, vol. 3, pp. 2854–2861, 2015.

[128] M. Schneider and B. Deml, "Analysis of a multimodal human-robot-interface in terms of mental workload," in *Advances in ergonomic design of systems, products and processes*, pp. 247–260, Springer, Berlin, Heidelberg, 2017.

[129] S. C. Bommer and M. Fendley, "A theoretical framework for evaluating mental workload resources in human systems design for manufacturing operations," *International Journal of Industrial Ergonomics*, vol. 63, pp. 7–17, 2018.

[130] M. M. Moniri, F. A. E. Valcarcel, D. Merkel, and D. Sonntag, "Human gaze and focus-of-attention in dual reality human-robot collaboration," in *2016 12th International Conference on Intelligent Environments (IE)*, pp. 238–241, IEEE, 2016.

[131] S. Lemaignan, F. Garcia, A. Jacq, and P. Dillenbourg, "From real-time attention assessment to "with-me-ness" in human-robot interaction," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 157–164, Ieee, 2016.

[132] U. Franke and J. Brynielsson, "Cyber situational awareness–a systematic review of the literature," *Computers & security*, vol. 46, pp. 18–31, 2014.

[133] L. Paletta, A. Dini, C. Murko, S. Yahyanejad, M. Schwarz, G. Lodron, S. Ladstätter, G. Paar, and R. Velik, "Towards real-time probabilistic evaluation of situation awareness from human gaze in human-robot interaction," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 247–248, 2017.

[134] L. Paletta, A. Dini, C. Murko, S. Yahyanejad, and U. Augsdörfer, "Estimation of situation awareness score and performance using eye and head gaze for human-robot collaboration," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pp. 1–3, 2019.

[135] J. Marescaux, J. Leroy, F. Rubino, M. Smith, M. Vix, M. Simone, and D. Mutter, "Transcontinental robot-assisted remote telesurgery: feasibility and potential applications," *Annals of surgery*, vol. 235, no. 4, p. 487, 2002.

[136] S. Liu, Y. Xie, Y. Jia, N. Xi, and Y. Li, "Effect of training on the quality of teleoperator (qot)," in *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, pp. 1928–1933, IEEE, 2015.

[137] C. Ju and H. I. Son, "Evaluation of haptic feedback in the performance of a teleoperated unmanned ground vehicle in an obstacle avoidance scenario," *International Journal of Control, Automation and Systems*, vol. 17, no. 1, pp. 168–180, 2019.

[138] L. Peternel, N. Tsagarakis, and A. Ajoudani, "A human–robot co-manipulation approach based on human sensorimotor information," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 7, pp. 811–822, 2017.

[139] S. L. Delp, F. C. Anderson, A. S. Arnold, P. Loan, A. Habib, C. T. John, E. Guendelman, and D. G. Thelen, "Opensim: open-source software to create and analyze dynamic simulations of movement," *IEEE transactions on biomedical engineering*, vol. 54, no. 11, pp. 1940–1950, 2007.

[140] L. Peternel, C. Fang, N. Tsagarakis, and A. Ajoudani, "A selective muscle fatigue management approach to ergonomic human-robot co-manipulation," *Robotics and Computer-Integrated Manufacturing*, vol. 58, pp. 69–79, 2019.

[141] G. Adamides, C. Katsanos, Y. Parmet, G. Christou, M. Xenos, T. Hadzilacos, and Y. Edan, "Hri usability evaluation of interaction modes for a teleoperated agricultural robotic sprayer," *Applied ergonomics*, vol. 62, pp. 237–246, 2017.

[142] C. Stanton, A. Bogdanovych, and E. Ratanasena, "Teleoperation of a humanoid robot using full-body motion capture, example movements, and machine learning," in *Proc. Australasian Conference on Robotics and Automation*, 2012.

[143] J. Allspaw, J. Roche, A. Norton, and H. Yanco, "Teleoperating a humanoid robot with virtual reality," in *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for Human-Robot Interaction*, 2018.

[144] C. Yang, J. Luo, C. Liu, M. Li, and S.-L. Dai, "Haptics electromyogrphy perception and

learning enhanced intelligence for teleoperated robot," *IEEE Transactions on Automation Science and Engineering*, 2018.

[145] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[146] W. Abdulla, "Mask r-cnn for object detection and instance segmentation on keras and tensorflow." `https://github.com/matterport/Mask_RCNN`, 2017.

[147] J. Kannala, J. Heikkilä, and S. S. Brandt, "Geometric camera calibration," *Wiley Encyclopedia of Computer Science and Engineering*, pp. 1–11, 2007.

[148] Y. Che, H. Culbertson, C.-W. Tang, S. Aich, and A. M. Okamura, "Facilitating human-mobile robot communication via haptic feedback and gesture teleoperation," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 7, no. 3, pp. 1–23, 2018.

[149] S. Chernova and A. L. Thomaz, "Robot learning from human teachers," vol. 8, ch. 10, pp. 1–121, Morgan & Claypool Publishers, 2014.

[150] N. K. Vøllestad, "Measurement of human muscle fatigue," *Journal of neuroscience methods*, vol. 74, no. 2, pp. 219–227, 1997.

[151] V. Gladwell and J. Coote, "Heart rate at the onset of muscle contraction and during passive muscle stretch in humans: a role for mechanoreceptors," *The Journal of physiology*, vol. 540, no. 3, pp. 1095–1102, 2002.

[152] M. Cifrek, V. Medved, S. Tonković, and S. Ostojić, "Surface emg based muscle fatigue evaluation in biomechanics," *Clinical biomechanics*, vol. 24, no. 4, pp. 327–340, 2009.

[153] P. W. Hodges and B. H. Bui, "A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography," *Electroencephalography and*

*Clinical Neurophysiology/Electromyography and Motor Control*, vol. 101, no. 6, pp. 511–519, 1996.

[154] C. E. Boettcher, K. A. Ginn, and I. Cathers, "Standard maximum isometric voluntary contraction tests for normalizing shoulder muscle emg," *Journal of orthopaedic research*, vol. 26, no. 12, pp. 1591–1597, 2008.

[155] C. J. De Luca, "The use of surface electromyography in biomechanics," *Journal of applied biomechanics*, vol. 13, no. 2, pp. 135–163, 1997.

[156] G. V. Dimitrov, T. I. Arabadzhiev, K. N. Mileva, J. L. Bowtell, N. Crichton, and N. A. Dimitrova, "Muscle fatigue during dynamic contractions assessed by new spectral indices," *Medicine and science in sports and exercise*, vol. 38, no. 11, p. 1971, 2006.

[157] E. Rappleye, "Gender ratio of nurses across 50 states," May 2015.

[158] STATISTICSTATS, "Male nursing statistics 2019," August 2019.

[159] NCSBN, "The 2015 National Nursing Workforce Survey," tech. rep., NCSBN, 2016.

[160] J. S. Budden, P. Moulton, K. J. Harper, M. Brunell, and R. Smiley, "The 2015 national nursing workforce survey," *Journal of Nursing Regulation*, vol. 7, no. 1, pp. S1–S90, 2016.

[161] M. C. Linn and A. C. Petersen, "Emergence and characterization of sex differences in spatial ability: A meta-analysis," *Child development*, pp. 1479–1498, 1985.

[162] S. D. Moffat, "Aging and spatial navigation: what do we know and where do we go?," *Neuropsychology review*, vol. 19, no. 4, p. 478, 2009.

[163] N. Paperno, M. Rupp, E. M. Maboudou-Tchao, J. A. A. Smither, and A. Behal, "A Predictive Model for Use of an Assistive Robotic Manipulator: Human Factors Versus Performance in

Pick-and-Place/Retrieval Tasks," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 6, pp. 846–858, 2016.

[164] I. Verner and S. Gamer, "Spatial Learning of Novice Engineering Students Through Practice of Interaction with Robot-Manipulators," in *Online Engineering & Internet of Things*, vol. 22, pp. 359–366, 2018.

[165] M. Draelos, B. Keller, C. Toth, A. Kuo, K. Hauser, and J. Izatt, "Teleoperating robots from arbitrary viewpoints in surgical contexts," in *IEEE International Conference on Intelligent Robots and Systems (IROS)*, (Vancouver), pp. 2549–2555, 2017.

[166] C. Wang, Y. Tian, S. Chen, Z. Tian, T. Jiang, and F. Du, "Predicting performance in manually controlled rendezvous and docking through spatial abilities," *Advances in Space Research*, vol. 53, pp. 362–369, 2014.

[167] A. W. Johnson, K. R. Duda, T. B. Sheridan, and C. M. Oman, "A Closed-Loop Model of Operator Visual Attention, Situation Awareness, and Performance Across Automation Mode Transitions," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, no. 2, pp. 229–241, 2017.

[168] S. J. Czaja, N. Charness, A. D. Fisk, C. Hertzog, S. N. Nair, W. A. Rogers, and J. Sharit, "Factors predicting the use of technology: Findings from the center for research and education on aging and technology enhancement (create).," *Psychology and aging*, vol. 21, no. 2, p. 333, 2006.

[169] R. D. Ellis and J. C. Allaire, "Modeling computer interest in older adults: The role of age, education, computer knowledge, and computer anxiety," *Human Factors*, vol. 41, no. 3, pp. 345–355, 1999.

[170] N. Charness, C. L. Kelley, E. A. Bosman, and M. Mottram, "Word-processing training and

retraining: Effects of adult age, experience, and interface.," *Psychology and aging*, vol. 16, no. 1, p. 110, 2001.

[171] S. J. Cutler, J. Hendricks, and A. Guyer, "Age differences in home computer availability and use," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 58, no. 5, pp. S271–S280, 2003.

[172] J. L. Dyck and J. A.-A. Smither, "Age differences in computer anxiety: The role of computer experience, gender and education," *Journal of educational computing research*, vol. 10, no. 3, pp. 239–248, 1994.

[173] M. Giuliani, M. Scopelliti, and F. Fornara, "Coping strategies and technology in later life," *Companions: Hard Problems and Open Challenges in Robot-Human Interaction*, p. 46, 2005.

[174] M. V. Giuliani, M. Scopelliti, and F. Fornara, "Elderly people at home: technological help in everyday activities," in *IEEE Workshop on Robot and Human Interactive Communication, (RO-MAN)*, (Nashville, US), pp. 365–370, IEEE, 2005.

[175] G. M. Jay and S. L. Willis, "Influence of direct computer experience on older adults' attitudes toward computers," *Journal of Gerontology*, vol. 47, no. 4, pp. P250–P257, 1992.

[176] R. R. Mackie and C. D. Wylie, "Factors influencing acceptance of computer-based innovations," in *Handbook of human-computer interaction*, pp. 1081–1106, Elsevier, 1988.

[177] R. W. Morrell, C. B. Mayhorn, and J. Bennett, "A survey of world wide web use in middle-aged and older adults," *Human Factors*, vol. 42, no. 2, pp. 175–182, 2000.

[178] S. Carty, "Many cars tone deaf to women's voices," *AOL Autos*, 2011.

[179] A. Howard and J. Borenstein, "The ugly truth about ourselves and our robot creations: The problem of bias and social inequity," *Science and Engineering Ethics*, vol. 24, pp. 1521–1536, Oct 2018.

[180] E. C. van Oost, N. Oudshoorn, and T. Pinch, "Materialized gender: how shavers configure the users' feminity and masculinity," *How users matter. The co-construction of users and technology*, pp. 193–208, 2003.

[181] K. Hauser and R. Shaw, "How medical robots will help treat patients in future outbreaks," *IEEE Spectrum*, 2020.

[182] Y. Yoshitake, H. Ue, M. Miyazaki, and T. Moritani, "Assessment of lower-back muscle fatigue using electromyography, mechanomyography, and near-infrared spectroscopy," *European journal of applied physiology*, vol. 84, no. 3, pp. 174–179, 2001.

[183] M. Georgi, C. Amma, and T. Schultz, "Recognizing hand and finger gestures with imu based motion and emg based muscle activity sensing.," in *Biosignals*, pp. 99–108, 2015.

[184] T. Apriantono, H. Nunome, Y. Ikegami, and S. Sano, "The effect of muscle fatigue on instep kicking kinetics and kinematics in association football," *Journal of sports sciences*, vol. 24, no. 9, pp. 951–960, 2006.

[185] A. Takács, D. Á. Nagy, I. Rudas, and T. Haidegger, "Origins of surgical robotics: From space to the operating room," *Acta Polytechnica Hungarica*, vol. 13, no. 1, pp. 13–30, 2016.

[186] L. Peppoloni, F. Brizzi, E. Ruffaldi, and C. A. Avizzano, "Augmented reality-aided telepresence system for robot manipulation in industrial manufacturing," in *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*, pp. 237–240, 2015.

[187] E. Ackerman, "Oculus rift-based system brings true immersion to telepresence robots," April 2015.

[188] A. Kristoffersson, S. Coradeschi, and A. Loutfi, "A review of mobile robotic telepresence," *Advances in Human-Computer Interaction*, vol. 2013, 2013.

[189] T. N. Nguyen, H. T. Nguyen, *et al.*, "Real-time video streaming with multi-camera for a telepresence wheelchair," in *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 1–5, IEEE, 2016.

[190] D. Whitney, E. Rosen, D. Ullman, E. Phillips, and S. 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)*, pp. 1–9, IEEE, 2018.

[191] J. de León, M. Garzón, D. Garzón, E. Narváez, J. del Cerro, and A. Barrientos, "From video games multiple cameras to multi-robot teleoperation in disaster scenarios," in *2016 International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pp. 323–328, IEEE, 2016.

[192] S. H. Seo, D. J. Rea, J. Wiebe, and J. E. Young, "Monocle: interactive detail-in-context using two pan-and-tilt cameras to improve teleoperation effectiveness," in *2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pp. 962–967, IEEE, 2017.

[193] M. Barnes, L. R. Elliott, J. Wright, A. Scharine, and J. Chen, "Human-robot interaction design research: From teleoperations to human-agent teaming," tech. rep., CCDC Army Research Laboratory Aberdeen Proving Ground United States, 2019.

[194] L. Almeida, P. Menezes, and J. Dias, "Interface transparency issues in teleoperation," *Applied Sciences*, vol. 10, no. 18, p. 6232, 2020.

[195] T. Huk, "Who benefits from learning with 3d models? the case of spatial ability," *Journal of computer assisted learning*, vol. 22, no. 6, pp. 392–404, 2006.

[196] T. Piumsomboon, A. Day, B. Ens, Y. Lee, G. Lee, and M. Billinghurst, "Exploring enhancements for remote mixed reality collaboration," in *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, pp. 1–5, 2017.

[197] N. Rudigkeit and M. Gebhard, "Amicus—a head motion-based interface for control of an assistive robot," *Sensors*, vol. 19, no. 12, p. 2836, 2019.

[198] D. Nicolis, M. Palumbo, A. M. Zanchettin, and P. Rocco, "Occlusion-free visual servoing for the shared autonomy teleoperation of dual-arm robots," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 796–803, 2018.

[199] S. Radmard, A. Moon, and E. A. Croft, "Impacts of visual occlusion and its resolution in robot-mediated social collaborations," *International Journal of Social Robotics*, vol. 11, no. 1, pp. 105–121, 2019.

[200] G. Yang, S. Wang, J. Yang, and B. Shen, "Viewpoint selection strategy for a life support robot," in *2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR)*, pp. 82–87, IEEE, 2018.

[201] B. Calli, W. Caarls, M. Wisse, and P. Jonker, "Viewpoint optimization for aiding grasp synthesis algorithms using reinforcement learning," *Advanced Robotics*, vol. 32, no. 20, pp. 1077–1089, 2018.

[202] T. Patten, M. Zillich, R. Fitch, M. Vincze, and S. Sukkarieh, "Viewpoint evaluation for online 3-d active object classification," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 73–81, 2015.

[203] A. Saran, B. Lakic, S. Majumdar, J. Hess, and S. Niekum, "Viewpoint selection for visual failure detection," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5437–5444, IEEE, 2017.

[204] H. Cash and T. J. Prescott, "Improving the visual comfort of virtual reality telepresence for robotics," in *International Conference on Social Robotics*, pp. 697–706, Springer, 2019.

[205] K. Carnegie and T. Rhee, "Reducing visual discomfort with hmds using dynamic depth of field," *IEEE computer graphics and applications*, vol. 35, no. 5, pp. 34–41, 2015.

[206] J. Sandoval, M. A. Laribi, and S. Zeghloul, "Autonomous robot-assistant camera holder for minimally invasive surgery," in *IFToMM International Symposium on Robotics and Mechatronics*, pp. 465–472, Springer, 2019.

[207] M. Ito and K. Sekiyama, "Optimal viewpoint selection for cooperative visual assistance in multi-robot systems," in *2015 IEEE/SICE International Symposium on System Integration (SII)*, pp. 605–610, IEEE, 2015.

[208] G. Leifman, E. Shtrom, and A. Tal, "Surface regions of interest for viewpoint selection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 12, pp. 2544–2556, 2016.

[209] C. Gebhardt, S. Stevšić, and O. Hilliges, "Optimizing for aesthetically pleasing quadrotor camera motion," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, pp. 1–11, 2018.

[210] K. Lan and K. Sekiyama, "Autonomous robot photographer system based on aesthetic composition evaluation using yohaku," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 101–106, IEEE, 2019.

[211] C. Huang, F. Gao, J. Pan, Z. Yang, W. Qiu, P. Chen, X. Yang, S. Shen, and K.-T. Cheng, "Act: An autonomous drone cinematography system for action scenes," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7039–7046, IEEE, 2018.

[212] Y. Tao, Q. Wang, W. Chen, Y. Wu, and H. Lin, "Similarity voting based viewpoint selection

for volumes," in *Computer graphics forum*, vol. 35, pp. 391–400, Wiley Online Library, 2016.

[213] T. Abe, N. Raison, N. Shinohara, M. S. Khan, K. Ahmed, and P. Dasgupta, "The effect of visual-spatial ability on the learning of robot-assisted surgical skills," *Journal of surgical education*, vol. 75, no. 2, pp. 458–464, 2018.

[214] A. Shafti, P. Orlov, and A. A. Faisal, "Gaze-based, context-aware robotic system for assisted reaching and grasping," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 863–869, IEEE, 2019.

[215] R. Lokesh and R. Ranganathan, "Haptic assistance that restricts the use of redundant solutions is detrimental to motor learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2020.

[216] M. M. Hayhoe, "Vision and action," *Annual review of vision science*, vol. 3, pp. 389–413, 2017.

[217] C.-L. Li, M. P. Aivar, D. M. Kit, M. H. Tong, and M. M. Hayhoe, "Memory and visual search in naturalistic 2d and 3d environments," *Journal of vision*, vol. 16, no. 8, pp. 9–9, 2016.

[218] D. K. Das, M. Laha, S. Majumder, and D. Ray, "Stable and consistent object tracking: An active vision approach," in *Advanced Computational and Communication Paradigms*, pp. 299–308, Springer, 2018.

[219] F. R. Danion and J. R. Flanagan, "Different gaze strategies during eye versus hand tracking of a moving target," *Scientific reports*, vol. 8, no. 1, pp. 1–9, 2018.

[220] E. Rezunenko, K. van der El, D. M. Pool, M. M. van Paassen, and M. Mulder, "Relating human gaze and manual control behavior in preview tracking tasks with spatial occlusion,"

in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3440–3445, IEEE, 2018.

[221] H. A. Trukenbrod, S. Barthelmé, F. A. Wichmann, and R. Engbert, "Spatial statistics for gaze patterns in scene viewing: effects of repeated viewing," *Journal of vision*, vol. 19, no. 6, pp. 5–5, 2019.

[222] J. S. Diamond, D. M. Wolpert, and J. R. Flanagan, "Rapid target foraging with reach or gaze: The hand looks further ahead than the eye," *PLoS computational biology*, vol. 13, no. 7, p. e1005504, 2017.

[223] N. Abekawa and H. Gomi, "Online gain update for manual following response accompanied by gaze shift during arm reaching," *Journal of neurophysiology*, vol. 113, no. 4, pp. 1206–1216, 2015.

[224] J. R. Kuntz, J. M. Karl, J. B. Doan, M. Grohs, and I. Q. Whishaw, "Two types of memory-based (pantomime) reaches distinguished by gaze anchoring in reach-to-grasp tasks," *Behavioural brain research*, vol. 381, p. 112438, 2020.

[225] M. R. Thulasiram, R. W. Langridge, H. H. Abbas, and J. J. Marotta, "Eye–hand coordination in reaching and grasping vertically moving targets," *Experimental brain research*, vol. 238, pp. 1433–1440, 2020.

[226] S. K. Allani, B. John, J. Ruiz, S. Dixit, J. Carter, C. Grimm, and R. Balasubramanian, "Evaluating human gaze patterns during grasping tasks: robot versus human hand," in *Proceedings of the ACM Symposium on Applied Perception*, pp. 45–52, 2016.

[227] J. A. Navia, M. Dicks, J. van der Kamp, and L. M. Ruiz, "Gaze control during interceptive actions with different spatiotemporal demands.," *Journal of experimental psychology: human perception and performance*, vol. 43, no. 4, p. 783, 2017.

[228] H. Zhao, D. Straub, and C. A. Rothkopf, "The visual control of interceptive steering: How do people steer a car to intercept a moving target?," *Journal of vision*, vol. 19, no. 14, pp. 11–11, 2019.

[229] S. P. Swinnen, Y. Li, N. Wenderoth, N. Dounskaia, W. Byblow, C. Stinear, and J. Wagemans, "Perception—action coupling during bimanual coordination: The role of visual perception in the coalition of constraints that govern bimanual action," *Journal of motor behavior*, vol. 36, no. 4, pp. 394–398, 2004.

[230] J. Kurz, M. Hegele, M. Reiser, and J. Munzert, "Impact of task difficulty on gaze behavior in a sequential object manipulation task," *Experimental brain research*, vol. 235, no. 11, pp. 3479–3486, 2017.

[231] S. Barton, S. Steinmetz, G. Diaz, J. Matthis, and B. Fajen, "The visual control of walking over complex terrain with flat versus raised obstacles," *Journal of Vision*, vol. 17, no. 10, pp. 707–707, 2017.

[232] J. Dominguez-Zamora, S. Gunn, and D. Marigold, "Does uncertainty about the terrain explain gaze behavior during visually guided walking?," *Journal of Vision*, vol. 17, no. 10, pp. 709–709, 2017.

[233] F. C. Fortenbaugh, J. C. Hicks, L. Hao, and K. A. Turano, "High-speed navigators: Using more than what meets the eye," *Journal of Vision*, vol. 6, no. 5, pp. 3–3, 2006.

[234] S. N. Hamid, B. Stankiewicz, and M. Hayhoe, "Gaze patterns in navigation: Encoding information in large-scale environments," *Journal of Vision*, vol. 10, no. 12, pp. 28–28, 2010.

[235] K. L. Macuga, A. C. Beall, R. S. Smith, and J. M. Loomis, "Visual control of steering in curve driving," *Journal of vision*, vol. 19, no. 5, pp. 1–1, 2019.

[236] J. Ibbotson, C. MacKenzie, C. Cao, and A. Lomax, "Gaze patterns in laparoscopic surgery," *Studies in Health Technology and Informatics*, pp. 154–160, 1999.

[237] J. L. Griffith, P. Voloschin, G. D. Gibb, and J. R. Bailey, "Differences in eye-hand motor coordination of video-game users and non-users," *Perceptual and motor skills*, vol. 57, no. 1, pp. 155–158, 1983.

[238] B. I. Bertenthal, J. L. Rose, and D. L. Bai, "Perception–action coupling in the development of visual control of posture.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 23, no. 6, p. 1631, 1997.

[239] B. W. Tatler, M. M. Hayhoe, M. F. Land, and D. H. Ballard, "Eye guidance in natural vision: Reinterpreting salience," *Journal of vision*, vol. 11, no. 5, pp. 5–5, 2011.

[240] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vision research*, vol. 49, no. 10, pp. 1295–1306, 2009.

[241] J. Jovancevic-Misic and M. Hayhoe, "Adaptive gaze control in natural environments," *Journal of Neuroscience*, vol. 29, no. 19, pp. 6234–6238, 2009.

[242] J. S. Matthis and B. R. Fajen, "Humans exploit the biomechanics of bipedal gait during visually guided walking over complex terrain," *Proceedings of the Royal Society B: Biological Sciences*, vol. 280, no. 1762, p. 20130700, 2013.

[243] V. Navalpakkam, C. Koch, A. Rangel, and P. Perona, "Optimal reward harvesting in complex perceptual environments," *Proceedings of the National Academy of Sciences*, vol. 107, no. 11, pp. 5232–5237, 2010.

[244] A. C. Schütz, J. Trommershäuser, and K. R. Gegenfurtner, "Dynamic integration of information about salience and value for saccadic eye movements," *Proceedings of the National Academy of Sciences*, vol. 109, no. 19, pp. 7547–7552, 2012.

[245] M. H. Tong, O. Zohar, and M. M. Hayhoe, "Control of gaze while walking: task structure, reward, and uncertainty," *Journal of vision*, vol. 17, no. 1, pp. 28–28, 2017.

[246] S. Ghahghaei and P. Verghese, "Efficient saccade planning requires time and clear choices," *Vision research*, vol. 113, pp. 125–136, 2015.

[247] P. Han, D. R. Saunders, R. L. Woods, and G. Luo, "Trajectory prediction of saccadic eye movements using a compressed exponential model," *Journal of vision*, vol. 13, no. 8, pp. 27–27, 2013.

[248] G. Diaz, J. Cooper, C. Rothkopf, and M. Hayhoe, "Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task," *Journal of vision*, vol. 13, no. 1, pp. 20–20, 2013.

[249] G. Diaz, J. Cooper, and M. Hayhoe, "Memory and prediction in natural gaze control," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 368, no. 1628, p. 20130064, 2013.

[250] D. M. Wolpert and M. S. Landy, "Motor control is decision-making," *Current opinion in neurobiology*, vol. 22, no. 6, pp. 996–1003, 2012.

[251] D. Liu and E. Todorov, "Evidence for the flexible sensorimotor strategies predicted by optimal feedback control," *Journal of Neuroscience*, vol. 27, no. 35, pp. 9354–9368, 2007.

[252] D. C. Knill, A. Bondada, and M. Chhabra, "Flexible, task-dependent use of sensory feedback to control hand movements," *Journal of Neuroscience*, vol. 31, no. 4, pp. 1219–1237, 2011.

[253] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 185–207, 2012.

[254] M. M. Hayhoe and J. S. Matthis, "Control of gaze in natural environments: effects of rewards and costs, uncertainty and memory in target selection," *Interface focus*, vol. 8, no. 4, p. 20180009, 2018.

[255] W. F. Helsen, D. Elliott, J. L. Starkes, and K. L. Ricker, "Coupling of eye, finger, elbow, and shoulder movements during manual aiming," *Journal of motor behavior*, vol. 32, no. 3, pp. 241–248, 2000.

[256] D. Elliott, W. F. Helsen, and R. Chua, "A century later: Woodworth's (1899) two-component model of goal-directed aiming.," *Psychological bulletin*, vol. 127, no. 3, p. 342, 2001.

[257] R. Bajcsy, Y. Aloimonos, and J. K. Tsotsos, "Revisiting active perception," *Autonomous Robots*, vol. 42, no. 2, pp. 177–196, 2018.

[258] A. B. Yu and J. M. Zacks, "Transformations and representations supporting spatial perspective taking," *Spatial Cognition & Computation*, vol. 17, no. 4, pp. 304–337, 2017.

[259] J. L. Wright, J. Y. Chen, and M. J. Barnes, "Human–automation interaction for multiple robot control: the effect of varying automation assistance and individual differences on operator performance," *Ergonomics*, vol. 61, no. 8, pp. 1033–1045, 2018.

[260] B. P. DeJong, J. E. Colgate, and M. A. Peshkin, "Improving teleoperation: reducing mental rotations and translations," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, vol. 4, pp. 3708–3714, IEEE, 2004.

[261] C. Wang, Y. Tian, S. Chen, Z. Tian, T. Jiang, and F. Du, "Predicting performance in manually controlled rendezvous and docking through spatial abilities," *Advances in Space Research*, vol. 53, no. 2, pp. 362–369, 2014.

[262] G. Desmarais, M. Meade, T. Wells, and M. Nadeau, "Visuo-haptic integration in object

identification using novel objects," *Attention, Perception, & Psychophysics*, vol. 79, no. 8, pp. 2478–2498, 2017.

[263] E. G. Roelofsen, J. Bosga, D. A. Rosenbaum, M. W. Nijhuis-van der Sanden, W. Hullegie, R. van Cingel, and R. G. Meulenbroek, "Haptic feedback helps bipedal coordination," *Experimental brain research*, vol. 234, no. 10, pp. 2869–2881, 2016.

[264] C. Bozzacchi, R. Volcic, and F. Domini, "Grasping in absence of feedback: systematic biases endure extensive training," *Experimental brain research*, vol. 234, no. 1, pp. 255–265, 2016.

[265] M. W. Wijntjes, R. Volcic, S. C. Pont, J. J. Koenderink, and A. M. Kappers, "Haptic perception disambiguates visual perception of 3d shape," *Experimental brain research*, vol. 193, no. 4, pp. 639–644, 2009.

[266] S. Lacey and K. Sathian, "Visuo-haptic object perception," *Multisensory Perception*, pp. 157–178, 2020.

[267] A. Sengül, G. Rognini, M. van Elk, J. E. Aspell, H. Bleuler, and O. Blanke, "Force feedback facilitates multisensory integration during robotic tool use," *Experimental brain research*, vol. 227, no. 4, pp. 497–507, 2013.

[268] F. G. Hamza-Lup, C. M. Bogdan, D. M. Popovici, and O. D. Costea, "A survey of visuo-haptic simulation in surgical training," *arXiv preprint arXiv:1903.03272*, 2019.

[269] R. Volcic and I. Camponogara, "How do vision and haptics combine in multisensory grasping?," *Journal of Vision*, vol. 18, no. 10, pp. 64–64, 2018.

[270] C. W. Nielsen, M. A. Goodrich, and R. W. Ricks, "Ecological interfaces for improving mobile robot teleoperation," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 927–941, 2007.

[271] "Logitech c310 hd webcam, 720p video with noise reducing mic."

[272] "Aw615 webcam."

[273] P. Praveena, L. Molina, Y. Wang, E. Senft, B. Mutlu, and M. Gleicher, "Understanding control frames in multi-camera robot telemanipulation," in *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 432–440, 2022.

[274] R. Reilink, G. de Bruin, M. Franken, M. A. Mariani, S. Misra, and S. Stramigioli, "Endoscopic camera control by head movements for thoracic surgery," in *2010 3rd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics*, pp. 510–515, IEEE, 2010.

[275] C. Carreto, D. Gêgo, and L. Figueiredo, "An eye-gaze tracking system for teleoperation of a mobile robot," *Journal of Information Systems Engineering & Management*, vol. 3, no. 2, p. 16, 2018.

[276] A. Roncone, U. Pattacini, G. Metta, and L. Natale, "A cartesian 6-dof gaze controller for humanoid robots.," in *Robotics: science and systems*, vol. 2016, 2016.

[277] S. J. Vine, R. S. Masters, J. S. McGrath, E. Bright, and M. R. Wilson, "Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills," *Surgery*, vol. 152, no. 1, pp. 32–40, 2012.

[278] P. Robotics, "A new mobile base offers seamless and self-detecting navigation to the robot.," 2021.

[279] M. Boyer, M. L. Cummings, L. B. Spence, and E. T. Solovey, "Investigating mental workload changes in a long duration supervisory control task," *Interacting with Computers*, vol. 27, no. 5, pp. 512–520, 2015.

[280] L. Xiong, C. B. Chng, C. K. Chui, P. Yu, and Y. Li, "Shared control of a medical robot with

haptic guidance," *International journal of computer assisted radiology and surgery*, vol. 12, no. 1, pp. 137–147, 2017.

[281] J. Aleotti, G. Micconi, S. Caselli, G. Benassi, N. Zambelli, M. Bettelli, and A. Zappettini, "Detection of nuclear sources by uav teleoperation using a visuo-haptic augmented reality interface," *Sensors*, vol. 17, no. 10, p. 2234, 2017.

[282] T.-C. Lin, A. U. Krishnan, and Z. Li, "Shared autonomous interface for reducing physical effort in robot teleoperation via human motion mapping," in *to appear in the 2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[283] H. Boessenkool, D. A. Abbink, C. J. Heemskerk, F. C. van der Helm, and J. G. Wildenbeest, "A task-specific analysis of the benefit of haptic shared control during telemanipulation," *IEEE Transactions on Haptics*, vol. 6, no. 1, pp. 2–12, 2012.

[284] Z. Makhataeva and H. A. Varol, "Augmented reality for robotics: A review," *Robotics*, vol. 9, no. 2, p. 21, 2020.

[285] D. P. Losey, C. G. McDonald, E. Battaglia, and M. K. O'Malley, "A review of intent detection, arbitration, and communication aspects of shared control for physical human–robot interaction," *Applied Mechanics Reviews*, vol. 70, no. 1, 2018.

[286] A. K. Tanwani and S. Calinon, "A generative model for intention recognition and manipulation assistance in teleoperation," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 43–50, IEEE, 2017.

[287] A. Hosseini, F. Richthammer, and M. Lienkamp, "Predictive haptic feedback for safe lateral control of teleoperated road vehicles in urban areas," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, pp. 1–7, IEEE, 2016.

[288] M. Corno, L. D'avico, G. Panzani, and S. M. Savaresi, "A haptic-based, safety-oriented, braking assistance system for road bicycles," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1189–1194, IEEE, 2017.

[289] R. Dang, J. Wang, S. E. Li, and K. Li, "Coordinated adaptive cruise control system with lane-change assistance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2373–2383, 2015.

[290] L. Profumo, L. Pollini, and D. A. Abbink, "Direct and indirect haptic aiding for curve negotiation," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1846–1852, IEEE, 2013.

[291] A. Balachandran, M. Brown, S. M. Erlien, and J. C. Gerdes, "Predictive haptic feedback for obstacle avoidance based on model predictive control," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 1, pp. 26–31, 2015.

[292] M. Itoh, T. Inagaki, and H. Tanaka, "Haptic steering direction guidance for pedestrian-vehicle collision avoidance," in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3327–3332, IEEE, 2012.

[293] D. Powell and M. K. O'Malley, "The task-dependent efficacy of shared-control haptic guidance paradigms," *IEEE transactions on haptics*, vol. 5, no. 3, pp. 208–219, 2012.

[294] C. Masone, P. R. Giordano, H. H. Bülthoff, and A. Franchi, "Semi-autonomous trajectory generation for mobile robots with integral haptic shared control," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6468–6475, IEEE, 2014.

[295] A. Kucukyilmaz and Y. Demiris, "Learning shared control by demonstration for personalized wheelchair assistance," *IEEE transactions on haptics*, vol. 11, no. 3, pp. 431–442, 2018.

[296] S. Scheggi, M. Aggravi, F. Morbidi, and D. Prattichizzo, "Cooperative human-robot haptic navigation," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2693–2698, IEEE, 2014.

[297] H. Saeidi, J. R. Wagner, and Y. Wang, "A mixed-initiative haptic teleoperation strategy for mobile robotic systems based on bidirectional computational trust analysis," *ieee Transactions on Robotics*, vol. 33, no. 6, pp. 1500–1507, 2017.

[298] S. A. Bowyer, B. L. Davies, and F. R. y Baena, "Active constraints/virtual fixtures: A survey," *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 138–157, 2013.

[299] C. P. Quintero, M. Dehghan, O. Ramirez, M. H. Ang, and M. Jagersand, "Flexible virtual fixture interface for path specification in tele-manipulation," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5363–5368, IEEE, 2017.

[300] A. Franchi, C. Secchi, M. Ryll, H. H. Bulthoff, and P. R. Giordano, "Shared control: Balancing autonomy and human assistance with a group of quadrotor uavs," *IEEE Robotics & Automation Magazine*, vol. 19, no. 3, pp. 57–68, 2012.

[301] H. I. Son, A. Franchi, L. L. Chuang, J. Kim, H. H. Bulthoff, and P. R. Giordano, "Human-centered design and evaluation of haptic cueing for teleoperation of multiple mobile robots," *IEEE transactions on cybernetics*, vol. 43, no. 2, pp. 597–609, 2013.

[302] D. Sieber, S. Musić, and S. Hirche, "Multi-robot manipulation controlled by a human with haptic feedback," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2440–2446, IEEE, 2015.

[303] S. Saliceti, J. Ortiz, A. Cardellino, L. Rossi, and J.-G. Fontaine, "Fusion of tactile sensing and haptic feedback for unknown object identification aimed to tele-manipulation," in *2010 IEEE Conference on Multisensor Fusion and Integration*, pp. 205–210, IEEE, 2010.

[304] M. L. Rusch, M. C. Schall Jr, J. D. Lee, J. D. Dawson, and M. Rizzo, "Augmented reality cues to assist older drivers with gap estimation for left-turns," *Accident Analysis & Prevention*, vol. 71, pp. 210–221, 2014.

[305] M. L. Rusch, M. C. Schall Jr, P. Gavin, J. D. Lee, J. D. Dawson, S. Vecera, and M. Rizzo, "Directing driver attention with augmented reality cues," *Transportation research part F: traffic psychology and behaviour*, vol. 16, pp. 127–137, 2013.

[306] M. T. Phan, I. Thouvenin, and V. Frémont, "Enhancing the driver awareness of pedestrian using augmented reality cues," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1298–1304, IEEE, 2016.

[307] K. R. Dillman, T. T. H. Mok, A. Tang, L. Oehlberg, and A. Mitchell, "A visual interaction cue framework from video game environments for augmented reality," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[308] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 316–324, 2018.

[309] M. Zolotas, J. Elsdon, and Y. Demiris, "Head-mounted augmented reality for explainable robotic wheelchair assistance," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1823–1829, IEEE, 2018.

[310] T. Kassuba, C. Klinge, C. Hölig, B. Röder, and H. R. Siebner, "Vision holds a greater share in visuo-haptic object recognition than touch," *Neuroimage*, vol. 65, pp. 59–68, 2013.

[311] T. R. Schneider, G. Buckingham, and J. Hermsdörfer, "Visual cues, expectations, and sensorimotor memories in the prediction and perception of object dynamics during manipulation," *Experimental brain research*, vol. 238, no. 2, pp. 395–409, 2020.

[312] B. Corbett, C. S. Nam, and T. Yamaguchi, "The effects of haptic feedback and visual distraction on pointing task performance," *International Journal of Human-Computer Interaction*, vol. 32, no. 2, pp. 89–102, 2016.

[313] G. D. Park and C. L. Reed, "Haptic over visual information in the distribution of visual attention after tool-use in near and far space," *Experimental brain research*, vol. 233, no. 10, pp. 2977–2988, 2015.

[314] C. Takahashi and S. J. Watt, "Optimal visual–haptic integration with articulated tools," *Experimental brain research*, vol. 235, no. 5, pp. 1361–1373, 2017.

[315] D. Hecht and M. Reiner, "Sensory dominance in combinations of audio, visual and haptic stimuli," *Experimental brain research*, vol. 193, no. 2, pp. 307–314, 2009.

[316] Y. Zheng and J. B. Morrell, "Comparison of visual and vibrotactile feedback methods for seated posture guidance," *IEEE transactions on haptics*, vol. 6, no. 1, pp. 13–23, 2012.

[317] A. Talasaz, A. L. Trejos, and R. V. Patel, "The role of direct and visual force feedback in suturing using a 7-dof dual-arm teleoperated system," *IEEE transactions on haptics*, vol. 10, no. 2, pp. 276–287, 2016.

[318] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review," *Psychonomic bulletin & review*, vol. 20, no. 1, pp. 21–53, 2013.

[319] G. Rauter, R. Sigrist, R. Riener, and P. Wolf, "Learning of temporal and spatial movement aspects: A comparison of four types of haptic control and concurrent visual feedback," *IEEE transactions on haptics*, vol. 8, no. 4, pp. 421–433, 2015.

[320] M. Ewerton, D. Rother, J. Weimar, G. Kollegger, J. Wiemeyer, J. Peters, and G. Maeda,

"Assisting movement training and execution with visual and haptic feedback," *Frontiers in neurorobotics*, vol. 12, p. 24, 2018.

[321] V. Girbés-Juan, V. Schettino, Y. Demiris, and J. Tornero, "Haptic and visual feedback assistance for dual-arm robot teleoperation in surface conditioning tasks," *IEEE Transactions on Haptics*, vol. 14, no. 1, pp. 44–56, 2020.

[322] D. Ni, A. Yew, S. Ong, and A. Nee, "Haptic and visual augmented reality interface for programming welding robots," *Advances in Manufacturing*, vol. 5, no. 3, pp. 191–198, 2017.

[323] M. Li, J. Konstantinova, E. L. Secco, A. Jiang, H. Liu, T. Nanayakkara, L. D. Seneviratne, P. Dasgupta, K. Althoefer, and H. A. Wurdemann, "Using visual cues to enhance haptic feedback for palpation on virtual model of soft tissue," *Medical & biological engineering & computing*, vol. 53, no. 11, pp. 1177–1186, 2015.

[324] Z. Wang, R. Zheng, T. Kaizuka, and K. Nakano, "The effect of haptic guidance on driver steering performance during curve negotiation with limited visual feedback," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 600–605, IEEE, 2017.

[325] N. Pedemonte, F. Abi-Farraj, and P. R. Giordano, "Visual-based shared control for remote telemanipulation with integral haptic feedback," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5342–5349, IEEE, 2017.

[326] V. Ho, C. Borst, M. M. van Paassen, and M. Mulder, "Increasing acceptance of haptic feedback in uav teleoperation by visualizing force fields," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3027–3032, IEEE, 2018.

[327] L. Meli, C. Pacchierotti, G. Salvietti, F. Chinello, M. Maisto, A. De Luca, and D. Prattichizzo, "Combining wearable finger haptics and augmented reality: User evaluation using

an external camera and the microsoft hololens," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4297–4304, 2018.

[328] A. Hong, H. H. Bülthoff, and H. I. Son, "A visual and force feedback for multi-robot teleoperation in outdoor environments: A preliminary result," in *2013 IEEE International Conference on Robotics and Automation*, pp. 1471–1478, IEEE, 2013.

[329] R. J. Kuiper, D. J. Heck, I. A. Kuling, and D. A. Abbink, "Evaluation of haptic and visual cues for repulsive or attractive guidance in nonholonomic steering tasks," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 5, pp. 672–683, 2016.

[330] J. S. Lee, Y. Ham, H. Park, and J. Kim, "Challenges, tasks, and opportunities in teleoperation of excavator toward human-in-the-loop construction automation," *Automation in Construction*, vol. 135, p. 104119, 2022.

[331] D. Crestani, K. Godary-Dejean, and L. Lapierre, "Enhancing fault tolerance of autonomous mobile robots," *Robotics and Autonomous Systems*, vol. 68, pp. 140–155, 2015.

[332] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge, "A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments," in *Computer graphics forum*, vol. 38, pp. 21–45, Wiley Online Library, 2019.

[333] T. Zhou, M. E. Cabrera, J. P. Wachs, T. Low, and C. Sundaram, "A comparative study for telerobotic surgery using free hand gestures," *Journal of Human-Robot Interaction*, vol. 5, no. 2, pp. 1–28, 2016.

[334] W. Pryor, B. P. Vagvolgyi, A. Deguet, S. Leonard, L. L. Whitcomb, and P. Kazanzides, "Interactive planning and supervised execution for high-risk, high-latency teleoperation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1857–1864, IEEE, 2020.

[335] P. C. Gloumeau, W. Stuerzlinger, and J. Han, "Pinnpivot: Object manipulation using pins in immersive virtual environments," *IEEE transactions on visualization and computer graphics*, vol. 27, no. 4, pp. 2488–2494, 2020.

[336] J. Guo, C. Liu, and P. Poignet, "A scaled bilateral teleoperation system for robotic-assisted surgery with time delay," *Journal of Intelligent & Robotic Systems*, vol. 95, no. 1, pp. 165–192, 2019.

[337] S. Frees, G. D. Kessler, and E. Kay, "Prism interaction for enhancing control in immersive virtual environments," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 14, no. 1, pp. 2–es, 2007.

[338] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu, "A handle bar metaphor for virtual object manipulation with mid-air interaction," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1297–1306, 2012.

[339] G. Du, G. Yao, C. Li, and P. X. Liu, "Natural human–robot interface using adaptive tracking system with the unscented kalman filter," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 1, pp. 42–54, 2019.

[340] E. You and K. Hauser, "Assisted teleoperation strategies for aggressively controlling a robot arm with 2d input," in *Robotics: science and systems*, vol. 7, p. 354, MIT Press USA, 2012.

[341] M. E. Cabrera, K. Dey, K. Krishnaswamy, T. Bhattacharjee, and M. Cakmak, "Cursor-based robot tele-manipulation through 2d-to-se2 interfaces," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4230–4237, IEEE.

[342] A. L. Orekhov, C. B. Black, J. Till, S. Chung, and D. C. Rucker, "Analysis and validation of a teleoperated surgical parallel continuum manipulator," *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 828–835, 2016.

[343] J. E. Solanes, A. Muñoz, L. Gracia, A. Martí, V. Girbés-Juan, and J. Tornero, "Teleoperation of industrial robot manipulators based on augmented reality," *The International Journal of Advanced Manufacturing Technology*, vol. 111, no. 3, pp. 1077–1097, 2020.

[344] S. S. White, K. W. Bisland, M. C. Collins, and Z. Li, "Design of a high-level teleoperation interface resilient to the effects of unreliable robot autonomy," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 11519–11524, IEEE, 2020.

[345] M. Draelos, B. Keller, C. Toth, A. Kuo, K. Hauser, and J. Izatt, "Teleoperating robots from arbitrary viewpoints in surgical contexts," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2549–2555, IEEE, 2017.

[346] M. Ewerton, O. Arenz, and J. Peters, "Assisted teleoperation in changing environments with a mixture of virtual guides," *Advanced Robotics*, vol. 34, no. 18, pp. 1157–1170, 2020.

[347] M. Aggravi, D. A. Estima, A. Krupa, S. Misra, and C. Pacchierotti, "Haptic teleoperation of flexible needles combining 3d ultrasound guidance and needle tip force feedback," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4859–4866, 2021.

[348] S. DiMaio, M. Hanuschik, and U. Kreaden, "The da vinci surgical system," in *Surgical robotics*, pp. 199–217, Springer, 2011.

[349] "Da vinci surgeon console." `https://www.intuitive.com/en-us/products-and-services/da-vinci/systems`. Accessed: 2022-04-28.

[350] H. Beik-Mohammadi, M. Kerzel, B. Pleintinger, T. Hulin, P. Reisich, A. Schmidt, A. Pereira, S. Wermter, and N. Y. Lii, "Model mediated teleoperation with a hand-arm exoskeleton in long time delays using reinforcement learning," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 713–720, IEEE, 2020.

[351] B. Fang, D. Guo, F. Sun, H. Liu, and Y. Wu, "A robotic hand-arm teleoperation system using human arm/hand with a novel data glove," in *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 2483–2488, IEEE, 2015.

[352] M. Li, Y. Zhuo, J. Chen, B. He, G. Xu, J. Xie, X. Zhao, and W. Yao, "Design and performance characterization of a soft robot hand with fingertip haptic feedback for teleoperation," *Advanced Robotics*, vol. 34, no. 23, pp. 1491–1505, 2020.

[353] S. Park, Y. Jung, and J. Bae, "A tele-operation interface with a motion capture system and a haptic glove," in *2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pp. 544–549, IEEE, 2016.

[354] M. Macchini, T. Havy, A. Weber, F. Schiano, and D. Floreano, "Hand-worn haptic interface for drone teleoperation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 10212–10218, IEEE, 2020.

[355] A. Eilering, G. Franchi, and K. Hauser, "Robopuppet: Low-cost, 3d printed miniatures for teleoperating full-size robots," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1248–1254, IEEE, 2014.

[356] L. Ke, A. Kamat, J. Wang, T. Bhattacharjee, C. Mavrogiannis, and S. S. Srinivasa, "Telemanipulation with chopsticks: Analyzing human factors in user demonstrations," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 11539–11546, IEEE, 2020.

[357] S. Sakr, T. Daunizeau, D. Reversat, S. Régnier, and S. Haliyo, "An ungrounded master device for tele-microassembly," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–9, IEEE, 2018.

[358] E. Bozgeyikli and L. L. Bozgeyikli, "Evaluating object manipulation interaction techniques

in mixed reality: Tangible user interfaces and gesture," in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 778–787, IEEE, 2021.

[359] T. Stoyanov, R. Krug, A. Kiselev, D. Sun, and A. Loutfi, "Assisted telemanipulation: A stack-of-tasks approach to remote manipulator control," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–9, IEEE, 2018.

[360] D. Dajles, F. Siles, *et al.*, "Teleoperation of a humanoid robot using an optical motion capture system," in *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, pp. 1–8, IEEE, 2018.

[361] A. Handa, K. Van Wyk, W. Yang, J. Liang, Y.-W. Chao, Q. Wan, S. Birchfield, N. Ratliff, and D. Fox, "Dexpilot: Vision-based teleoperation of dexterous robotic hand-arm system," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9164–9170, IEEE, 2020.

[362] S. Chamorro, J. Collier, and F. Grondin, "Neural network based lidar gesture recognition for realtime robot teleoperation," in *2021 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pp. 98–103, IEEE, 2021.

[363] S. Bier, R. Li, and W. Wang, "A full-dimensional robot teleoperation platform," in *2020 11th International Conference on Mechanical and Aerospace Engineering (ICMAE)*, pp. 186–191, IEEE, 2020.

[364] R. Luz, J. Corujeira, L. Grisoni, F. Giraud, J. L. Silva, and R. Ventura, "On the use of haptic tablets for ugv teleoperation in unstructured environments: System design and evaluation," *IEEE Access*, vol. 7, pp. 95443–95454, 2019.

[365] L. Yang, Y. Chen, Z. Liu, K. Chen, and Z. Zhang, "Adaptive fuzzy control for teleoper-

ation system with uncertain kinematics and dynamics," *International Journal of Control, Automation and Systems*, vol. 17, no. 5, pp. 1158–1166, 2019.

[366] M. Kamezaki, J. Yang, H. Iwata, and S. Sugano, "Visibility enhancement using autonomous multicamera controls with situational role assignment for teleoperated work machines," *Journal of Field Robotics*, vol. 33, no. 6, pp. 802–824, 2016.

[367] A. Valiton, H. Baez, N. Harrison, J. Roy, and Z. Li, "Active telepresence assistance for supervisory control: A user study with a multi-camera tele-nursing robot," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021.

[368] C. R. Evans, M. G. Medina, and A. M. Dwyer, "Telemedicine and telerobotics: from science fiction to reality," *Updates in surgery*, vol. 70, no. 3, pp. 357–362, 2018.

[369] C. González, J. E. Solanes, A. Munoz, L. Gracia, V. Girbés-Juan, and J. Tornero, "Advanced teleoperation and control system for industrial robots based on augmented virtuality and haptic feedback," *Journal of Manufacturing Systems*, vol. 59, pp. 283–298, 2021.

[370] O. Tokatli, P. Das, R. Nath, L. Pangione, A. Altobelli, G. Burroughes, E. T. Jonasson, M. F. Turner, and R. Skilton, "Robot-assisted glovebox teleoperation for nuclear industry," *Robotics*, vol. 10, no. 3, p. 85, 2021.

[371] D. Ryu, C.-S. Hwang, S. Kang, M. Kim, and J.-B. Song, "Wearable haptic-based multi-modal teleoperation of field mobile manipulator for explosive ordnance disposal," in *IEEE International Safety, Security and Rescue Rototics, Workshop, 2005.*, pp. 75–80, IEEE, 2005.

[372] S. Haddadin, L. Johannsmeier, and F. D. Ledezma, "Tactile robots as a central embodiment of the tactile internet," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 471–487, 2018.

[373] S. Parsa, H. A. Maior, A. R. E. Thumwood, M. L. Wilson, M. Hanheide, and A. G. Esfahani, "The impact of motion scaling and haptic guidance on operators' workload and performance

in teleoperation," in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1–7, 2022.

[374] R. M. Aronson, T. Santini, T. C. Kübler, E. Kasneci, S. Srinivasa, and H. Admoni, "Eye-hand behavior in human-robot shared manipulation," in *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 4–13, IEEE, 2018.

[375] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," in *Situational awareness*, pp. 9–42, Routledge, 2017.

[376] S. N. Young and J. M. Peschel, "Review of human–machine interfaces for small unmanned systems with robotic manipulators," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 2, pp. 131–143, 2020.

[377] C. Feng, Y. Xiao, A. Willette, W. McGee, and V. R. Kamat, "Vision guided autonomous robotic assembly and as-built scanning on unstructured construction sites," *Automation in Construction*, vol. 59, pp. 128–138, 2015.

[378] P. Gliesche, T. Krick, M. Pfingsthorn, S. Drolshagen, C. Kowalski, and A. Hein, "Kinesthetic device vs. keyboard/mouse: a comparison in home care telemanipulation," *Frontiers in Robotics and AI*, vol. 7, p. 561015, 2020.

[379] W. A. Isop, C. Gebhardt, T. Nägeli, F. Fraundorfer, O. Hilliges, and D. Schmalstieg, "High-level teleoperation system for aerial exploration of indoor environments," *Frontiers in Robotics and AI*, vol. 6, p. 95, 2019.

[380] T.-C. Lin, A. U. Krishnan, and Z. Li, "Comparison of haptic and augmented reality visual cues for assisting tele-manipulation," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 9309–9316, IEEE, 2022.

[381] M. E. Walker, H. Hedayati, and D. Szafir, "Robot teleoperation with augmented reality virtual surrogates," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 202–210, IEEE, 2019.

[382] H. Fang, S. Ong, and A. Nee, "Interactive robot trajectory planning and simulation using augmented reality," *Robotics and Computer-Integrated Manufacturing*, vol. 28, no. 2, pp. 227–237, 2012.

[383] S. H. Seo, J. E. Young, and P. Irani, "How are your robot friends doing? a design exploration of graphical techniques supporting awareness of robot team members in teleoperation," *International Journal of Social Robotics*, vol. 13, pp. 725–749, 2021.

[384] T.-C. Lin, A. U. Krishnan, and Z. Li, "How people use active telepresence cameras in telemanipulation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021.

[385] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Speeded up detection of squared fiducial markers," *Image and vision Computing*, vol. 76, pp. 38–47, 2018.

[386] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer, "Generation of fiducial marker dictionaries using mixed integer linear programming," *Pattern recognition*, vol. 51, pp. 481–491, 2016.

[387] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 895–935, 2020.

[388] H. Admoni and S. Srinivasa, "Predicting user intent through eye gaze for shared autonomy," in *2016 AAAI Fall Symposium Series*, 2016.

[389] Y. Razin and K. Feigh, "Learning to predict intent from gaze during robotic hand-eye coordination," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[390] H. C. Ravichandar, A. Kumar, and A. Dani, "Gaze and motion information fusion for human intention inference," *International Journal of Intelligent Robotics and Applications*, vol. 2, no. 2, pp. 136–148, 2018.

[391] A. Kadian, J. Truong, A. Gokaslan, A. Clegg, E. Wijmans, S. Lee, M. Savva, S. Chernova, and D. Batra, "Sim2real predictivity: Does evaluation in simulation predict real-world performance?," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6670–6677, 2020.

[392] F.-J. Chu, R. Xu, L. Seguin, and P. A. Vela, "Toward affordance detection and ranking on novel objects for real-world robotic manipulation," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4070–4077, 2019.

[393] D. Wilkie, J. Van Den Berg, and D. Manocha, "Generalized velocity obstacles," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5573–5578, IEEE, 2009.

[394] J. Dufek, X. Xiao, and R. R. Murphy, "Best viewpoints for external robots or sensors assisting other robots," *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 4, pp. 324–334, 2021.

[395] X. Gao, J. Silvério, E. Pignat, S. Calinon, M. Li, and X. Xiao, "Motion mappings for continuous bilateral teleoperation," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5048–5055, 2021.

[396] K. Krejtz, A. T. Duchowski, A. Niedzielska, C. Biele, and I. Krejtz, "Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze," *PloS one*, vol. 13, no. 9, p. e0203629, 2018.

[397] E. H. Hess and J. M. Polt, "Pupil size in relation to mental activity during simple problem-solving," *Science*, vol. 143, no. 3611, pp. 1190–1192, 1964.

[398] A. D. Souchet, S. Philippe, D. Lourdeaux, and L. Leroy, "Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review," *International Journal of Human–Computer Interaction*, vol. 38, no. 9, pp. 801–824, 2022.

[399] L. McAtamney and E. N. Corlett, "Rula: a survey method for the investigation of work-related upper limb disorders," *Applied ergonomics*, vol. 24, no. 2, pp. 91–99, 1993.

[400] D. C. Ackland, S. Roshan-Zamir, M. Richardson, and M. G. Pandy, "Muscle and joint-contact loading at the glenohumeral joint after reverse total shoulder arthroplasty," *Journal of Orthopaedic Research*, vol. 29, no. 12, pp. 1850–1858, 2011.

[401] R. Hale, D. Dorman, and R. V. Gonzalez, "Individual muscle force parameters and fiber operating ranges for elbow flexion–extension and forearm pronation–supination," *Journal of biomechanics*, vol. 44, no. 4, pp. 650–656, 2011.

[402] G. M. Kontakis, K. Steriopoulos, J. Damilakis, and E. Michalodimitrakis, "The position of the axillary nerve in the deltoid muscle: A cadaveric study," *Acta orthopaedica Scandinavica*, vol. 70, no. 1, pp. 9–11, 1999.

[403] assessmentday, "Spatial reasoning free test," 2022.

[404] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[405] M. D. Coovert, T. Lee, I. Shindev, and Y. Sun, "Spatial augmented reality as a method for a

mobile robot to communicate intended movement," *Computers in Human Behavior*, vol. 34, pp. 241–248, 2014.

[406] S. Arevalo Arboleda, F. Rücker, T. Dierks, and J. Gerken, "Assisting manipulation and grasping in robot teleoperation with augmented reality visual cues," in *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–14, 2021.

[407] C. Piyavichayanon, M. Koga, E. Hayashi, and S. Chumkamon, "Collision-aware ar telemanipulation using depth mesh," in *2022 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 386–392, IEEE, 2022.

[408] J. Larsson, M. Broxvall, and A. Saffiotti, "An evaluation of local autonomy applied to teleoperated vehicles in underground mines," in *2010 IEEE International Conference on Robotics and Automation*, pp. 1745–1752, IEEE, 2010.

[409] J. F. Mullen, J. Mosier, S. Chakrabarti, A. Chen, T. White, and D. P. Losey, "Communicating inferred goals with passive augmented reality and active haptic feedback," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 8522–8529, 2021.

[410] K. Chintamani, A. Cao, R. D. Ellis, and A. K. Pandya, "Improved telemanipulator navigation during display-control misalignments using augmented reality cues," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 1, pp. 29–39, 2009.

[411] S. Arévalo Arboleda, T. Dierks, F. Rücker, and J. Gerken, "Exploring the visual space to improve depth perception in robot teleoperation using augmented reality: the role of distance and target's pose in time, success, and certainty," in *IFIP Conference on Human-Computer Interaction*, pp. 522–543, Springer, 2021.

[412] M. Walker, Z. Chen, M. Whitlock, D. Blair, D. A. Szafir, C. Heckman, and D. Szafir, "A

mixed reality supervision and telepresence interface for outdoor field robotics," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2345–2352, IEEE, 2021.

[413] C. Brooks and D. Szafir, "Visualization of intended assistance for acceptance of shared control," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 11425–11430, IEEE, 2020.

[414] I. Jang, J. Carrasco, A. Weightman, and B. Lennox, "Intuitive bare-hand teleoperation of a robotic manipulator using virtual reality and leap motion," in *Annual Conference Towards Autonomous Robotic Systems*, pp. 283–294, Springer, 2019.

[415] B. Bejczy, R. Bozyil, E. Vaičekauskas, S. B. K. Petersen, S. Bøgh, S. S. Hjorth, and E. B. Hansen, "Mixed reality interface for improving mobile manipulator teleoperation in contamination critical applications," *Procedia Manufacturing*, vol. 51, pp. 620–626, 2020.

[416] Y. Chen, B. Zhang, J. Zhou, and K. Wang, "Real-time 3d unstructured environment reconstruction utilizing vr and kinect-based immersive teleoperation for agricultural field robots," *Computers and Electronics in Agriculture*, vol. 175, p. 105579, 2020.

[417] L. Penco, N. Scianca, V. Modugno, L. Lanari, G. Oriolo, and S. Ivaldi, "A multimode teleoperation framework for humanoid loco-manipulation: An application for the icub robot," *IEEE Robotics & Automation Magazine*, vol. 26, no. 4, pp. 73–82, 2019.

[418] C. Rognon, A. R. Wu, S. Mintchev, A. Ijspeert, and D. Floreano, "Haptic guidance with a soft exoskeleton reduces error in drone teleoperation," in *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*, pp. 404–415, Springer, 2018.

[419] C. G. Welker, V. L. Chiu, A. S. Voloshina, S. H. Collins, and A. M. Okamura, "Teleoperation

of an ankle-foot prosthesis with a wrist exoskeleton," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 5, pp. 1714–1725, 2020.

[420] F. Falck, K. Larppichet, and P. Kormushev, "De vito: A dual-arm, high degree-of-freedom, lightweight, inexpensive, passive upper-limb exoskeleton for robot teleoperation," in *Annual Conference Towards Autonomous Robotic Systems*, pp. 78–89, Springer, 2019.

[421] P. Herbin and M. Pajor, "The torque control system of exoskeleton exoarm 7-dof used in bilateral teleoperation system," in *AIP Conference Proceedings*, vol. 2029, p. 020020, AIP Publishing LLC, 2018.

[422] E. Triantafyllidis, C. Mcgreavy, J. Gu, and Z. Li, "Study of multimodal interfaces and the improvements on teleoperation," *IEEE Access*, vol. 8, pp. 78213–78227, 2020.

[423] P. Schmaus, D. Leidner, T. Krüger, R. Bayer, B. Pleintinger, A. Schiele, and N. Y. Lii, "Knowledge driven orbit-to-ground teleoperation of a robot coworker," *IEEE Robotics and Automation Letters*, vol. 5, no. 1, pp. 143–150, 2019.

[424] R. M. Aronson and H. Admoni, "Gaze complements control input for goal prediction during assisted teleoperation," in *Robotics science and systems*, 2022.

[425] D. A. Abbink, T. Carlson, M. Mulder, J. C. De Winter, F. Aminravan, T. L. Gibo, and E. R. Boer, "A topology of shared control systems—finding common ground in diversity," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 5, pp. 509–525, 2018.

[426] S. A. Mostafa, M. S. Ahmad, and A. Mustapha, "Adjustable autonomy: a systematic literature review," *Artificial Intelligence Review*, vol. 51, pp. 149–186, 2019.