Open-Minded

DOCTORAL THESIS

Towards a Human-Robot Interaction Design for People with Motor Disabilities by Enhancing the Visual Space

by Stephanie Arévalo Arboleda

Towards a Human-Robot Interaction Design for People with Motor Disabilities by Enhancing the Visual Space

Von der Fakultät für Ingenieurwissenschaften

Abteilung Informatik und Angewandte Kognitionswissenschaft

der Universität Duisburg-Essen

zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften (Dr.-Ing.)

genehmigte kumulative Dissertation

von

Stephanie G. Arévalo Arboleda
aus
Quito - Ecuador

1. Gutachter: Prof. Dr.-Ing. Jürgen Ziegler
2. Gutachter: Prof. Dr. rer. nat. Jens Gerken
3. Gutachter: Prof. Dr.-Ing. Maic Masuch

Tag der mündlichen Prüfung: 10-08-2022

ABSTRACT

People with motor disabilities experience several physical limitations that affect not only their activities of daily living but their integration into the labor market. Human-Robot Collaboration presents opportunities to enhance human capabilities and counters physical limitations through different interaction paradigms and technological devices. However, little is known about the needs, expectations, and perspectives of people with motor disabilities within a human-robot collaborative work environment.

In this thesis, we aim to shed light on the perspectives of people with motor disabilities when designing a teleoperation concept that could enable them to perform manipulation tasks in a manufacturing environment. First, we provide the concerns of different people with motor disabilities, social workers, and caregivers about including a collaborative robotic arm in assembly lines. Second, we identify specific opportunities and potential challenges in hands-free interaction design for robot control. Third, we present a multimodal hands-free interaction for robot control that uses augmented reality to display the user interface. On top of that, we propose a feedback concept that provides augmented visual cues to aid robot operators in gaining a better perception of the location of the objects in the workspace and improve performance in pick-and-place tasks.

We present our contributions through six studies with people with and without disabilities, and the empirical findings are reported in eight publications. Publications I, II, and IV aim to extend the research efforts of designing human-robot collaborative spaces for people with motor disabilities. Publication III sheds a light on the reasoning for hands-free modality choices and Publication VIII evaluates a hands-free teleoperation concept with an individual with motor disabilities. Publications V - VIII explore augmented reality to present a user interface that facilitates hands-free robot control and uses augmented visual cues to address depth perception issues improving thus performance in pick-and-place tasks.

Our findings can be summarized as follows. We point out concerns grouped into three themes: the robot fitting in the social and organizational structure, human-robot synergy, and human-robot problem management. Additionally, we provide five lessons learned derived from the pragmatic use of participatory design for people with motor disabilities, (1) approach participants through different channels and allow for multidisciplinarity in the research team, (2) consider the relationship between social dependencies in the selection of a participatory design technique, (3) plan for early exposure to robots and other technology, (4) take into account all opinions in design sessions, and (5) acknowledge that ethical implications go beyond consent. Also, we introduce findings about the nature of modality choices in hands-free interaction, which point to the user's own abilities and individual experiences as determining factors in interaction evaluation. Finally, we present and evaluate a possible hands-free multimodal interaction design for robot control using augmented reality and augmented visual cues. We propose that augmented visual cues can improve depth perception and performance in pick-and-place tasks. Thus, we evaluated our designs of visual cues by taking into account depth-related variables (target's distance and pose) and subjective certainty. Our results highlight that shorter distances and a clear pose lead to higher success, faster grasping time, and higher certainty. In addition, we re-designed our augmented visual cues considering visualization techniques and monocular cues that could be used to enhance the visual space for robot teleoperation. Our results demonstrate that our augmented visual cues can assist robot control and increase accuracy in pick-and-place tasks.

In conclusion, our findings on people with motor disabilities in a human-robot collaborative workplace, a hands-free multimodal interaction design, and augmented visual cues can extend the knowledge about using mixed reality in human-robot interaction. Further, these contributions have the potential to promote future research to design inclusive environments for people with disabilities.

Keywords Human-robot interaction, Augmented reality, Hands-free modalities, Participatory design, People with disabilities, Interaction design, Empirical studies.

PREFACE

Along this work, I could not feel closer to the saying “It takes a village,” as this is the result of the joint effort and support from different people and institutions.

First, I would like to express my gratitude to Prof. Jürgen Ziegler. Also, I am fondly grateful to the project director and my co-supervisor Prof. Jens Gerken for his guidance, support, openness, and patience during these years.

I would also like to thank the MIA project partners who made possible the development of this research. The staff and incredible workers of Büngern Technik who opened up their doors, allowed us to carry out co-design processes, conduct experiments, and helped us get closer to the lives of people with disabilities. Additionally, I would like to acknowledge the support from fellow project members from the Group of Sensors and Actuators of the Department of Electrical Engineering and Applied Physics, Prof. Marion Gebhard and Dr. Lukas Wohle.

I thank all my co-authors and colleagues from the HCI Group at the Westphalian University of Applied Sciences. Their support while designing and conducting various experiments contributed a great deal to the success of this research.

At last but not least, I am enormously grateful to all the people close to my heart. To my whole family who trusted and supported all my decisions in my academic career in so many different ways. To my dear friends, Achmad Maulana for the support in our brainstorming sessions and Paulina Duran Miranda for taking some time to proofread some publications and this thesis. To those who held me and provided emotional support when I needed it the most. To my boyfriend Henrik Dreisewerd for standing by me during the stressful time along my Ph.D. and easing my living as a foreigner in Germany. To my best friend Patricia Canizares for the endless hours on the phone listening and sending me all her love from the other hemisphere. To Dr. Abdelrahaman Aly whose kind words helped me to trundle during the last year of my Ph.D. To all my friends who in one or another way supported me.

From the bottom of my heart, thank you all.

Stephanie Arevalo Arboleda

TABLE OF CONTENTS

1. INTRODUCTION.......... 1
   1.1 An Intersection of Human-Robot Interaction, Augmented Reality, and People with Disabilities 1
   1.2 Background 3
   1.3 Problem Formulation 5
2. THEORETICAL BACKGROUND 7
   2.1 User-Centered Design in Human-Robot Interaction 7
   2.2 Human-Robot Collaboration in Assembly Environments 10
   2.3 Robot Teleoperation and Human-Robot Interaction 12
      2.3.1 Hands-free Multimodal Teleoperation 14
   2.4 Mixed Reality for Human-Robot Collaboration 16
3. RESEARCH SCOPE AND APPROACH 18
   3.1 Objective 1. Understand HRC for PMD at the Workplace 19
   3.2 Objective 2. Understand hands-free modality choices and hands-free multimodal teleoperation. 19
   3.3 Objective 3. Investigate how AR could enhance the visual space for co-located teleoperation and improve task performance. 20
4. CONTRIBUTIONS 22
   4.1 Synopsis of Publication I. 22
   4.2 Synopsis of Publication II. 23
   4.3 Synopsis of Publication III. 25
   4.4 Synopsis of Publication IV. 26
   4.5 Synopsis of Publication V. 28
   4.6 Synopsis of Publication VI. 29
   4.7 Synopsis of Publication VII. 30
   4.8 Synopsis of Publication VIII. 32
5. DISCUSSION & LIMITATIONS 34
   5.1 Objective 1. Understanding HRC for PMD at the Workplace 34
   5.2 Objective 2. Understand hands-free modality choices and hands-free multimodal teleoperation. 35
   5.3 Objective 3. Using AR to enhance the visual space for co-located teleoperation and improve task performance 36
   5.4 Limitations 38
   5.5 Future Directions 39
6. BIBLIOGRAPHY 41
7. APPENDIX 50
   7.1 Paper I. Understanding Human-Robot Collaboration for People with Mobility Impairments at the Workplace, a Thematic Analysis. 50

7.2 Paper II. Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects. 57
7.3 Paper III. What's behind a choice? Understanding Modality Choices under Changing Environmental Conditions.. 67
7.4 Paper IV. Opportunities and Challenges in Mixed-Reality for an Inclusive Human-Robot Collaboration Environment. 79
7.5 Paper V. There's More than Meets the Eye: Enhancing Robot Control through Augmented Visual Cues. 84
7.6 Paper VI. 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. 88
7.7 Paper VII. Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues. 111
7.8 Paper VIII. Does One Size Fit All? A Case Study to Discuss Findings of an Augmented Hands-Free Robot Teleoperation Concept for People with and without Motor Disabilities. 126

LIST OF PUBLICATIONS

I. **Arévalo Arboleda, Stephanie**; Pascher, Max; Lakhnati, Younes; and Gerken, Jens. 2020. Understanding Human-Robot Collaboration for People with Mobility Impairments at the Workplace, a Thematic Analysis. In 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE, 561–566, 2020. DOI=10.1109/RO-MAN47096.2020.9223489.

II. **Arévalo Arboleda Stephanie**; Pascher Max; Baumeister Annalies; Klein Barbara; and Gerken Jens. 2021. Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects. In The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA 2021). Association for Computing Machinery, New York, NY, USA, 147–155. DOI: 10.1145/3453892.3458044.

III. **Arévalo Arboleda Stephanie**; Miller Stanislaw; Janka Martha; and Gerken Jens. 2019. What's behind a choice? Understanding Modality Choices under Changing Environmental Conditions. In 2019 International Conference on Multimodal Interaction (ICMI '19). Association for Computing Machinery, New York, NY, USA, 291–301. DOI: 10.1145/3340555.3353717.

IV. **Arévalo Arboleda, Stephanie**; Pascher, Max; and Gerken, Jens. 2018. Opportunities and Challenges in Mixed-Reality for an Inclusive Human-Robot Collaboration Environment. In 1st. International Workshop on Virtual, Augmented, and Mixed Reality for Human-Robot Interaction (VAM-HRI), held in conjunction with ACM/IEEE Human-Robot Interaction (HRI '18).

V. **Arévalo Arboleda, Stephanie**; Dierks, Tim; Rücker, Franziska; and Gerken, Jens. 2020. There's More than Meets the Eye: Enhancing Robot Control through Augmented Visual Cues. In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20). Association for Computing Machinery, New York, NY, USA, 104–106. DOI:10.1145/3371382.3378240.

VI. **Arévalo Arboleda, Stephanie;** Dierks, Tim; Rücker, Franziska; and Gerken, Jens. 2021. 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. Human-Computer Interaction – INTERACT 2021. INTERACT 2021. Lecture Notes in Computer Science, vol 12932. Springer, Cham. DOI: 10.1007/978-3-030-85623-6_31.

VII. **Arévalo Arboleda, Stephanie**; Rücker, Franziska; Dierks, Tim; and Gerken, Jens. (Eds. 2021). Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '21), ACM, NY. DOI: 10.1145/3411764.3445398.

VIII. **Arévalo Arboleda, Stephanie**; Becker, Marvin; and Gerken, Jens. 2022. Does One Size Fit All? A Case Study to Discuss Findings of an Augmented Hands-Free Robot Teleoperation Concept for People with and without Motor Disabilities. Technologies, 10(1), 4. DOI:10.3390/technologies10010004.

LIST OF FIGURES

Figure 1. Adaptation of the research activities in cognitive HRI and paradigms of HRI by Mutlu et al. (2016).. 10

Figure 2. Graphical Representation of different levels of task performance in HRC. Adapted from Bauer et al. (2016). 11

LIST OF TABLES

Table 1. Categorization of tasks at Büngern Technik 4
Table 2. Summary of Publications and their relation to the RQs. 18

LIST OF ACRONYMS

AR	Augmented Reality
ARHMD	Augmented Reality Head-Mounted Display
CASA	Computers are Social Actors
DIX	Disability Interaction
HCI	Human-Computer Interaction
HMD	Head-Mounted Display
HRC	Human-Robot Collaboration
HRI	Human-Robot Interaction
MIA	German Acronym for Human-Robot Interaction at the Workplace for People with Physical Impairments
MR	Mixed Reality
PD	Participatory Design
PMD	People with Motor Disabilities
RQ	Research Question
SA	Situation Awareness
UCD	User-Centered Design
UI	User Interface
VR	Virtual Reality
WHO	World-Health Organization

1. INTRODUCTION

> "For most of us, technology makes things easier. For a person with a disability, it makes things possible."
>
> Judy Heumann. American Disability Rights Activist.

This thesis aims to contribute to the efforts of combining *Augmented Reality* (AR) for *Human-Robot Collaboration* (HRC) within the context of User-Centered Human-Robot Interaction for *People with Motor Disabilities* (PMD).

The following chapters structure this thesis as follows: first, we provide a general outline and introduce the problem space to then present relevant related work; this facilitates establishing the research scope and approach of the thesis; then, we provide brief synopses of the contributions made; finally, the discussion chapter puts in context the objectives of this thesis relative to the contributions.

Clarification: This thesis uses the pronoun "we" to ease the reading flow and avoid unnecessary use of passive voice. In addition, it acknowledges the multiple efforts of several people that contributed to its development. However, the conceptualization, design, and part of the implementation are the original work of the author.

1.1 An Intersection of Human-Robot Interaction, Augmented Reality, and People with Disabilities

The *World Health Organization* (WHO), in its global disability action plan 2014-2021, reports that one in seven people have a disability globally, and this number will only increase with time (WHO 2015). Its action plan states that "people with disabilities have unique insights about their disability and situation but have been excluded from decision-making process about issues that directly affect their lives." Carrying out research together with people with disabilities has long been a matter of concern in the social sciences, where social and ethical aspects have been highlighted (Sałkowska 2018). However, it has also sparked designing technology that supports the different abilities of people with disabilities.

Oswal (2019) mentions the opportunities to gain knowledge from the affordances of the other senses that people with disabilities use to design technology. Further, Newell et al. (2011) introduced "user-sensitive inclusive design," where design is oriented toward changing and diversifying the user's abilities. Here, a seminal paradigm is presented by Holloway (2019), who introduces the term, *Disability Interaction* (DIX), which considers disability as a source of innovation and calls to “design disruptive technologies for inclusion.”

“Disabilities are part of the human condition” as everyone is exposed to having some type of disability at any given point in life and the decay of abilities that comes naturally with aging (World Health Organization 2011). A specific sub-group of people with disabilities, PMD, have physical limitations as they cannot fully or partially use their upper or lower limbs, but most of them maintain their cognitive abilities. To a certain extent, these physical limitations can also be experienced by people without disabilities, for example, contextual constraints in hands-busy situations. That is why the development of technologies or assistive devices that counter physical limitations could benefit people with disabilities and extend the abilities of people without disabilities. An example of this is presented by Holloway (2019), who proposes assistive technologies as “an external coupling with the user that becomes constant in their life.” It indicates that developing assistive technologies for people with disabilities may be transferred to other users and even become universal, such as the invention of the typewriter or the commercial email client initially developed to assist deaf and blind people in communication.

Human-Robot Interaction (HRI) offers a range of potential solutions for people with disabilities in their activities of daily living and work-related environments. HRI is oriented toward understanding, designing, and evaluating robotic systems used by humans (Goodrich and Schultz 2007). Further, HRI brings several interaction paradigms that enhance the abilities of PMD by allowing them to perform tasks that could only be achieved with the help of others. For instance, robotic arms enable PMD to perform reaching and manipulating tasks by themselves. Manipulation tasks involve a fundamental interaction between human operators and robots that is common in assembly environments. Assembly environments have been the pioneers of using industrial collaborative robots (Akella et al. 1999). Humans collaborating with robots brings flexibility to assembly environments and assists the human counterpart in exhaustive or monotonous tasks (Krüger et al. 2009). For PMD, a robotic system provides a workspace where manipulation and grasping tasks can be performed by a robot, e.g., robotic arms. Stöhr et al. (2018) present a model that allows users with different disabilities to control a robot and perform manufacturing tasks. Further, their model encourages the exploration of flexible multimodal interfaces for robot control.

In HRI, a crucial factor is the *User Interface* (UI) since it allows operators to interact with the robot. It should take advantage of human skills by providing information that allows to understand the system status, behavior, and facilitates intervention (Villani et al. 2018). Krüger et al. (2009) divide interfaces into (1) remote interfaces, which may use visual, gestures, and/or voice; (2) physical interfaces, which use haptic, displays, force feedback systems, and *Head-Mounted Displays* (HMDs) for interaction. An approach to designing an interface for robot control is by considering virtual and augmented reality.

AR through HMDs promotes new HRI paradigms such as hands-free modalities (gaze, head movements, and speech). The widespread use of *Augmented Reality Head Mounted Displays* (ARHMDs), such as the Microsoft HoloLens[1], has expanded the visualization and interaction possibilities by combining virtual and real elements. However, it also brings a new set of challenges to HRI. Some challenges related to the manner of presenting visual information within a UI for robot control or expanding the visual space to allow teleoperation and improve HRC.

HRC assembly environments can be enriched by adding interactable virtual objects that coexist in the real world. Ong et al. (2008) presented a review of uses of AR in manufacturing, where one of the most common applications found was providing instructions in the operator's line of sight that can be used for training, maintenance, repair, or inspection. Wang et al. (2016) present a survey of AR in assembly research that categorizes it into guidance, training and design, simulation, and planning. For PMD, AR in HRC could improve awareness of the environment, as it could enhance the visual space and address issues such as depth perception.

1.2 Background

This thesis is part of the research efforts of the MIA project (Mensch-Roboter Interaktion im Arbeitsleben bewegungseingeschränkter Personen, FKZ: 13FH011IX6), which is the German acronym for Human-Robot Interaction at the Workplace for People with Physical Impairments. It builds upon previous work (MeroSy 2015-2018[2]) on robot control for PMD and extends it to include AR technologies that could enhance human capabilities and enrich HRC.

One of the cooperating institutions of the MIA project is a sheltered workshop (Büngern Technik[3]). A sheltered workshop is an institution that offers a safe employment and training environment for people with disabilities. Having direct contact with this institution allowed us to explore a possible HRC at the workplace for PMD. Further, the MIA project and the sheltered workshop provided the foundations for this thesis by serving as inspiration (1) to opt for a *Participatory Design* (PD) approach and (2) to define a common HRI in certain workplaces, i.e., pick-and-place. The following paragraphs elaborate on this process.

[1] https://www.microsoft.com/en-us/hololens
[2] https://www.interaktive-technologien.de/projekte/merosy
[3] https://buengern-technik.de/de/

As part of the project, we performed a set of visits to Büngern Technik. The main goal was to better understand the type of work that is performed and get closer to the different stakeholders (people with disabilities, social workers, supervisors, and caregivers). We established a panel of discussion about shared goals and stakeholders' involvement along the design process of a human-robot collaborative workstation. This influenced the decision of using a PD approach and consider the involvement of the stakeholders along the design process. Further, this helped to establish a “lead user group” of five PMD. However, due to advanced cognitive disabilities and difficulties in communication, two of them did not participate further in the design process.

The discussions and observations at the different workstations of the sheltered workshop are summarized in Table 1. These revealed that an HRC setting for PMD may involve a pick-and-place interaction. Hence, we considered the tasks from the “Assembly” category, as there is a wide range of sub-tasks involved, e.g., picking, manipulation, and placing.

Assembly tasks may require a certain degree of complexity when handling certain pieces. These could benefit from human-supervisory control of robots, teleoperation, and semi-automated operation. Further, for PMD, this opens a working space that enables them to teleoperate robotic arms and perform assembly tasks.

Table 1. *Categorization of tasks at Büngern Technik*

Task Category	Description	Type of Workstations
Customization and Judgement	High level of judgment and cautiousness.	Candle fabrication, Sanding, waxing, and painting of wooden pieces.
Fine Detail	Detailed manual work. Handicrafts	Greeting cards and car filter fabrication
Assembly	An object is picked from an initial location, then it is manipulated and changed, to finally place it in a different location.	Packaging, Sorting, Stickering, wooden toys assembly, metal parts assembly for beds and cars

1.3 Problem Formulation

PMD experience a series of physical limitations that could be counteracted through HRC by enabling them to physically manipulate objects in their surroundings. However, there is still room for exploration of how PMD could actually interact and collaborate with a robot, especially in a workplace environment. This raises a set of questions that we grouped as (1) PMD-related, how do they feel about collaborating with a robot and sharing duties? How do the individual differences of PMD influence the use of technology and the interaction possibilities? (2) interaction-modalities-related, which interaction modalities can still be used? and how could these be mapped within HRC? What type of interaction challenges arise from an HRC setup at the workplace? And (3) robot-control-related, how PMD can control a robotic system? How can AR be used to support robot control for manipulation tasks? What other elements should be considered in robot control to improve performance in manipulation tasks?

First, we aim to expand the knowledge about HRC for PMD in a workplace environment. Philips & Zhao (1993) presented that one of the reasons people with disabilities do not use a technological aid lay in the researchers' neglection of their opinions during the design process. Therefore, to better understand how people with disabilities could work together with a robot, we involve PMD along the design process using a mix of user-centered design and PD.

Second, we investigate hands-free interaction modalities to design an interaction suitable for PMD. People with limited limb mobility use alternative senses to interact with the environment (Markham and Greenough 2004) and have different levels of physical abilities with their upper limbs. The challenge arises on how to design a suitable hands-free teleoperation concept. This requires understanding different input and output modalities available to PMD and the technology to leverage those for the interaction. People have large differences in abilities and preferences in terms of modes of communication, whereby a multimodal interface may support this diversity (Oviatt and Cohen 2015). Further, hands-free interaction requires a deeper understanding of multimodal interaction, i.e., the combination of different modalities to overcome certain limitations.

Third, the interaction that takes place when controlling a robot for manipulation tasks is different from the one with a computer or mobile device, as the robot does manipulate the real world. Therefore, it requires direct attention to a point of interest in space, where the manipulation task takes place. However, this is usually not the place where visual output is presented, as it is often presented in an external display detached from the robot, especially in teleoperation. Teleoperating a robotic arm can be an arduous task that involves expertise and attention, concentration, and spatial abilities—these relate to “the ability to generate, retain, retrieve, and transform” spatial relations of objects (Lohman 1996). Spatial abilities are hindered for PMD who cannot acquire visual information by moving around the workspace. Hence, we explore the potential of using AR technologies for robot teleoperation by presenting a UI through augmented reality and providing visual feedback to achieve better task performance. A pivotal work in this area is the work of Walker et al. (2018) and Williams et al. (2019), who presented a design framework for AR in HRI. These works have been an influential factor in most of our contributions and provide a base for our designs that enhance the visual space for robot operators.

To summarize, in this thesis, we approach issues related to HRC for PMD by understanding the needs and expectations of HRC for PMD, designing a possible hands-free teleoperation concept, and using AR to support robot control and performance in manipulation tasks. In addition, these issues structure our approach and are aligned with our contributions.

2. THEORETICAL BACKGROUND

This chapter extends the related work contained in the publications of this thesis. It elaborates on the conceptual framework of topics related to (1) user-centered design in HRI, (2) HRC in assembly environments, (3) robot teleoperation, and (4) mixed reality in HRC.

First, in this thesis' publications, we considered a PD approach and elaborated on the related literature for each paper. However, in this chapter, we go beyond PD to elaborate on *User-Centered Design* (UCD), as it is a framework that can provide solutions for a specific target (Duque et al. 2019). Second, we narrow the research efforts in HRC down to assembly environments. This relates closely to the focus of this thesis and bounds the findings from our publications. Third, we present research about robot teleoperation in the context of PMD through hands-free multimodal teleoperation. Fourth, we extend the research that has been done not only using AR for HRC but from a broader view and present research that has been done using mixed reality for HRC.

2.1 User-Centered Design in Human-Robot Interaction

UCD is a term that "describes design processes" where end-users participate actively in decisions that influence a design (Abras et al. 2004). In UCD the user is placed at the center of the design process (Wilkinson and Angeli 2014), involving aspects related to understanding the user's tasks, ecosystems, and context of use. Dorrington et al. (2016) state that UCD is a bridge between different mental models of a user and a designer that are translated into a usable solution meeting the needs of users. When considering a UCD approach, it is important to identify and involve not only the people to whom a design is targeted (primary users) but people that might be in constant contact with the primary users and be influenced by the use of a solution (secondary users) (Bartneck 2020). It is also relevant to recognize people that might be indirectly affected by the use of the design, i.e., tertiary users, and involve them as well when appropriate.

UCD is a well-known approach in the *Human-Computer Interaction* (HCI) community. UCD guidelines, goals, and techniques have been applied to HRI considering the particularities of a human-robot environment. One of the first approaches in HRI comes from (Granda et al. 1990) which provides stages for human-robot systems (1) bounded autonomy, (2) teleoperation, (3) supervised autonomy, (4) adaptive autonomy, and (5) virtual symbiosis. All these stages are related to the level of autonomy of robots and the level of intervention of the human counterpart in task performance. Moreover, the works of Yanco & Drury (2002) integrated HCI notions to the nascent field of HRI. They presented HRI as a subset of HCI as robots are "computing-intensive systems." However, differences between HRI and HCI can come in terms of long-term autonomy (which may involve long-term interaction, robot survival in the real world, and an integrated interface and control) together with the ability to interact and learn from people (Breazeal 2004). Breazeal (2004) argues that HRI should not be merely evaluated from the human's perspective but acknowledge humans and robots as parts with particular goals, which could lead to a relationship that is beneficial to both parties.

Yanco & Drury (2002) provided a taxonomy for HRI that considers (1) the human-robot composition, (2) the amount of required interaction, (3) decision support provided for the user, and (4) space-time location. This classification goes beyond the level of autonomy of robotic systems to include the composition of a human-robot team and the level of interaction among them, plus the time/space ratio where the interaction takes place. Later, this taxonomy was extended to include the social nature of the task (human interaction roles and human-robot physical proximity), task type, and morphology of the robot (Yanco and Drury 2004).

Yanco and Drury's (2004) extension shows the vast scope of HRI as a field. There are different aspects that can be considered in HRI and may be domain-dependent. For instance, a design framework for social robots (Bartneck and Forlizzi 2004), physical HRI (Santis et al. 2008), psychological categories in HRI (Kahn et al. 2007), the level of assistance for optimal working conditions (Schmidtler et al. 2015). Onnasch & Roesler (2021) present a more recent taxonomy to structure HRI considering (1) the interaction context, (2) the robot classification, and (3) the team classification. Their first category, "interaction context," considers the different fields of application, e.g., industry, therapy, education, military, etc.; and the exposure to the robot (embodied, depicted) and the type of setting (field, laboratory). Their second category, "the robot classification," involves the robot task specification, e.g., information exchange, precision, manipulation, different types of stimulation, transport, etc.; robot morphology that takes into account appearance, communication, movement, and context; and degree of robot autonomy in terms of information acquisition and analysis, decision-making, and action implementation. Their third category, "the team classification," contemplates the human role (supervisor, operator, collaborator, cooperator, bystander), the team composition, the communication channel divided into input (electronic, mechanical, acoustic, optic) and output (tactile, acoustic, visual), and the proximity of the team, temporal (synchronous, asynchronous) and physical (following, touching, approaching, passing, avoidance, none). This classification merges previous classifications presented by other authors to help structure HRI.

Kim et al. (2011) establish two approaches in HRI research "Robot Centered" and "User Centered." They mention that "Robot Centered HRI" focuses on designing for improving robots and their interaction with a human counterpart. Whilst a User-Centered HRI emphasizes the needs and perceptions of robot users when interacting with a robot. Additionally, a User-Centered HRI investigates ways of interaction that favor users' objectives and aims at a better overall experience. This, in turn, aligns with a broader definition of technology-centered design, "technologies that let humans adapt" vs. user-centered design, "design technologies to fit the capabilities of humans" (Endsley and Jones 2012).

Designing an interaction environment is connected to the degree that natural human interaction is incorporated with technology to allow a continuous flow of information (Papageorgiou et al. 2019). This is of particular interest in HRI since it can improve human-robot co-existence and facilitate the implementation of collaborative environments. Also, Kim et al. (2011) present a set of considerations when designing for a user-centered HRI. Considerations that relate (1) to the elements involved in the design process according to a rhetorical classification of mental processes such as: "aesthetic contextuability", "operational contextuability" and "social contextuability". Here, social and humanoid robots give special attention to "aesthetic and social contextuability", e.g., PARO[4], Pepper[5], NAO[6], whilst industrial robots tend to focus on "operational contextuability", by considering elements related to robot manipulation, movement patterns, and usability. This thesis considers elements related to the operation contextuability of a user-centered HRI.

In HRI, human factors can involve capturing expectations and information processing before, during, and after an interaction. Human factors can be defined as the physical, psychological, and social characteristics of humans and the extent to which they influence or are influenced by socio-technical systems (Badke-Schaub et al. 2012). One area within HRI that emphasizes the context of the interaction between humans and robots in the real world as a complex cognitive system is Cognitive HRI. Cognitive HRI aims to improve interaction by understanding human mental models and developing cognitive models for robots (Mutlu et al. 2016). Figure 1 shows the role of the human and robot in performing tasks from the cognitive HRI perspective. This illustration places the task as the center of the interaction and the different approaches behind the robot and the human committing action towards the development of the task. The importance of how the task is handled and the challenges of it from the human and the robot's side are also main areas of interest in the human-centered design of robots (Kidd et al. 1992). Further, Kidd et al. (1992) emphasized human skills in robotic systems, where robotic systems should enhance human skills and abilities

[4] http://www.parorobots.com/
[5] https://www.softbankrobotics.com/emea/en/pepper
[6] https://www.softbankrobotics.com/emea/en/nao

to achieve higher levels of productivity, effectiveness, and safety, releaving the human counterpart from tedious tasks.

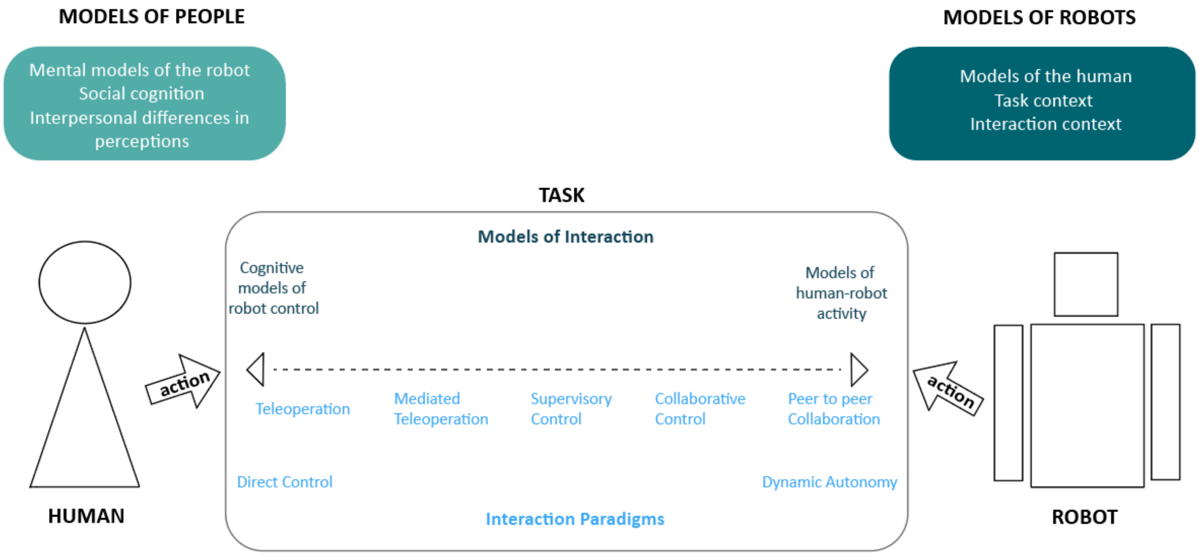


Figure 1. *Adaptation of the research activities in cognitive HRI and paradigms of HRI by Mutlu et al. (2016).*

2.2 Human-Robot Collaboration in Assembly Environments

According to the ISO 8373:2012 (International Organization for Standardization 2012) HRC is characterized by the opportunity for direct physical human-robot contact. HRC addresses research where humans and robots are considered partners in carrying out joint goals (Bauer et al. 2008). The goal of HRC within assembly environments is to create an environment that counters the inherent limitations of both humans and robots through working collectively (Michalos et al. 2015).

Human-robot relationships in assembly environments involve a set of actions performed in a shared workspace that would lead to a goal. Fong et al. (2003) presented four aspects that HRC must consider: self-awareness, in terms of the capabilities of the human and the robot; self-reliance to take care of its own safety; the capacity for dialogue in terms of human-robot communication, and adaptability to collaborate with people with different skills, expertise, and knowledge.

Current assembly environments go beyond the use of industrial robots which needed safety fences and a clear separation between humans and the robot. The widespread of lightweight collaborative robots, e.g., Kuka LBR iiwa[7], Kinova[8], Universal Robots[9], enables robots to work in a safe shared workspace in close contact with humans.

[7] https://www.kuka.com/en-de/products/robot-systems/industrial-robots/lbr-iiwa
[8] https://www.kinovarobotics.com/en/robots-industry-solutions
[9] https://www.universal-robots.com/

Nevertheless, safety still is the most important enabling factor in assembly environments.

Safety can be approached in terms of collision prevention by optimizing timing and intervention (Michalos et al. 2015). Another component of safety in HRI is situation awareness, it involves determining when it is an appropriate time for action, and what information needs to be communicated to the human counterpart (Unhelkar et al. 2020). Additionally, trust is a factor that can determine the achievement of successful human-robot cooperation, since it characterizes how prompt are humans to interact with their robot counterpart (Charalambous et al. 2016).

Krüger et al. (2009) state the relevance of designing intelligent automation devices to improve workers' occupational safety and health in human-robot assembly environments. They divide these environments into workplace sharing systems and workplace and time-sharing systems. In workplace sharing systems, both humans and robots perform separately either handling or assembly tasks within the same workspace. Here, the HRI is limited to collision avoidance. In workplace and time sharing systems, humans and robots are able to jointly perform handling or assembly tasks. The HRI here is more complex and may involve force torque sensors among others.

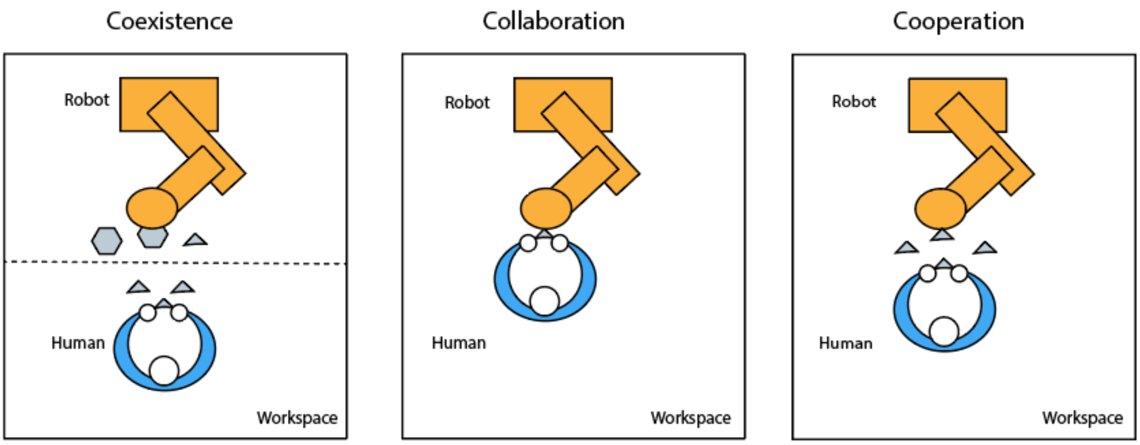


Figure 2. *Graphical Representation of different levels of task performance in HRC. Adapted from Bauer et al. (2016).*

Figure 2 presents different levels of human-robot involvement in task allocation. Human-robot coexistence is often discussed in HRC and it can be defined as humans and robots sharing a workspace that might perform actions towards the same goal without contact or coordination (Luca and Flacco 2012). Kolbeinsson et al. (2019) highlight that cooperation and collaboration are often used interchangeably. They describe cooperation as sequential actions towards a shared goal, conversely to collaboration, described as shared sequential actions towards a shared goal.

In HRC, task planning and allocation are factors that should be carefully evaluated in human-robot assembly environments. Weßkamp et al. (2019) state that task allocation should not only consider the abilities of humans and robots but include industrial science to estimate production. They propose to optimize human-robot work by considering workload from the human side, layout of the workstation, and material flow. In this regard, Tan et al. (2009) found a correlation between mental workload and both robot motion speed and distance. For instance, people experienced an increase in mental workload when the robot moved faster and at a closer distance.

Another relevant aspect of HRC is *Situation Awareness* (SA), which can be defined as "knowing what is going on around you" (Endsley and Garland 2000). SA helps humans to make decisions considering three levels, (1) perception of elements in the environment (e.g., size, location, colors, dynamics), (2) comprehension of the current situation—after the first level, it is important to understand the relationship and significance of elements within the environment, that intervene to fulfill a goal, and (3) projection of future status—this is given by the ability to foresee future actions or states, which could be achieved through the acquired knowledge in the previous levels (Endsley 1995). In HRC, SA helps keep the human counterpart informed about the environment, regardless of the level of automation, which may or may not lead to better performance when achieving joint goals. SA and performance may not be directly correlated (Scholtz 2003), as SA is also related to other factors such as working memory, perceptual-motor ability, and visual and temporal processing ability (Gugerty and Tirre 2000). However, an SA-oriented design for HRC may complement UCD practices helping the human counterpart to better acquire and process relevant information over time (Riley et al. 2010). Riley (2010) proposes some guidelines that consider different factors grouped per task, system, environment, individual/team, external world, and user interface design, as all of them play a part in achieving effective HRC.

2.3 Robot Teleoperation and Human-Robot Interaction

Humans and robots can have specific manners of interacting with each other. One of the earliest manners of humans interacting with robots is teleoperation. Sheridan (1992) defined teleoperation as the "direct and continuous human control of the teleoperator," where the teleoperator was understood as the machine. Teleoperation together with supervisory control have been considered a paradigm for HRI (Goodrich and Schultz 2007). HRI can be understood as "the study of the humans, robots, and the ways they influence each other" (Fong et al. 2003). A more recent approach presents HRI as a "field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans" (Goodrich and Schultz 2007).

Scholtz (2003) presents a classification of the human role in HRI wherein humans can act as supervisors, operators, mechanics, teammates, or bystanders. The supervisor and teammate role would have the same interaction and relationship as the one in human-human interaction. The operator role would need to teleoperate a robot or intervene to change a robot's behavior. The mechanic role would need to physically adjust the components of the robot. The bystander role may not directly interact with the robot but needs to understand robot behavior and the consequences of the robot's actions. Frennert (2020) groups the mechanic and operator roles in one and specifies the main difference against the supervisory role as follows. On the one hand, the operator role involves the human physically or remotely operating the robot, to enable that, the human needs to perceive, process, and interpret the status of the robot and make decisions. On the other hand, the supervisory role needs to handle information about the situation, understand robot tasks, and be able to intervene when needed. Independently of the human’s relationship with the robot, the user experience needs to be appropriately designed and evaluated for a better interaction flow (Frennert 2020).

Gleeson et al. (2013) analyzed assembly tasks to better design human-robot communication in industrial environments. They classified and dissected assembly tasks into several subtasks: part acquisition, part manipulation, and part operation. Part acquisition involves selecting and locating parts from the supply area and moving them to the workspace. Part manipulation includes different steps discerning and aligning parts to allow other part placement and insertions. Part operations encompass different actions after part placement, e.g., tightening and fastening, which require torque and precision. Here, HRI can address how humans and robots plan the task, interact with the parts, and manage problems.

When analyzing HRI in assembly tasks, it is essential to discuss different levels of robot autonomy. Beer et al. (2014) present a framework to categorize the levels of autonomy of a robotic system based on the interaction between humans and robots. They consider the allocation of human-robot functions (sense, plan, act) from Endsley & Kaber (1999) during task execution to present ten levels of robot autonomy:

1. Manual teleoperation, where the human is in charge of sensing, planning, and acting towards a task.
2. Action support, where the human is in charge of sensing and planning while the robot commits an action (act).
3. Assisted teleoperation, where the human senses the environment together with the robot, the human plans, and the human or the robot might intervene in the task to act according to changes in the environment.
4. Batch processing, where humans and robots sense the environment, the human plans and the robot implements the task.

5. Decision Support, where both humans and robots sense the environment, plan together, and the human commands the robot to act.
6. Shared control with human initiative, where the robot can sense, plan, and act according to a task while the human monitors the robot and may intervene.
7. Shared control with robot initiative, where the robot senses, plans, and acts to perform a task but can prompt the human for assistance when needed.
8. Supervisory control, where the robot senses, plans, and acts while the human monitors the robot and may change the goal or plan.
9. Executive control, where the human provides an abstract goal, and the robot senses, plans, and acts to reach that goal.
10. Full Autonomy, where the robot performs autonomously senses, plans, and acts to perform a task without human intervention.

These different levels of autonomy of robotic systems should be taken into consideration when designing a human-robot collaborative workspace, as they will also determine the amount of human intervention and the workload required.

2.3.1 Hands-free Multimodal Teleoperation

Multiple modalities have been explored in hands-free robot control. For instance, facial muscle movements (Tamura et al. 2010), brain interfaces (Iáñez et al. 2009), and tongue interfaces (Kim et al. 2013). However, most of these involve high complexity and variability in their tracking, which often leads to errors and hinders their use in real manufacturing environments.

A pertinent question in control interfaces is the use of multiple input modalities versus a single modality. A goal of multimodal interfaces is to integrate complementary modalities in a way that each modality contributes its strengths while counteracting the weaknesses of the other modalities (Cohen et al. 1989). Salem et al. (2011) conducted a study to evaluate multimodality vs unimodality in a human-robot collaborative task. Their results showed that the robot’s behavior was perceived more positively (active, communicative, competent) when interacting multimodally. In assistive HRI, the capabilities of robots may be dynamically adjusted to the human’s level of skill to compensate for human limitations and lead to increased performance (Modares et al. 2016).

When considering hands-free multimodality, it is crucial to understand the structure of each modality individually. Each modality processes information differently depending on the human sense that it is related to (Norris 2009). For instance, people naturally use gaze to explore their surroundings and combine it with other modalities to manipulate objects (Sibert and Jacob 2000). Majaranta et al. (2012) present the potential of gaze-based interaction in assistive technologies and highlight the user’s preference for gaze control. Tracking the movements of the eyes can be used as an input method that replaces the mouse and keyboard (Scott 2010), as eye gaze carries deictic and spatial reference information (Qvarfordt 2017). However, the eye's saccadic movements can also lead to inaccuracies (Ware and Mikaelian 1987). These can activate actions involuntarily by looking at controls (Midas touch problem) (Jacob 1990). Therefore, eye gaze could be combined with other explicit input methods (Jacob and Stellmach 2016).

Gaze is often supported by head movements to reach further distances and stabilize the gaze (Tweed et al. 1995). Hansen et al. (2018) evaluated click and dwell interaction by using gaze, head, and mouse for pointing and concluded that gaze is the least accurate but the fastest, and head pointing is more effective than gaze. Another example is "Head Movement and Gaze Input Cascaded Pointing," which uses the strength of the speed of gaze for pointing and combines it with head movements to position a mouse pointer (Kurauchi et al. 2015). Quian & Teather (2017) compared head-based vs. eye-based selection in a *Virtual Reality* (VR) environment. Their results showed that head-only input performed best in terms of accuracy, outperforming eye-only selection and even the combination of head and eye selection.

Humans naturally use verbal descriptions to describe spatial knowledge (Belouaer et al. 2013). Bolt (1980), in his prominent "Put-that-there" manuscript, showed the potential of using speech and gestures for pointing. Griffin and Bock (2000) showed that people look at an object for 900 ms before naming it. Further, speech regularly expresses the "what" while pointing expresses "where" (Elepfandt 2012). However, using speech presents a tradeoff between natural language and unnatural commands. The use of natural language translates into long commands whilst unnatural commands are rather short and straightforward. Elepfandt (2012) conducted a study that revealed people's preference for short commands as they are direct.

A common human pattern of interaction with the environment is the combination of gaze, head movements, and speech. This combination is used in pointing and selecting, two actions that can be understood from perceptions and intention (Oviatt and Cohen 2015). Pointing is classified as a deictic gesture (Arnheim and McNeill 1994) and is defined as a gesture that specifies a direction from the person's perspective (Kita 2008). Pointing involves evaluating one's position and distance in space. For instance, when we spot an object in the environment, it is natural to move one's head in the direction where that object is located first, to then gaze at it. This movement and eye gaze fixations reveal the intention over an object (Jacob 1991). Then, if we want to specify the object we are currently looking at within the environment (selection), we point at it with the index finger. Here, selection can be expressed as the action of concentrating attention on a specific object and alienating the rest of the objects in the environment. This can be explained through visual attention, when pointing, there is an unconscious saccadic eye movement that precedes the arm movement to guide the trajectory (Neggers and Bekkering 2001; Bernardos et al.

2016). Monsell & Driver (2000) explain selection in terms of visual attention and deem it as the first step towards behavioral goals. Additionally, a verbal comment about its features can be added to communicate a message about an object.

Using head motion, eye gaze, and speech as hands-free input modalities for robot control interfaces has gained popularity to compensate for hands-busy situations or as an opportunity for PMD. Jia et al. (2007) explored the use of head gestures to execute wheelchair motion control commands. Laddi et al. (2015) focused on employing solely head orientation (roll and pitch) to control the speed and direction of a mobility cart. In teleoperation, Jackowski et al. (2016) presented a head motion-based interface displayed on an external display to perform pick-and-place tasks. Park et al. (2021) proposed using eye gaze and head movements as a hands-free interaction to perform manipulation tasks with a robotic arm. Krupke et al. (2018) compared the use of speech, combined with gestures or head orientation in robot control. Their results found head orientation more precise, faster, and with a lower cognitive load. Also, Ruzaij et al. (2017) paired head movements with speech to allow users to use and switch between those two modalities to control a wheelchair. They tested this combination in a study where users controlled a wheelchair in indoor and outdoor environments. Their results showed that each voice command was better suited for binary actions, e.g., commands to turn on/off lights, and head movements were better suited for motion controls, e.g., speeding up/down. Further, manipulating spatial information using different modalities simultaneously could only improve how this information is perceived and used (Oviatt and Cohen 2015).

2.4 Mixed Reality for Human-Robot Collaboration

Science fiction has long been a predictor and inspiration to present a world where technology is a stem part of our day-to-day activities. A world where advanced computing systems bring closer virtual and real worlds is not merely a part of science fiction anymore. Milgram & Kishino (1994) presented a reality-virtuality continuum where on one end there is the real environment and on the other the virtual environment, i.e., VR, which is an immersive virtual environment with only virtual elements that may or may not be based on the real-world. In the mid-points, they locate AR and augmented virtuality, and present both as parts of *Mixed Reality* (MR). On the one hand, AR showcases virtual elements that are presented in the real world and interact with other real or virtual elements. On the other hand, augmented virtuality merges real-world elements into the virtual world.

Azuma (1997) defines AR as a system that: (1) combines real and virtual elements, (2) has virtual elements that are interactive in real-time, and (3) has a 3D registration of virtual and real objects. This definition aligns with a more recent conception of MR as an interaction concept that merges virtual and real elements in a unified interaction space (Young et al. 2011). Within this space, Young et al. (2011) consider robotic systems as MR entities, since robots can be virtual as they are advanced computational systems that deal with digital information and real since they can physically sense the world and manipulate elements of it. We believe that this conception of robotic systems presents opportunities to create blended human-robot collaborative environments with seamless human-robot communication.

In virtual environments, human-robot communication presents a safe space that allows collaboration. Simulating manufacturing environments in VR allows for testing different types of interaction that might be dangerous for human operators in a real environment with large industrial robots (Seth et al. 2011). However, virtual environments also present limitations related to the fidelity and credibility of assembly processes (Wang et al. 2016). This has led to further exploring AR to provide visual information for robot operators. Hedayati et al. (2018) presented different interfaces to provide information about the robot’s camera to the operator. Walker et al. (2018) presented different visualization techniques that can allow operators to better interpret robot motion. Regarding safety, Hietanen et al. (2020) proposed a model for a safe HRC in manufacturing environments. Their work presents a UI using AR to show danger zones in the workspace that change dynamically based on the robot's movements. Similarly, Vogel et al. (2017) used a projection-based UI with interactive buttons to control the robot to perform human-robot shared assembly tasks, e.g., the human counterpart inserts screws in a mounting plate and a robotic arm tightens them and a color-coded danger zone alerts the presence of human hands. These works show the potential of AR to provide visual feedback and facilitate teleoperation by informing operators about the different robot states and allowing for flexible HRC.

Communicating spatial information of objects in the workspace is another use of AR in assembly lines. Sukan et al. (2014) examined visual guidelines called ParaFrustum to represent viewing directions that will help users accommodate their heads and line of sight to improve the view of target objects. Additionally, sharing visual references of objects between humans and robots is crucial. Gao et al. (2020) propose a graph matching method that provides visual and spatial information that helps connect visually the robot’s perspective with the human’s one.

3. RESEARCH SCOPE AND APPROACH

The central concept of this thesis involves empirical studies, interaction design, and engineering through novel interaction concepts that lean towards further advancing technology without leaving aside the human factor for HRI.

This research is of an inductive nature since it emphasizes the study of different phenomena in HRI grounded on observations, design workshops, development of prototypes, and empirical studies. This information is used to infer knowledge that can be used to make probabilistic estimations about other situations (Hayes et al. 2010). Also, the results contained in this thesis are intended to serve as an instrument that contributes to the research efforts about hands-free multimodal interaction and the use of AR in HRI.

This thesis is presented through main objectives involving different *Research Questions* (RQ) and their corresponding publications, see Table 1. We provide two types of contributions: (1) empirical studies, based on a mixed-methods approach, and (2) practical artifacts, created through design and engineering. The **empirical** findings aim to generate knowledge about PMD in an HRC context (Objectives 1, 2, 3). Further, we provide **design and engineering** contributions about using AR in HRI to enhance robot teleoperation not only for PMD but also for a wider user group (Objective 3).

Table 2. *Summary of Publications and their relation to the RQs.*

Objective	**Link to the research question**	**Related Publications**
1. Understanding HRC for PMD at the workplace.	RQ1. What should be considered when designing HRC for PMD?	I, II
2. Understanding hands-free modality choices and hands-free multimodal teleoperation.	RQ2. Which factors do users consider in hands-free modality choices? RQ3. What is the experience evoked by an augmented hands-free teleoperation concept for PMD?	III, VIII

3. Investigate how AR could enhance the visual space for co-located teleoperation and improve task performance.	RQ4. How can AR improve depth perception in co-located teleoperation? RQ5. How to improve task performance in teleoperation by using AR?	IV, V, VI, VI, VII

In this research, we used a mixed-methods approach. The first objective is mostly of qualitative nature and serves as a base to identify aspects related to the needs, preferences, and expectations of end-users (PMD) to design a type of interaction for robot control. The second and third objectives combine qualitative and quantitative analysis to better understand hands-free multimodality, explore possible AR interfaces for robot control, and further evaluate the robot operator's visual space to enhance teleoperation.

3.1 Objective 1. Understand HRC for PMD at the Workplace

PMD have long been excluded from industrial work-related activities due to physical constraints and a lack of adequate environments that may allow them to rejoin the working force.

Our approach is to expand the knowledge about PMD in a workplace environment by identifying their needs, expectations, and concerns (RQ1). In order to achieve this, we aim to identify factors to be considered when designing a human-robot collaborative workstation (Publication I. Understanding Human-Robot Collaboration for People with Mobility Impairments at the Workplace, a Thematic Analysis). Additionally, we consider it pertinent to report our experiences through a set of challenges and lessons learned that could serve other researchers when designing robotic assistive environments (Publication II. Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects).

3.2 Objective 2. Understand hands-free modality choices and hands-free multimodal teleoperation.

Publications I and II pointed to the relevance of the individual differences in abilities and expectations of PMD for a human-robot collaborative workstation. These factors hint at the necessity for a flexible robot control interface that adapts to those abilities. Hence, before designing an interaction concept for robot control, we considered it crucial to better understand the factors that lead people to choose an input modality to interact with an interface within a range of other possible modalities.

Hands-free modalities have been widely explored in multimodal interaction in terms of efficiency and combination. However, little is known about the nature of modality choices and how they relate to interaction evaluation (RQ3, Publication III).

Publication III. What's behind a choice? Understanding Modality Choices under Changing Environmental Conditions, revealed the importance of the user's characteristics (user's own abilities) and the consequences for interaction (people's personal experience) when evaluating an interaction. That is why we aimed to design a teleoperation concept that (1) simplifies teleoperation and (2) considers the characteristics of PMD. Thus, we chose a multimodal approach that uses speech and head orientation (yaw and pitch) to teleoperate a robotic arm through an AR UI. Further, we aimed to better understand the perception of this concept by capturing the experiences of a person with motor disabilities (RQ4, Publication VIII. Does One Size Fit All? A Case Study to Discuss Findings of an Augmented Hands-Free Robot Teleoperation Concept for People with and without Motor Disabilities).

3.3 Objective 3. Investigate how AR could enhance the visual space for co-located teleoperation and improve task performance.

This objective contains the main contributions of this thesis. Publications I - III helped us understand the needs of PMD and provided us a basis to design a hands-free multimodal robot control concept. Moreover, we considered it relevant to identify different aspects of designing an interface for robot control and how AR supports presenting robot controls within the operator's line of sight (Publication IV. Opportunities and Challenges in Mixed-Reality for an Inclusive Human-Robot Collaboration Environment). Publication IV brought up more questions about the operator's visual space in co-located teleoperation, which is the core of this thesis.

In our first designs, we created a virtual environment that allows us to augment the robot for picking and the environment for placing. Additionally, we evaluate the role of situation awareness in co-located teleoperation (Publication V. There's More than Meets the Eye: Enhancing Robot Control through Augmented Visual Cues). The results from Publication V inspired further exploring visualization techniques to design visual cues. We aimed to help position the gripper above a target object to successfully grasp it. This led us to evaluate the influence of depth-related variables such as the distance and the target's pose (Publication VI. 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).

Publications V and VI prompted us to keep exploring different designs of cues. That is why we took into account the nature of monocular cues to provide augmented visual cues (Publication VII. Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues). On top of that, our augmented visual cues considered evaluating indirect visual cues vs. explicit ones. Further, we aimed to evaluate their effect not only on depth perception but on performance in pick-and-place tasks.

4. CONTRIBUTIONS

This chapter comprises summaries of each publication contained in this thesis to provide an overview of the research that was carried out.

4.1 Synopsis of Publication I.

Understanding Human-Robot Collaboration for People with Mobility Impairments at the Workplace, a Thematic Analysis

Arévalo Arboleda, Stephanie, Pascher, Max, Lakhnati, Younes, and Gerken, Jens, 2020. Understanding Human-Robot Collaboration for People with Mobility Impairments at the Workplace, a Thematic Analysis. In 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE, 561–566, 2020. DOI=10.1109/RO-MAN47096.2020.9223489.

Background

There are a series of limitations in the lives of people with mobility impairments that hinder not only their activities of daily living but their economic participation. However, human-robot collaboration creates an environment that considers the strengths of both humans and robots for safe collaboration in different workplace settings. Here, manufacturing industries have already adopted hybrid human-robot collaborative workstations that may allow people with motor disabilities to control robotic arms through hands-free modalities.

Approach

We took a closer look at the needs and preferences of people with mobility impairments to design a human-robot cooperative environment in a workplace setting. To enable this, we approached potential end-users (primary users), as well as social workers and supervisors (secondary users) from a sheltered workshop (an institution that employs people with disabilities). At this institution, we collected information using a PD technique called Future Workshop (Jungk and Müllert 1996).

A Future Workshop aims to identify common problems in a group and develop together a vision of the future. A Future Workshop has four different phases. First, a preparation phase, where we explained the protocol, the workshop's goals, and introduced a manufacturing scenario through videos and pictures. Second, a critique phase, where we encouraged discussion about potential foreseeable problems and concerns in the presented scenario. Third, a fantasy phase, wherein we presented hands-free multimodal interaction and intervention techniques for teleoperating a robotic arm.

Fourth, a final implementation phase, wherein we summarized the outcomes of the session and shared thoughts about the feasibility of some ideas. Our Future Workshop was held in two separate sessions: one with primary users and another with secondary users. The gathered information was analyzed using thematic analysis.

Results

The results pointed out not only different needs and preferences but brought to light underlying concerns and resulted in three different themes (1) the robot fitting in the social and organizational structure, (2) human-robot synergy, and (3) human-robot problem management.

The first theme concerns a robot adapting to the existent environment without drastically altering the social interactions that take place. Further, even an industrial robot is perceived as a social actor since it can enable the inclusion of people with different disabilities in assembly tasks which is an area that has often excluded people with mobility impairments. The second theme relates to the robot's adaptability to perform not only different tasks but to allow for people with different capabilities to operate it. Thereby, it would ideally combine different levels of automation and allow people with mobility impairments to supervise assembly lines (supervisory control of robots through semi-autonomous control of robotic arms). Additionally, this can create a sense of agency. The third theme revealed that people with mobility impairments lean toward relying on the robot to detect or assist human operators to solve human-robot problems deriving from the interaction.

To conclude, these resulting themes aim to show the views of an underrepresented group of people and influence design decisions about a human-robot collaborative workplace for people with mobility impairments.

4.2 Synopsis of Publication II.

Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects

Arevalo Arboleda Stephanie, Pascher Max, Baumeister Annalies, Klein Barbara, and Gerken Jens. 2021. Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects. In The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA 2021). Association for Computing Machinery, New York, NY, USA, 147–155. DOI: 10.1145/3453892.3458044.

Background

Research in human-robot collaboration has focused on technology-related aspects, thereby relegating those that are human-related. When designing a human-robot collaborative space for people with motor disabilities, a user-centered approach has the potential to create an environment where designers and users exchange ideas. That is why we considered participatory design techniques to design human-robot collaborative environments in three different scenarios: a home setting (assisting eating and drinking), a workplace (teleoperation of robotic arms), and addressing a mobility issue (crossing streets in a sensor-enhanced wheelchair).

Approach

Having different scenarios in HRC for people with motor disabilities allowed us to explore different methodologies and expand our view on the challenges that may arise. In the home project, our goal was to assist people with motor disabilities with food consumption. Here, we involved 15 participants with quadriplegia and their caregivers contacted through a hospital. We used interviews, surveys, and in-situ participatory observations, wherein we employed pictures, videos, google cardboard, VR simulations, and a Kuka iiwa robotic arm. In the workplace project, we focused on developing an interaction concept that allows people with motor disabilities to teleoperate robotic arms for assembly tasks. To achieve that, we partnered with a sheltered workshop and involved 3 primary users (people with motor disabilities) and 6 secondary users (social workers, engineers, supervisors). We chose to use participatory observations and design workshops, where we employed pictures, videos, the Microsoft HoloLens, and a Kuka iiwa robotic arm. In the crossing streets aid project, the goal was to design an assistive robotic wheelchair that allows wheelchair users to safely cross streets. To reach out to potential participants, we got involved in a Reddit community and also contacted 4 participants with different motor disabilities from a sheltered workshop. The methods we used were interviews and participatory observations. The materials used were videos and virtual reality simulations.

Results

In developing these projects, we encountered different challenges related to the pragmatic use of participatory design for people with motor disabilities. We grouped these challenges into three topics: (1) participants, (2) methods, and (3) materials. Regarding (1) participants, we encountered challenges in terms of planning and participation, and preferences of stakeholders. Concerning (2) methods, we found challenges using various participatory design techniques. Pertaining to (3) materials, we experienced challenges related to technology exposure. Finally, we also discussed ethical, legal, and social implications as a crucial topic when designing for people with disabilities.

These challenges led in turn to five lessons related to each challenge type: With regards to planning and participation, we learned to (1) approach participants through different channels and allow for multidisciplinarity in the research team. Regarding the preferences of stakeholders, (2) we should take into account all opinions and manage design sessions to favor the flow of ideas and avoid bottleneck discussions. Concerning using various participatory design techniques, (3) we highlight the consideration of social dependencies in the selection of a particular technique. Concerning planning and participation and using various participatory techniques, (4) it is imperative to plan for early exposure to robots and other technology. This can have a drastic effect not only on the technology to be developed but its perception. Finally, related to ethical, legal, and social implications, (5) we emphasize acknowledging that ethical implications go beyond consent.

In conclusion, we report our experiences using participatory design with people with motor disabilities in human-robot collaboration. These lessons could serve early-career researchers as guidelines when designing human-robot collaboration for people with disabilities.

4.3 Synopsis of Publication III.

What's behind a choice? Understanding Modality Choices under Changing Environmental Conditions

Arevalo Arboleda Stephanie, Miller Stanislaw, Janka Martha, and Gerken Jens. 2019. What's behind a choice? Understanding Modality Choices under Changing Environmental Conditions. In 2019 International Conference on Multimodal Interaction (ICMI '19). Association for Computing Machinery, New York, NY, USA, 291–301. DOI: 10.1145/3340555.3353717.

Background

Hands-free modalities have been widely explored in multimodal interaction in terms of efficiency and combination. However, little is known about user behavior in an environment that provides freedom to choose from various modalities. That is why we took a closer look at the factors that influence input modality choices. For that, we considered the model of Jameson & Kristenssson (2017) that takes into account features of the situation (characteristics of the environment), consequences for interaction (objective aspects of user’s performance and subjective responses to interaction), and user characteristics (individual skills and preferences) as variables that influence the evaluation of an interaction. Our goal was to better understand the nature of modality choices by evaluating the impact of each variable on the user’s choices.

Approach

We carried out a study with 12 participants to explore modality choices in a hands-free interaction environment. Here, participants could choose and combine freely three hands-free modalities (Gaze, Head Movements, Speech) to execute basic point-and-select actions in a 2D interface. To evaluate modality choices, we manipulated the environment (features of the situation) by presenting different constraints related to time, complexity, noise, a visual constraint, a physical constraint, and a baseline condition without any added constraint. We expected that participants would select a modality based on the type of environment that they were experiencing.

Results

Our results show that the impact of features of the situation is rather minimal compared to the consequences for interaction and user characteristics. Regarding user characteristics, our results show that personal preference is still a defining factor of modality choices—participants prime modalities based on their level of proficiency (individual skill, frequency of use). Concerning consequences for interaction, the real experience—determined by the speed, success rate, and error when using a particular modality, affects the evaluation of a modality, namely the subjective experience, and also impacts the frequency of modality use.

The evaluation of the modality switching behavior revealed that users avoid changing modalities. Particularly under conditions that should prompt modality switching, for instance, when the input modality is affected, e.g., using speech in a noisy environment. Also, when users make a modality switch, user characteristics and consequences of the experienced interaction have a higher impact on the choice made compared to the changes in environmental conditions. Finally, when users switch modalities, we identified different types of switching behaviors: users who deliberately try to find and choose an optimal modality (single switcher), users who try to find optimal combinations of modalities (multiple switcher), and a switching behavior triggered by error occurrence (error biased switcher).

In conclusion, this work provides an analysis that sheds light on the factors that influence the reasoning behind modality priming, which should be considered when designing for multimodal interaction.

4.4 Synopsis of Publication IV.

Opportunities and Challenges in Mixed-Reality for an Inclusive Human-Robot Collaboration Environment.

Arévalo Arboleda, Stephanie, Pascher, Max, and Gerken, Jens. 2018. Opportunities and Challenges in Mixed-Reality for an Inclusive Human-Robot Collaboration Environment. In 1st. International Workshop on Virtual, Augmented, and Mixed Reality for Human-Robot Interaction (VAM-HRI), held in conjunction with ACM/IEEE Human-Robot Interaction (HRI 2018).

Background

Mixed reality is a promising approach to human-robot interaction. A possible application of mixed reality is to enhance teleoperation by enhancing the visual space of the robot operator. In this work, we present some opportunities and challenges in interaction design to achieve an inclusive human-robot collaborative environment. We envision a workplace environment where people with severe motor impairments could control robotic arms using mixed reality. A challenge in robot teleoperation is the lack of a simple user interface that allows robot operators to have robot controls within their line of sight, avoiding unnecessary gaze switches.

Approach

We consider that mixed reality could reduce the gap between input and output interaction spaces by providing an interface within the operator’s line of sight using augmented reality head-mounted displays such as the Microsoft HoloLens. Further, presenting information through these types of displays could reduce cognitive load by addressing divided attention. However, using mixed reality to present a user interface also introduces some challenges

Results

We present some challenges and opportunities related to the design of a human-robot collaborative environment for people with severe motor impairments. One of them arises from the way the information is laid to the operator in the user interface. That is why, we propose three designs of user interfaces: (1) a centralized interface similar to the interfaces displayed on control pads, (2) a distributed interface with robot controls attached to the real-world objects, (3) an egocentric interface, providing an egocentric view of the environment from the robot’s perspective.

Another challenge comes in terms of input modalities. People with severe motor impairments have limited to no mobility of their upper and lower limbs, leaving aside traditional hand-based modalities. This leads us to consider using other modalities such as gaze, head movements, and speech. These should not be considered as single interaction possibilities but as complementary leaning towards robot control by using a multimodal approach. However, each of the aforementioned modalities comes with its own complexities in a human-robot interaction environment, e.g, degrees of freedom of the robotic arm, the type of gaze interaction (smooth pursuit, fixations), and controlling multiple robotic arms.

Taken together, this work presents possible opportunities and challenges of using mixed reality in human-robot collaboration to present a user interface that allows robot control.

4.5 Synopsis of Publication V.

There's More than Meets the Eye: Enhancing Robot Control through Augmented Visual Cues

Arévalo Arboleda, Stephanie, Dierks, Tim, Rücker, Franziska, and Gerken, Jens. 2020. There's More than Meets the Eye: Enhancing Robot Control through Augmented Visual Cues. In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20). Association for Computing Machinery, New York, NY, USA, 104–106. DOI:10.1145/3371382.3378240.

Background

In teleoperation, the robot operator can be distant or detached from the scene where manipulation actions occur, affecting the perception of distance and object visual relationships. Augmented reality can enrich the visual space by providing information about the environment and the robot to the operator.

Approach

Our approach explores the design space through visual cues using augmented reality (augmented visual cues) to improve the operator‘s situation awareness and task performance in pick-and-place tasks. For that, we considered Walker et al. (2018) design framework for augmented reality in human-robot interaction augmenting the robot, augmenting the environment, and augmenting the user interface. We augment the robot to assist in picking objects and augment the environment to place the grasped objects in a specific area. Augmenting the robot aims to improve the understanding of the position of the robot’s gripper relative to the object to be grasped. For this, we virtually project a green beam starting at the grippers’ fingers until it reaches a contact surface. Additionally, we show a reflection of the grip region–presented through a shadow over the contact surface, and change the beam’s color to red in case of potential collisions. For augmenting the environment, we present the future status of an object when trying to place it in a specific location. Specifically, we show a “ghost projection” of the grasped object–displayed at the cursor’s position, i.e., where the operator is currently pointing. Also, we show the “ghost projection” in green to indicate that the target position is within reach and red when it is not. This allows not only to see where the object would be placed but aids in evaluating its spatial placement. We evaluated these design concepts in a within-subjects experiment with 10 participants. We implemented a co-located space with a virtual robot arm and virtual objects. We evaluated a peg-in-hole task, using augmented visual cues and the lack of them, with three different shaped objects of increasing difficulty levels. We analyzed the impact of our designs of cues on situation awareness using a Situation Awareness Global Assessment Technique (SAGAT) (Endsley 1995) and task performance by collecting the number of actions (commands used to move, pick, rotate, or place) and success rate (yes, no).

Results

Our results show that there is no significant difference in the average number of actions, success rate, or situation awareness during picking. A reason for this may be that the task complexity did not demand additional visual cues. Whereas in placing, our augmented visual cues resulted in a higher success rate with a higher number of actions–this increase may be necessary to improve the success rate, i.e., making adjustments based on richer feedback. Regarding situation awareness, we found a significant difference favoring the visual cues in terms of projection of the future state. Also, our designs showed positive opinions regarding certainty and usefulness, suggesting that this approach could be effective but should be tested with tasks requiring higher complexity levels.

4.6 Synopsis of Publication VI.

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

Arévalo Arboleda, Stephanie, Dierks, Tim, Rücker, Franziska, & Gerken, Jens. 2021. 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. Human-Computer Interaction – INTERACT 2021. INTERACT 2021. Lecture Notes in Computer Science, vol 12932. Springer, Cham. DOI: 10.1007/978-3-030-85623-6_31.

Background

The teleoperation of a robotic arm can be a challenging task that requires high levels of attention, expertise, and a good understanding of the environment, the robot, and the objects in it. In co-located spaces, robot operators acquire this information visually by walking around the area or having different angles from a camera. However, visually exploring the environment is compromised when the operators have a fixed viewpoint due to physical mobility or environmental restrictions. This affects how objects are perceived in space and depth perception.

Approach

In robot teleoperation for pick-and-place tasks, spatial abilities are a determinant factor. Thus, we aim to (1) support depth perception abilities by using augmented reality and (2) balance the limitations of having fixed viewpoints by gaining a better understanding of the role of pose and distance in co-located workspaces. For this, we propose three designs of visual cues that build upon each other and consider design elements that could improve depth perception. These aim to facilitate the positioning of the gripper above a target object before attempting to grasp it. The designs we propose include a virtual circle (Circle), virtual extensions (Extensions) from the gripper's fingers, and a mixed reality colormap (Colors). The Circle uses casting shadows. It casts the real grip region through a virtual representation displayed under the real gripper through a circle with the same diameter as the grip region. The Extensions build upon the Circle and are a symbolic enhancement of the gripper's fingers. These present a virtual elongation of the end effector, i.e., we display two planes that start from the gripper's fingers and one from the center of the gripper towards the workspace. The Colors build upon the Circle and Extensions and are grounded on the use of physical and virtual landmarks that connect the gripper and the workspace by matching colors. We present small circles at the end of the virtual extensions that adopt the color displayed on the physical colormap. In order to evaluate these designs of cues and investigate the relationship between distance and the target's pose, we conducted a within-subjects experiment with 24 participants. We collected measures of time, success, and perceived certainty while performing grasping tasks.

Results

Our results show that a shorter distance leads to higher success, faster grasping time, and higher certainty. Concerning the target object's pose, a clear pose leads to higher success and certainty, but interestingly slower task times. Regarding the design of cues, our results reveal that the simplicity of the Circle cue leads to the highest success and outperforms the most complex cue, Colors. Additionally, our findings indicate that the level of certainty depends on the distance rather than the type of cue. This work can serve as an initial analysis to further explore distance, pose, and certainty when designing to improve depth perception within a co-located teleoperation context.

4.7 Synopsis of Publication VII.

Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues

Arévalo Arboleda, Stephanie, Rücker, Franziska, Dierks, Tim, and Gerken, Jens, (Eds. 2021). Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '21), ACM, NY. DOI: 10.1145/3411764.3445398.

Background

A prevalent problem in teleoperation is distance and depth estimation in manipulation tasks. The human eye acquires information about distance and depth through monocular and binocular cues. Monocular cues are used to perceive distances within 3 m and they can be further divided into static or dynamic (acquired by motion). The absence of dynamic cues reduces the ability to accurately estimate the location of objects. In co-located scenarios where the operators have a fixed viewpoint and are distant from the workstation, they experience depth misestimations that affect accuracy in manipulation tasks.

Approach

We present two designs of visual cues using augmented reality named Augmented Visual Cues (AVC) that can provide better awareness of objects' position and depth in a co-located workspace. Both designs are inspired by the monocular cues of lighting and interposition. The first design, basic cues, provides indirect hints about certain characteristics of the workspace and the objects on it. The basic cues' main characteristic combines physical and virtual landmarks by presenting a physical colormap displayed on the workspace and a virtual colormap. The real and virtual colormap can help operators realize the position of the gripper relative to a target by matching virtual with real colors. The second design, advanced cues, provides explicit hints for interaction that are environmentally aware. Here, we use object and pose recognition to provide explicit hints. The main characteristic of the advanced cues is a virtual copy of the grasped object and a virtual robot gripper to point at potential collisions. In order to evaluate these designs, we carried out a user study with a between-subjects design. Here, participants wore the Microsoft HoloLens 1 to teleoperate a robotic arm. We evaluated our designs against a baseline condition (without cues) in a pick-and-place task that required some level of precision. We hypothesized about the improvement of task performance at different levels in objective and subjective measures using both designs of augmented visual cues.

Results

Our findings show that both of our designs of AVC improve task performance in terms of accuracy. Our subjective measures hinted at nuances in participants' perception of self-performance and the usefulness of both designs of AVC. This led us to further explore the participants' views through sentiment analysis. Both designs of AVC gathered positive sentiments about the system. Nonetheless, we found differences in participants' feelings towards teleoperating a robot with basic and advanced cues. In particular, our advanced cues evoked a positive sentiment while the basic cues prompted rather mixed sentiments. A possible reason behind this can be found in the results of our thematic content analysis. These results revealed that our design of basic cues was regarded as complex, and we attribute this to the number of hints presented to the operator. Overall, our findings show that both basic and advanced cues offer a promising approach to enhancing the visual space. They help to understand the relation between the environment and the robot in the workspace, thus, assisting the teleoperation of a robotic arm and increasing accuracy.

4.8 Synopsis of Publication VIII.

Does one size fit all? A Case Study to Discuss Findings of an Augmented Hands-free Robot Teleoperation Concept for People with and without Motor Disabilities

Arévalo Arboleda, Stephanie; Becker, Marvin; and Gerken, Jens. Does One Size Fit All? A Case Study to Discuss Findings of an Augmented Hands-Free Robot Teleoperation Concept for People with and without Motor Disabilities. 2022. Technologies, 10(1), 4. DOI:10.3390/technologies10010004.

Background

Many people with motor disabilities maintain the ability to move their heads, use eye gaze, and speak to different extents. People with motor disabilities often use these abilities to interact with the environment. We capitalize on that and consider them as potential hands-free modalities that can be used to teleoperate robotic arms and manipulate objects. Added to that, we consider that an augmented interface for robot control can ease teleoperation. Combining multimodal hands-free interaction with augmented reality has the potential to increase autonomy for people with motor disabilities. However, the experiences evoked by hands-free multimodal interaction and augmented reality can vary significantly for people with and without motor disabilities.

Approach

We present a perspective from a person with multiple sclerosis, Miss L., on a multimodal hands-free teleoperation concept that uses a robotic arm, augmented reality, and multimodal robot control. She teleoperated a robotic arm using a virtual menu with robot controls together with head movements, voice commands, and augmented visual cues presented through the Microsoft HoloLens 2. The task involved the manipulation of an “L-shaped object,” and we collected subjective measures to evaluate the robot's workload, usability, and perception. Also, we carried out an interview.

We present the findings from Miss L.’s experience and put them in context with those of people without disabilities from a previous study. We divided these findings into two topics, hands-free multimodal teleoperation and using augmented visual cues. Additionally, we provide learning points when evaluating interaction designs with people with and without disabilities.

Results

Miss L.’s impressions suggest the potential of using augmented reality for human-robot interaction for people with motor disabilities. Overall the subjective measures collected point at positive impressions of the interaction concept. However, we also report the challenges that Miss L. experienced related to specific design decisions of our interaction concept.

When comparing the findings of people without disabilities and Miss L., we found that both did not experience problems using head orientation. However, the main difference lies in the use of voice commands. She expressed that her voice is rather feeble and led to failures in recognizing the voice commands. Concerning the use of augmented reality to present the robot control interface, Miss L. also reported problems identifying the virtual cursor—needed to use the dwell buttons to control the robot. Nevertheless, Miss L. and people without disabilities reported improvements in depth perception derived from the use of augmented visual cues.

These similarities and differences led us to provide some learning points when comparing the experiences of people with and without disabilities. First, re-evaluate the experimental metrics. We highlight the importance of subjective measures and qualitative analysis, as these can highlight aspects that challenge the usability of a solution. Second, beware of the bias. People with disabilities may be eager to be included in design decisions as they are often excluded. However, this could lead to a rather positive bias when evaluating interaction concepts or other technological solutions. Third, consider variability in abilities evoking different experiences. A base interaction that allows for adaptations may be a solution that considers the individual abilities of people with and without disabilities.

5. DISCUSSION & LIMITATIONS

This section is dedicated to reviewing the findings of each publication. First, it sets each publication and its findings in context with the objectives of this research. Second, it extends discussions about specific findings from each publication. Finally, it presents general limitations and suggests some directions for future work.

5.1 Objective 1. Understanding HRC for PMD at the Workplace

In our publications I and II, we aimed to expand the knowledge about HRC for PMD. First by evaluating the needs and expectations of a human-robot collaborative workplace, and second by presenting some challenges and lessons learned while designing with and for PMD.

In Publication I, we present potential concerns of PMD in a human-robot collaborative workplace, which were grouped into three themes: the robot fitting in the social and organizational structure, human-robot synergy, and human-robot problem management. These themes point out concerns about how a robot will alter pre-existent social and practical interactions. This concern may be inherent to sheltered workshops, as these are non-traditional workplaces with a strong sense of community that fosters support and teamwork. However, we believe that the presence of a robot may indeed influence the social interactions taking place at any workplace. The *Computers are Social Actors* (CASA) framework indicates that humans apply social norms to computers as they would do to human peers (Nass et al. 1994). However, these social norms towards computers have changed due to the advances and widespread of technology which in turn affect interaction (Gambino et al. 2020). Our findings point out that even industrial robotic arms bring concerns about social interactions highlighting inclusion and adaptability. We consider that robotic arms in workplace settings can be inclusion enablers for PMD. This provides robotic arms with social facilitation characteristics, namely, improving the individual performance in the presence of a co-actor (robotic arm). This may align with Gambino et al.'s (2020) remarks that argue that social interactions with robots have changed over time, bringing new social affordances that do not necessarily follow human-to-human interactions.

In publication II, we report the challenges and lessons learned when applying user-centered design with PMD for HRC. We examined different potential scenarios for HRC and provided lessons learned related to planning and participation, preferences of stakeholders, using various PD techniques, and ethical, legal, and social implications. In our publication, we did not consider culture or age as demographic factors that may have influenced our findings. Especially within the participation, technology exposure, and ethical legal and social implications topics. Our participants were located in Germany and our findings might represent expectations, behaviors, and perspectives influenced by the German culture. For instance, a study reported that germans rated robots less trustworthy than Chinese or Korean (Li et al. 2010). Another study about robot acceptance in industrial settings showed that Germany presents notable differences from China, Japan, or the USA. These differences involve a stronger correlation between social implications and robot acceptance (Bröhl et al. 2019).

Our findings from publications I and II contribute to the body of knowledge that aims to design inclusive environments through HRC. We consider that we present a perspective of PMD, their caregivers, and designers, which must be considered when designing technology. Further, we believe that PD and user-centered robotics open up possibilities to design technologies and interactions that go beyond being assistive towards designing inclusive technologies.

5.2 Objective 2. Understand hands-free modality choices and hands-free multimodal teleoperation.

In publication III, we investigated the nature of modality choices and their relation to evaluating a given interaction. Our findings pointed to the user's own abilities and individual experiences as determining factors when evaluating an interaction. Additionally, Publication I and II showed high individual differences in abilities and expectations between PMD. Publication VIII demonstrates the findings of Publications I, II, and III in a case study with an individual with motor disabilities and points out the absence of a "one size fits all" solution in multimodal interfaces. Further, we consider that Publication III and VIII contribute to the efforts of increasing the knowledge of multimodal interaction and thus can serve as a baseline to design adaptable interfaces.

A multimodal context-sensitive UI could provide a solution that adapts to users' current abilities and experiences. In the last years, "plasticity of user interface design" or adaptive UI are terms that have become popular to describe adaptable interfaces. On the one hand, the attribute of plasticity is given to UIs that can be adapted to different users with different abilities or the same user with changing abilities, yet the UI will have a base state that it will always come back to (Miraz et al. 2021). On the other hand, adaptive UIs are closely related to intelligent UIs, which can be defined as interfaces that can learn from the user and adapt to fit the individual characteristics of users (Miraz et al. 2021). In order to achieve these types of adaptations, Oppermann (1994) stated three essential adaptivity components: afferential, which collects previous user data; efferential, which is related to seamless automation to the users; and inferential, related to continued monitoring of interaction. Designing UIs with these components may be a goal that can be achievable with the widespread of artificial intelligence and machine learning approaches.

5.3 Objective 3. Using AR to enhance the visual space for co-located teleoperation and improve task performance

Publication IV - VII explore the operator's visual space by using AR to facilitate teleoperation and improve task performance.

AR in HRI takes a crucial role as it considers real and virtual elements related to the environment, the user, and the robot. Publication IV presents variations of a UI for teleoperation, one with user-anchored interface elements and another with environment-anchored interface elements. These designs contributed to the analysis presented by Williams et al. (2019), who introduced a conceptual framework that considers the plane of interaction and the MR interaction design elements.

Another crucial aspect of robot teleoperation is how the UI could be controlled. A multimodal approach to robot teleoperation may complement the use of some modalities. We initially considered three hands-free modalities gaze, head movements, and speech. However, we opted for using head movements and speech as input modalities. Head movements complement eye gaze for pointing and can be coordinated with speech (Qvarfordt 2017). Further, using speech for direct manipulation tasks, such as those involved in teleoperation, is a multimodal approach that reduces the limitations of a single modality use (Cohen et al. 1989). We did not pursue using gaze as an interaction modality due to the Midas-touch problem (Jacob 1990). Also, robot teleoperation entails per se high levels of attention and expertise, which affect mental workload. Additionally, an assembly environment could involve many visual distractors, e.g., other workers, different materials, and other machinery. Previous research highlights the limits of gaze interaction in the presence of visual distractors when performing tasks that involve a high mental workload (Trosterer and Dzaack 2011). Head movements and speech were used as interaction modalities in Publications V-VIII.

Publication V focused on enhancing the environment and the robot for the operator. One of the main aspects was exploring SA variables in a virtual reality environment, wherein we provided visual cues and tested the lack of them to perform pick-and-place tasks. Here, we would like to further discuss the evaluation of SA. We used SAGAT (Situation Awareness Global Assessment Technique) which prompted results that did not find conclusive improvements in terms of SA. We attribute this to the queries used to evaluate SA. There is a rather ample range of results depending on the type of queries asked. Although this can be considered as part of the technique, it has been mentioned that SAGAT indicates behavior rather than SA (Kellogg and Whittaker 2000). Moreover, Mott et al. (2021) also reported inconclusive results when using similar techniques such as QASAGAT (Quantitative Analog Situation Awareness Global Assessment Technique) and SPASA (Short Post-Assessment of Situation Awareness) (Gatsoulis et al. 2008). Looking retrospectively, we believe that we could have used different and complementary techniques to evaluate SA. However, our findings align with those of Mott et al. (2021) and point out that scenarios, type, and level of difficulty of tasks are crucial when evaluating SA. Further, more complex environments could allow more insightful findings on the different levels of SA (perception, comprehension, and projection) (Endsley 1995).

Building upon Publication V, in Publication VI we considered visualization techniques that can be used to enhance the visual space for robot teleoperation. Additionally, we considered depth-related variables (distance and target's pose) to better understand the role of certainty in picking tasks. Certainty is a controversial factor that can easily fluctuate depending on other subjective factors such as attention. Spatial visual attention, i.e., awareness of elements in the environment, can influence certainty. Rahnev et al. (2011) carried out a series of experiments and found that humans overestimate what their senses perceive in the absence of attention. We did not evaluate levels of attention in our experiment. In retrospect, for a better understanding of certainty, we could have measured attention by considering the pre-and-post-stimulus allocation of attention resources (Wyart and Tallon-Baudry 2009). We believe that fluctuating attention could have influenced participants' levels of certainty in our picking task, as they exhibited higher levels of attention during the first trials. There have been studies with mixed results about the association of attention and certainty, as well as with certainty in general. Some have found that attention has a rather strong direct effect on certainty (Denison et al. 2018; Zizlsperger et al. 2012) while others have found the opposite effect (Wilimzig et al. 2008). We believe that certainty is an ongoing area of interest for HCI researchers which should receive more attention when evaluating UI and performance.

Publication VII puts together our previous findings of hands-free multimodal teleoperation and improving depth perception through augmented visual cues. In our paper, we presented two types of cues: basic cues that suggested characteristics of the environment (distance and depth) related to the position of the gripper, and advanced cues that presented explicit cues by showing a virtual representation of the gripper's position on the workspace. Our findings indicated improvements in performance using both types of cues. We argue that teleoperation using our basic cues could be an asset to perform intervention tasks or even when a robotic arm is used for manipulation tasks in unknown (for a robotic system) workspaces, as it provides the operator higher control in teleoperation. Whilst in known or static workspaces, the advanced cues could simplify manipulation tasks. In fact, our advanced cues showed higher performance in the manipulation task and evoked higher user preference. Further, our advanced cues involve a higher level of automation when compared to the basic cues. Thereby, we locate the concept of advanced cues at the decision support level and our basic cues at the action support level within Beer et al.'s (2014) taxonomy of levels of robot autonomy for HRI.

Adjustable or variable autonomy involves varying levels of robot autonomy considering the capabilities of humans and robots (Luck et al. 2003). Considering adjustable autonomy in HRC can be a way to favor the inclusion of people with disabilities. Our designs of augmented visual cues do not intend to consider the abilities of PMD as a generalizable ensemble but take into account the different degrees of severity within a medical diagnosis or even changes in the energy level of an individual. Our experiences researching with and for PMD have shown that their level of energy and abilities are highly fluctuating over time. One example of this is presented in Publication VIII, where we report the experiences of a person with motor disabilities using not only our teleoperation concept but one of our designs of augmented visual cues. Our findings pointed out the influence of these fluctuations on the usefulness and usability of an interaction, in our case hands-free teleoperation and an AR UI. We are confident that our findings can further contribute to the efforts of creating human-centered "adjustable autonomy" in HRC.

In this chapter, we approached adaptability as a crucial factor in UI design and teleoperation. Regarding the needs and preferences of PMD, one of our main findings was adaptability as a social facilitator. In terms of multimodal interfaces, we discussed preferences for adaptable UI and the attribute of plasticity when designing for multimodality. Finally, concerning enhancing teleoperation through augmented visual cues, we brought to light the topic of adjustable autonomy, which would provide different levels of agency to robot operators.

We highlight a finding across experiments when using speech and head movements as input modalities. Inexperienced or first-time users of the Microsoft HoloLens did not experience significant issues using voice commands or head movements to teleoperate the robotic arm and promptly were able to perform the tasks. We consider that using hands-free teleoperation is a promising multimodal interaction not only for robot control but for interacting in MR environments.

5.4 Limitations

The use of current ARHMDs presents some limitations in HRI. Williams et al. (2020) highlight that ARHMDs do not cover peripheral vision, the average visual fidelity of virtual objects could be improved, and the narrow field of view does not allow for a fully immersive experience. Additionally, from our experiences, tasks that require precision are challenging when performed further than 2 m, since the fidelity of the virtual objects decreases, adding problems to depth perception. Further, using AR for manipulation tasks can add fatigue and eye strain to users. We experienced similar challenges to the ones found by Williams et al. (2020), e.g., focal rivalry (the human eye cannot focus on both virtual and real objects at the same time) and vergence accommodation conflicts (most lenses in ARHMD have fixed focal length but the human eyes turn inwards to triangulate on an object when it approaches).

We recognize that the use of voice commands is limiting in noisy environments or frustrating when commands are not executed or misunderstood (filtering problems). Further, PMD could have added cognitive or physical disabilities that limit their use of speech or ability to clearly articulate words. We use the speech modality as a possible approach for hands-free interaction, which we recognize might not work in all cases. Similar limitations may apply to the use of head movements as the interface would be challenging to use for people with temporal neck stiffness or permanent conditions. Hence, we advocate for a multimodal approach that can be adapted and changed to the personal and temporal requirements of PMD.

We did not include participants with motor disabilities in all of our studies due to (1) the high variability between participants and (2) the access to this group of participants during SARS-COVID. The high variability in abilities within PMD led us to consider a lead group in some of our studies. We do not intend to provide a solution that can be generalized to every person with a motor disability, but to invite reflection on the absence of a unique solution. Moreover, these findings present first-hand experiences of PMD which should be considered in the creation of inclusive environments in the workplace. Notwithstanding, carrying out studies with people without disabilities allows us to generalize the usefulness of our interaction designs to hands-busy situations.

5.5 Future Directions

There are different taxonomies in HRI (Yanco and Drury 2004) that could have been considered in this work. One of them is robot morphology, as it could determine social expectations. Publication I pointed to the relevance of the social and organizational in HRC for PMD. Hence, different robot morphologies could have been evaluated to extend the knowledge in this category. Additionally, Chita-Tegmark and Scheutz (2020) pointed out the potential of assistive robots as mediators for the social behavior management of people with health conditions. Thus, we consider that there is still space to explore the robot assistance's role in facilitating social interactions in HRC at the workplace, especially for people with disabilities.

Another taxonomy in HRI that could be considered is the physical space. Our scenario considers a co-located space for teleoperation. However, remote teleoperation could also be an option for PMD. This could minimize safety concerns, but it could exacerbate depth perception and SA issues (Graf and Hussmann 2020). Both factors could increase the mental workload, especially for PMD. Nevertheless, considering remote teleoperation could create new hypotheses and extend this research to include people with other types of disabilities who may not be able not physically attend to a workplace.

There is a wide range of levels of autonomy or the amount of intervention that could have been considered. An example of this is the level of emotional engagement that the level of autonomy involves in HRC. For instance, previous work has found that operators felt a higher social presence when teleoperating a robot than when engaging with an autonomous robot (Choi et al. 2014). These could bring a new set of research questions about the level of engagement of people with disabilities not only emotionally but to self-assess their own performance and the human-robot team performance.

Building upon the findings of depth perception (Pubications V – VII), we consider it relevant to look further into the visual perception of partially occluded objects.

A problem derived from having a fixed viewpoint is accurately determining the shape and pose of target objects in a loaded workspace to grasp them successfully. Regular workspaces can present a variety of elements that could be distributed in a cluttered manner. This causes some or various elements might partially or completely obstruct the view of other objects (object occlusion). Object occlusion hinders recognizing with precision the shape and pose of a target object which might lead to failures in grasping tasks or multiple trial-and-error attempts.

An option to tackle object occlusion is to provide robot operators with visual and auditory hints (multimodal hints). These hints aim to alert the robot operator about an occluded object's existence, shape and pose. This concept is inspired by real-world deictic gestures (pointing, verbal descriptions) used to indicate the presence of an object in a given space. These hints could be displayed using ARHMDs.

6. BIBLIOGRAPHY

Abras, Chadia; Maloney-Krichmar, Diane; Preece, Jenny (2004): User-centered design. In *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks* 37 (4), pp. 445–456.

Akella, P.; Peshkin, M.; Colgate, E.; Wannasuphoprasit, W.; Nagesh, N.; Wells, J. et al. (1999): Cobots for the automobile assembly line. In : Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C). International Conference on Robotics and Automation. Detroit, MI, USA, 10-15 May 1999: IEEE, pp. 728–733.

Arnheim, Rudolf; McNeill, David (1994): Hand and Mind: What Gestures Reveal about Thought. In *Leonardo* 27 (4), p. 358. DOI: 10.2307/1576015.

Azuma, Ronald T. (1997): A survey of augmented reality. In *Presence: Teleoperators & Virtual Environments* 6 (4), pp. 355–385.

Badke-Schaub, Petra; Hofinger, Gesine; Lauche, Kristina (Eds.) (2012): Human factors. Psychologie sicheren Handelns in Risikobranchen. 2. überarbeitete Aufl. Berlin: Springer.

Bartneck, C.; Forlizzi, J. (2004): A design-centred framework for social human-robot interaction. In : RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759). RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication. Kurashiki, Okayama, Japan, 20-22 Sept. 2004: IEEE, pp. 591–594.

Bartneck, Christoph (2020): Human-robot interaction. An introduction. Cambridge United Kingdom, New York NY USA: Cambridge University Press.

Bauer, Andrea; Wollherr, Dirk; Buss, Martin (2008): Human–Robot Collaboration: A Survey. In *Int. J. Human. Robot.* 05 (01), pp. 47–66. DOI: 10.1142/S0219843608001303.

Bauer, Wilhelm; Bender, Manfred; Braun, Martin; Rally, Peter; Scholtz, Oliver (2016): Lightweight robots in manual assembly–best to start simply. In *Frauenhofer-Institut für Arbeitswirtschaft und Organisation IAO, Stuttgart*.

Beer, Jenay M.; Fisk, Arthur D.; Rogers, Wendy A. (2014): Toward a framework for levels of robot autonomy in human-robot interaction. In *Journal of human-robot interaction* 3 (2), pp. 74–99. DOI: 10.5898/JHRI.3.2.Beer.

Belouaer, Lamia; Brosset, David; Claramunt, Christophe (2013): Modeling Spatial Knowledge from Verbal Descriptions. In Thora Tenbrink (Ed.): Spatial information theory. 11th international conference, COSIT 2013, Scarborough, UK, September 2-6, 2013, proceedings / Thora Tenbrink [and three others] (eds.), vol. 8116. Heidelberg: Springer (LNCS sublibrary: SL 1 - Theoretical computer science and general issues, 8116), pp. 338–357.

Bernardos, Ana M.; Gómez, David; Casar, José R. (2016): A Comparison of Head Pose and Deictic Pointing Interaction Methods for Smart Environments. In *International Journal of Human-Computer Interaction* 32 (4), pp. 325–351. DOI: 10.1080/10447318.2016.1142054.

Bolt, Richard A. (1980): “Put-that-there”. In *SIGGRAPH Comput. Graph.* 14 (3), pp. 262–270. DOI: 10.1145/965105.807503.

Breazeal, C. (2004): Social Interactions in HRI: The Robot View. In *IEEE Trans. Syst., Man, Cybern. C* 34 (2), pp. 181–186. DOI: 10.1109/TSMCC.2004.826268.

Bröhl, Christina; Nelles, Jochen; Brandl, Christopher; Mertens, Alexander; Nitsch, Verena (2019): Human–Robot Collaboration Acceptance Model: Development and Comparison for Germany, Japan, China and the USA. In *Int J of Soc Robotics* 11 (5), pp. 709–726. DOI: 10.1007/s12369-019-00593-0.

Charalambous, George; Fletcher, Sarah; Webb, Philip (2016): The Development of a Scale to Evaluate Trust in Industrial Human-robot Collaboration. In *Int J of Soc Robotics* 8 (2), pp. 193–209. DOI: 10.1007/s12369-015-0333-8.

Chita-Tegmark, Meia; Scheutz, Matthias (2020): Assistive Robots for the Social Management of Health: A Framework for Robot Design and Human-Robot Interaction Research. In *Int J of Soc Robotics*, pp. 1–21. DOI: 10.1007/s12369-020-00634-z.

Choi, Jung Ju; Kim, Yunkyung; Kwak, Sonya S. (2014): The autonomy levels and the human intervention levels of robots: The impact of robot types in human-robot interaction. In : The 23rd IEEE International Symposium on Robot and Human Interactive Communication. 2014 RO-MAN: The 23rd IEEE International Symposium on Robot and Human Interactive Communication. Edinburgh, UK, 25.08.2014 - 29.08.2014: IEEE, pp. 1069–1074.

Cohen, P. R.; Dalrymple, M.; Moran, D. B.; Pereira, F. C.; Sullivan, J. W. (1989): Synergistic use of direct manipulation and natural language. In *SIGCHI Bull.* 20 (SI), pp. 227–233. DOI: 10.1145/67450.67494.

Denison, Rachel N.; Adler, William T.; Carrasco, Marisa; Ma, Wei Ji (2018): Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. In *PNAS* 115 (43), pp. 11090–11095. DOI: 10.1073/pnas.1717720115.

Dorrington, Peter; Wilkinson, Christopher; Tasker, Lorna; Walters, Andrew (2016): User-Centered Design Method for the Design of Assistive Switch Devices to Improve User Experience, Accessibility, and Independence. In *Journal of Usability Studies* 11 (2), pp. 66–82.

Duque, Ezequiel; Fonseca, Guilherme; Vieira, Heitor; Gontijo, Gustavo; Ishitani, Lucila (2019): A systematic literature review on user centered design and participatory design with older people. In Rodrigo Laiola Guimaraes, Celso Alberto Saibel Santos, Vagner Figueredo de Santana, Artur Kronbauer (Eds.): Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems. IHC '19: XVIII Brazilian Symposium on Human Factors in Computing Systems. Vitória Espírito Santo Brazil, 22 10 2019 25 10 2019. New York, NY, USA: ACM, pp. 1–11.

Elepfandt, Monika (2012): Pointing and speech. In Lone Malmborg (Ed.): Proceedings of the 7th Nordic Conference on Human-Computer Interaction Making Sense Through Design. the 7th Nordic Conference. Copenhagen, Denmark, 10/14/2012 - 10/17/2012. New York, NY: ACM, p. 807.

Endsley, M. R.; Kaber, D. B. (1999): Level of automation effects on performance, situation awareness and workload in a dynamic control task. In *Ergonomics* 42 (3), pp. 462–492. DOI: 10.1080/001401399185595.

Endsley, Mica R. (1995): Toward a Theory of Situation Awareness in Dynamic Systems. In *Hum Factors* 37 (1), pp. 32–64. DOI: 10.1518/001872095779049543.

Endsley, Mica R.; Garland, Daniel J. (2000): Situation Awareness Analysis and Measurement: CRC Press.

Endsley, Mica R.; Jones, Debra G. (2012): Designing for situation awareness. An approach to user-centered design / Mica R. Endsley and Debra G. Jones. 2nd ed. Boca Raton, FL: CRC Press.

Fong, Terrence; Thorpe, Charles; Baur, Charles (2003): Collaboration, Dialogue, Human-Robot Interaction. In Raymond A. Jarvis, Alexander Zelinsky (Eds.): Robotics research. The Tenth International Symposium / Raymond A. Jarvis, Alexander Zelinsky (eds.), vol. 6. Berlin, London: Springer (Springer Tracts in Advanced Robotics, v. 6), pp. 255–266.

Frennert, Susanne (2020): A Human Factor Approach to HRI. In Adriana Tapus, Elisabeth André, Jean-Claude Martin, François Ferland, Mehdi Ammi (Eds.): Social robotics. 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015 : proceedings / Adriana Tapus, Elisabeth André, Jean-Claude Martin, François Ferland, Mehdi Ammi (eds.). Cham, 2019. Cham: Springer (LNCS sublibrary. SL 7, Artificial intelligence, 8239), pp. 311–321.

Gambino, Andrew; Fox, Jesse; Ratan, Rabindra (2020): Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm. In *HMC* 1, pp. 71–86. DOI: 10.30658/hmc.1.5.

Gao, Peng; Reily, Brian; Paul, Savannah; Zhang, Hao (2020): Visual Reference of Ambiguous Objects for Augmented Reality-Powered Human-Robot Communication in a Shared Workspace. In Jessie Y. C. Chen, Gino Fragomeni (Eds.): Virtual, Augmented and Mixed Reality. Design and Interaction, vol. 12190. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 550–561.

Gatsoulis, Yiannis; Virk, Gurvinder S.; Dehgani, A. A. (2008): The Influence of Human Factors on Task Performance: A Linear Approach. In L. Marques, A. de Almeida, M. O. Tokhi, G. S. VIRK (Eds.): Advances in Mobile Robotics. Proceedings of the Eleventh International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines. Coimbra, Portugal, 8 – 10 September 2008: WORLD SCIENTIFIC, pp. 481–488.

Gleeson, Brian; MacLean, Karon; Haddadi, Amir; Croft, Elizabeth; Alcazar, Javier (2013): Gestures for industry Intuitive human-robot communication from human observation. In : 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI). 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Tokyo, Japan: IEEE, pp. 349–356.

Goodrich, Michael A.; Schultz, Alan C. (2007): Human-Robot Interaction: A Survey. In *Foundations and Trends, in Human-Computer Interaction* 1 (3), pp. 203–275. DOI: 10.1561/1100000005.

Graf, Gaetano; Hussmann, Heinrich (2020): User Requirements for Remote Teleoperation-based Interfaces. In : 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. AutomotiveUI '20: 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. Virtual Event DC USA, 21 09 2020 22 09 2020. New York, NY, USA: ACM, pp. 85–88.

Granda, Thomas M.; Kirkpatrick, Mark; Julien, Tracye D.; Peterson, Larry A. (1990): The Evolutionary Role of Humans in the Human-Robot System. In *Proceedings of the Human Factors Society Annual Meeting* 34 (9), pp. 664–668. DOI: 10.1177/154193129003400919.

Griffin, Z. M.; Bock, K. (2000): What the eyes say about speaking. In *Psychological science* 11 (4), pp. 274–279. DOI: 10.1111/1467-9280.00255.

Gugerty, Leo J.; Tirre, William C. (2000): Individual differences in situation awareness. In *Situation awareness analysis and measurement*, pp. 249–276.

Hansen, John Paulin; Rajanna, Vijay; MacKenzie, I. Scott; Bækgaard, Per (2018): A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In Carlos Morimoto, Thies Pfeiffer (Eds.): Proceedings of the Workshop on Communication by Gaze Interaction - COGAIN '18. the Workshop. Warsaw, Poland, 6/15/2018 - 6/15/2018. New York, New York, USA: ACM Press, pp. 1–5.

Hayes, Brett K.; Heit, Evan; Swendsen, Haruka (2010): Inductive reasoning. In *Wiley interdisciplinary reviews. Cognitive science* 1 (2), pp. 278–292. DOI: 10.1002/wcs.44.

Hedayati, Hooman; Walker, Michael; Szafir, Daniel (2018): Improving Collocated Robot Teleoperation with Augmented Reality. In Takayuki Kanda, Selma Ŝabanović, Guy Hoffman, Adriana Tapus (Eds.): HRI'18. Proceedings of the 2018 ACM/IEEE International Conference on Human Robot Interaction : March 5-8, 2018, Chicago, Il, USA. the 2018 ACM/IEEE International Conference. Chicago, IL, USA, 3/5/2018 - 3/8/2018. New York, New York: Association for Computing Machinery, pp. 78–86.

Hietanen, Antti; Pieters, Roel; Lanz, Minna; Latokartano, Jyrki; Kämäräinen, Joni-Kristian (2020): AR-based interaction for human-robot collaborative manufacturing. In *Robotics and Computer-Integrated Manufacturing* 63, p. 101891. DOI: 10.1016/j.rcim.2019.101891.

Holloway, Catherine (2019): Disability interaction (DIX). In *interactions* 26 (2), pp. 44–49. DOI: 10.1145/3310322.

Iáñez, Eduardo; Furió, M. Clara; Azorín, José M.; Huizzi, José Alejandro; Fernández, Eduardo (2009): Brain-Robot Interface for Controlling a Remote Robot Arm. In J. Mira (Ed.): Third International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2009, Santiago De Compostela, Spain, June 22-2009. Proceedings / José Mira … [et al.] (eds), vol. 5602. Berlin: Springer-Verlag (Lecture notes in computer science, 0302-9743, 5601-5602), pp. 353–361.

International Organization for Standardization (2012): ISO 8373:2012(en), Robots and robotic devices — Vocabulary. Available online at https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:en, updated on 2/19/2021, checked on 2/19/2021.

Jackowski, A.; Gebhard, M.; Gräser, A. (2016): A Novel Head Gesture Based Interface for Hands-Free Control of a Robot. In : 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1–6.

Jacob, Rob; Stellmach, Sophie (2016): What you look at is what you get. In *interactions* 23 (5), pp. 62–65. DOI: 10.1145/2978577.

Jacob, Robert J. K. (1990): What You Look at Is What You Get: Eye Movement-Based Interaction Techniques. In : ACM Press, pp. 11–18.

Jacob, Robert J. K. (1991): The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at Is What You Get. In *ACM Transactions on Information Systems* 9 (2), pp. 152–169. DOI: 10.1145/123078.128728.

Jameson, Anthony; Kristensson, Per Ola (2017): Understanding and supporting modality choices. In Sharon Oviatt, Björn Schuller, Philip R. Cohen, Daniel Sonntag, Gerasimos Potamianos, Antonio Krüger (Eds.): The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1: ACM, pp. 201–238.

Jia, Pei; Hu, Huosheng H.; Lu, Tao; Yuan, Kui (2007): Head gesture recognition for hands-free control of an intelligent wheelchair. In *Industrial Robot* 34 (1), pp. 60–68. DOI: 10.1108/01439910710718469.

Jungk, Robert; Müllert, Norbert R. (1996): Future workshops. How to create desirable futures. London: Institute for Social Inventions.

Kahn, Peter H.; Ishiguro, Hiroshi; Friedman, Batya; Kanda, Takayuki; Freier, Nathan G.; Severson, Rachel L.; Miller, Jessica (2007): Psychological Benchmarks of Human–Robot Interaction. In *IS* 8 (3), pp. 363–390. DOI: 10.1075/is.8.3.04kah.

Kellogg, Wendy; Whittaker, Steve (Eds.) (2000): Proceedings of the 2000 ACM conference on Computer supported cooperative work - CSCW '00. the 2000 ACM conference. Philadelphia, Pennsylvania, United States. New York, New York, USA: ACM Press.

Kidd, Paul; Rahimi, Mansour; Karwowski, Waldemar (1992): Design of human-centered robotic systems. In : human robot interaction: London, UK: Taylor and Francis, pp. 225–241.

Kim, Jeonghee; Park, Hangue; Bruce, Joy; Sutton, Erica; Rowles, Diane; Pucci, Deborah et al. (2013): The tongue enables computer and wheelchair control for people with spinal cord injury. In *Science translational medicine* 5 (213), 213ra166. DOI: 10.1126/scitranslmed.3006296.

Kim, M.; Oh, K.; Choi, J.; Jung, J.; Kim, Y. (2011): User-Centered HRI: HRI Research Methodology for Designers. In Xiangyu Wang (Ed.): Mixed reality and human-robot interaction. Dordrecht: Springer Science+Business Media B.V (Intelligent systems, control and automation: science and engineering, 47), pp. 13–33.

Kita, Sotaro (2008): Pointing. Where Language, Culture, and Cognition Meet. Hove: Taylor and Francis.

Kolbeinsson, Ari; Lagerstedt, Erik; Lindblom, Jessica (2019): Foundation for a classification of collaboration levels for human-robot cooperation in manufacturing. In *Production & Manufacturing Research* 7 (1), pp. 448–471. DOI: 10.1080/21693277.2019.1645628.

Krüger, J.; Lien, T. K.; Verl, A. (2009): Cooperation of human and machines in assembly lines. In *CIRP Annals - Manufacturing Technology* 58 (2), pp. 628–646. Available online at https://www.academia.edu/28599686/Cooperation_of_human_and_machines_in_assembly_lines.

Krupke, Dennis; Steinicke, Frank; Lubos, Paul; Jonetzko, Yannick; Gorner, Michael; Zhang, Jianwei (2018): Comparison of Multimodal Heading and Pointing Gestures for Co-Located Mixed Reality Human-Robot Interaction. In : 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid: IEEE, pp. 1–9.

Kurauchi, Andrew; Feng, Wenxin; Morimoto, Carlos; Betke, Margrit (2015): HMAGIC: Head Movement and Gaze Input Cascaded Pointing. In : 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments - PETRA ’15: ACM Press, pp. 1–4.

Laddi, Amit; Bhardwaj, Vijay; Kapoor, Nishant; Pankaj, Dinesh; Kumar, Amod (2015): Unobtrusive head gesture based directional control system for patient mobility cart. In : Signal Processing, Computing and Control (ISPCC), 2015 International Conference on. 2015 International Conference on Signal Processing, Computing and Control (ISPCC), 2015/09: IEEE.

Li, Dingjun; Rau, P. L. Patrick; Li, Ye (2010): A Cross-cultural Study: Effect of Robot Appearance and Task. In *Int J of Soc Robotics* 2 (2), pp. 175–186. DOI: 10.1007/s12369-010-0056-9.

Lohman, David F. (1996): Spatial ability and g. In *Human abilities: Their nature and measurement* 97 (116), p. 1.

Luca, Alessandro de; Flacco, Fabrizio (2012): Integrated control for pHRI: Collision avoidance, detection, reaction and collaboration. In : BioRob Roma 2012. IEEE International Conference on Biomedical Robotics and Biomechatronics : June 24-27, 2012, Roma, Italy. 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob 2012). Rome, Italy, 6/24/2012 - 6/27/2012. [Piscataway, N.J.]: [IEEE], pp. 288–295.

Luck, Michael; D’Inverno, Mark; Munroe, Steve (2003): Autonomy: Variable and Generative. In Gerhard Weiss, Henry Hexmoor, Cristiano Castelfranchi, Rino Falcone (Eds.): Agent Autonomy, vol. 7. Boston, MA: Springer US (Multiagent Systems, Artificial Societies, and Simulated Organizations), pp. 11–28.

Majaranta, Päivi; Aoki, Hirotaka; Donegan, Mick; Hansen, Dan Witzner; Hansen, John Paulin; Hyrskykari, Aulikki; Räihä, Kari-Jouko (2012): Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies: IGI Global.

Markham, Julie A.; Greenough, William T. (2004): Experience-driven brain plasticity: beyond the synapse. In *Neuron glia biology* 1 (4), pp. 351–363. DOI: 10.1017/s1740925x05000219.

Michalos, George; Makris, Sotiris; Tsarouchi, Panagiota; Guasch, Toni; Kontovrakis, Dimitris; Chryssolouris, George (2015): Design Considerations for Safe Human-robot Collaborative Workplaces. In *Procedia CIRP* 37, pp. 248–253. DOI: 10.1016/j.procir.2015.08.014.

Milgram, Paul; Kishino, Fumio (1994): A taxonomy of mixed reality visual displays. In *IEICE TRANSACTIONS on Information and Systems* 77 (12), pp. 1321–1329.

Miraz, Mahdi H.; Ali, Maaruf; Excell, Peter S. (2021): Adaptive user interfaces and universal usability through plasticity of user interface design. In *Computer Science Review* 40, p. 100363. DOI: 10.1016/j.cosrev.2021.100363.

Modares, Hamidreza; Ranatunga, Isura; Lewis, Frank L.; Popa, Dan O. (2016): Optimized Assistive Human-Robot Interaction Using Reinforcement Learning. In *IEEE transactions on cybernetics* 46 (3), pp. 655–667. DOI: 10.1109/TCYB.2015.2412554.

Monsell, Stephen; Driver, Jon (2000): Control of cognitive processes. Attention and performance XVIII / edited by Stephen Monsell and Jon Driver. Cambridge, Mass., London: MIT Press (A Bradford book).

Mott, Terran; Williams, Thomas; Zhang, Hao; Reardon, Christopher (2021): You Have Time to Explore Over Here!: Augmented Reality for Enhanced Situation Awareness in Human-Robot Collaborative Exploration.

Mutlu, Bilge; Roy, Nicholas; Šabanović, Selma (2016): Cognitive Human–Robot Interaction. In Bruno Siciliano, Oussama Khatib (Eds.): Springer Handbook of Robotics. Cham: Springer International Publishing, pp. 1907–1934.

Nass, Clifford; Steuer, Jonathan; Tauber, Ellen R. (1994): Computers are social actors. In Catherine Plaisant (Ed.): Conference companion on Human factors in computing systems - CHI '94. Conference companion. Boston, Massachusetts, United States, 23/04/1994 - 27/04/1994. New York, New York, USA: ACM Press, p. 204.

Neggers, S. F.; Bekkering, H. (2001): Gaze anchoring to a pointing target is present during the entire pointing movement and is driven by a non-visual signal. In *Journal of neurophysiology* 86 (2), pp. 961–970. DOI: 10.1152/jn.2001.86.2.961.

Newell, A. F.; Gregor, P.; Morgan, M.; Pullin, G.; Macaulay, C. (2011): User-Sensitive Inclusive Design. In *Univ Access Inf Soc* 10 (3), pp. 235–243. DOI: 10.1007/s10209-010-0203-y.

Norris, Sigrid (2009): Analyzing multimodal interaction. A methodological framework. [Repr.]. New York, NY: Routledge. Available online at http://www.loc.gov/catdir/enhancements/fy0651/2003025307-d.html.

Onnasch, Linda; Roesler, Eileen (2021): A Taxonomy to Structure and Analyze Human–Robot Interaction. In *Int J of Soc Robotics* 13 (4), pp. 833–849. DOI: 10.1007/s12369-020-00666-5.

Oppermann, Reinhard (1994): Adaptively supported adaptability. In *International Journal of Human-Computer Studies* 40 (3), pp. 455–472. DOI: 10.1006/ijhc.1994.1021.

Oswal, Sushil K. (2019): Breaking the exclusionary boundary between user experience and access. In Julie Staggers, Daniel P. Richards, Tim Amidon, Ehren Pflugfelder (Eds.): Proceedings of the 37th ACM International Conference on the Design of Communication. SIGDOC '19: The 37th ACM International Conference on the Design of Communication. Portland Oregon, 04 10 2019 06 10 2019. New York, NY, USA: ACM, pp. 1–8.

Oviatt, Sharon; Cohen, Philip R. (2015): The Paradigm Shift to Multimodality in Contemporary Computer Interfaces. In *Synthesis Lectures on Human-Centered Informatics* 8 (3), pp. 1–243. DOI: 10.2200/S00636ED1V01Y201503HCI030.

Papageorgiou, Xanthi S.; Chalvatzaki, Georgia; Efthimiou, Eleni; Fotinea, Stavroula-Evita; Karavasili, Alexandra; Tzafestas, Costas S. et al. (2019): User Centered Design for an Intelligent Robotic Rollator. Proc. Workshop on Mobile Robot Assistants for the Elderly (MoRobAE) in 2019 IEEE Int'l Conference on Robotics and Automation (ICRA), Montreal, Canada, 2019.

Park, Kyeong-Beom; Choi, Sung Ho; Lee, Jae Yeol; Ghasemi, Yalda; Mohammed, Mustafa; Jeong, Heejin (2021): Hands-Free Human–Robot Interaction Using Multimodal Gestures and Deep Learning in Wearable Mixed Reality. In *IEEE Access* 9, pp. 55448–55464. DOI: 10.1109/ACCESS.2021.3071364.

Phillips, B.; Zhao, H. (1993): Predictors of assistive technology abandonment. In *Assistive Technology* 5 (1), pp. 36–45. DOI: 10.1080/10400435.1993.10132205.

Qian, Yuan Yuan; Teather, Robert J. (2017): The eyes don't have it. In Adalberto L. Simeone (Ed.): Proceedings of the 5th Symposium on Spatial User Interaction. the 5th Symposium. Brighton, United Kingdom, 10/16/2017 - 10/17/2017. New York, NY: ACM, pp. 91–98.

Qvarfordt, Pernilla (2017): Gaze-informed multimodal interaction. In Sharon Oviatt, Björn Schuller, Philip R. Cohen, Daniel Sonntag, Gerasimos Potamianos, Antonio Krüger (Eds.): The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1: ACM, pp. 365–402.

Rahnev, Dobromir; Maniscalco, Brian; Graves, Tashina; Huang, Elliott; Lange, Floris P. de; Lau, Hakwan (2011): Attention induces conservative subjective biases in visual perception. In *Nature neuroscience* 14 (12), pp. 1513–1515. DOI: 10.1038/nn.2948.

Riley, Jennifer M.; Strater, Laura D.; Chappell, Sheryl L.; Connors, Erik S.; Endsley, Mica R. (2010): Situation awareness in human-robot interaction: Challenges and user interface requirements. In *Human-Robot Interactions in Future Military Operations*, pp. 171–192.

Ruzaij, Mohammed Faeik; Neubert, Sebastian; Stoll, Norbert; Thurow, Kerstin (2017): Design and implementation of low-cost intelligent wheelchair controller for quadriplegias and paralysis patient. In : 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI). 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI). Herl'any, Slovakia. [Place of publication not identified]: IEEE, pp. 399–404.

Salem, Maha; Rohlfing, Katharina; Kopp, Stefan; Joublin, Frank (2011): A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction. In Matthias Scheutz (Ed.): IEEE RO-MAN, 2011. July 31 - Aug. 3 2011, Atlanta, Georgia, United States of America ; 20th IEEE International Symposium on Robot and Human Interactive Communication. 2011 RO-MAN: The 20th IEEE International Symposium on Robot and Human Interactive Communication. Atlanta, GA, USA, 7/31/2011 - 8/3/2011. IEEE International Symposium on Robot and Human Interactive Communication; IEEE RO-MAN. Piscataway, NJ: IEEE, pp. 247–252.

Sałkowska, Marta (2018): Carrying Out Research Among Persons with Disabilities and their Relatives – Selected Ethical Issues. In *ZOP* Special Issue, pp. 200–217. DOI: 10.4467/2543408XZOP.18.010.10066.

Santis, Agostino de; Siciliano, Bruno; Luca, Alessandro de; Bicchi, Antonio (2008): An atlas of physical human–robot interaction. In *Mechanism and Machine Theory* 43 (3), pp. 253–270. DOI: 10.1016/j.mechmachtheory.2007.03.003.

Schmidtler, Jonas; Knott, Verena; Hölzel, Christin; Bengler, Klaus (2015): Human Centered Assistance Applications for the working environment of the future. In *OER* 12 (3), pp. 83–95. DOI: 10.3233/OER-150226.

Scholtz, J. (2003): Theory and evaluation of human robot interactions. In Ralph H. Sprague (Ed.): Proceedings of the 36th annual Hawaii international conference on system sciences. 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the. Big Island, HI, USA, 1/9/2003 - 1/9/2003. IEEE Computer Society: IEEE, 10 pp.

Scott, Janet (2010): Snapshot of using eye gaze–a professional journey into the world of eye gaze. In *WilsonA. GowR.(Eds.), The eye have it*, pp. 7–13.

Seth, Abhishek; Vance, Judy M.; Oliver, James H. (2011): Virtual reality for assembly methods prototyping: a review. In *Virtual Reality* 15 (1), pp. 5–20. DOI: 10.1007/s10055-009-0153-y.

Sheridan, Thomas B. (1992): Telerobotics, automation, and human supervisory control.

Sibert, Linda E.; Jacob, Robert J. K. (2000): Evaluation of Eye Gaze Interaction. In : Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM (CHI '00), pp. 281–288.

Stöhr, Matthias; Schneider, Matthias; Henkel, Christian (2018): Adaptive Work Instructions for People with Disabilities in the Context of Human Robot Collaboration (2018 IEEE 16th International Conference on Industrial Informatics (INDIN)), pp. 301–308.

Sukan, Mengu; Elvezio, Carmine; Oda, Ohan; Feiner, Steven; Tversky, Barbara (2014): ParaFrustum. In Hrvoje Benko, Mira Dontcheva, Daniel Wigdor (Eds.): Proceedings of the 27th annual ACM symposium on User interface software and technology. UIST '14: The 27th Annual ACM Symposium on User Interface Software and Technology. Honolulu Hawaii USA, 05 10 2014 08 10 2014. New York, NY, USA: ACM, pp. 331–340.

Tamura, Hiroki; Manabe, Takao; Goto, Takafumi; Yamashita, Yuki; Tanno, Koichi (2010): A Study of the Electric Wheelchair Hands-Free Safety Control System Using the Surface-Electromygram of Facial Muscles. In Honghai Liu (Ed.): Intelligent robotics and applications. Third international conference, ICIRA 2010, Shanghai, China, November 10-12, 2010 : proceedings / edited by Honghai Liu … [et al.], vol. 6425. Berlin: Springer (LNCS sublibrary. SL 7, Artificial intelligence, 6424-6425. Lecture notes in artificial intelligence), pp. 97–104.

Tan, Jeffrey Too Chuan; Duan, Feng; Zhang, Ye; Watanabe, Kei; Kato, Ryu; Arai, Tamio (2009): Human-robot collaboration in cellular manufacturing: Design and development. In : 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. St Louis, Missouri, 10-15 October 2009. 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2009). St. Louis, MO, USA, 10/10/2009 - 10/15/2009. Piscataway, NJ: IEEE, pp. 29–34.

Trosterer, Sandra.; Dzaack, Jeronimo (Eds.) (2011): Challenges limits of gaze-including interaction. Proc. of the 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction.

Tweed, D.; Glenn, B.; Vilis, T. (1995): Eye-head coordination during large gaze shifts. In *Journal of neurophysiology* 73 (2), pp. 766–779. DOI: 10.1152/jn.1995.73.2.766.

Unhelkar, Vaibhav V.; Li, Shen; Shah, Julie A. (2020): Decision-Making for Bidirectional Communication in Sequential Human-Robot Collaborative Tasks. In Tony Belpaeme, James Young, Hatice Gunes, Laurel Riek (Eds.): Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. HRI '20: ACM/IEEE International Conference on Human-Robot Interaction. Cambridge United Kingdom, 23 03 2020 26 03 2020. New York, NY, USA: ACM, pp. 329–341.

Villani, Valeria; Pini, Fabio; Leali, Francesco; Secchi, Cristian (2018): Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. In *Mechatronics* 55, pp. 248–266. DOI: 10.1016/j.mechatronics.2018.02.009.

Vogel, Christian; Walter, Christoph; Elkmann, Norbert (2017): Safeguarding and Supporting Future Human-robot Cooperative Manufacturing Processes by a Projection- and Camera-based Technology. In *Procedia Manufacturing* 11, pp. 39–46. DOI: 10.1016/j.promfg.2017.07.127.

Walker, Michael; Hedayati, Hooman; Lee, Jennifer; Szafir, Daniel (2018): Communicating Robot Motion Intent with Augmented Reality. In Takayuki Kanda, Selma Ŝabanović, Guy Hoffman, Adriana Tapus (Eds.): HRI'18. Proceedings of the 2018 ACM/IEEE International Conference on Human Robot Interaction : March 5-8, 2018, Chicago, Il, USA. the 2018 ACM/IEEE International Conference. Chicago, IL, USA, 3/5/2018 - 3/8/2018. New York, New York: Association for Computing Machinery, pp. 316–324.

Wang, X.; Ong, S. K.; Nee, A. Y. C. (2016): A comprehensive survey of augmented reality assembly research. In *Adv. Manuf.* 4 (1), pp. 1–22. DOI: 10.1007/s40436-015-0131-4.

Ware, Colin; Mikaelian, Harutune H. (1987): An evaluation of an eye tracker as a device for computer input2. In John M. Carroll, Peter P. Tanner (Eds.): Human factors in computing systems and graphics interface. Joint conference : Papers. the SIGCHI/GI conference. Toronto, Ontario, Canada, 4/5/1987 - 4/9/1987. Association for Computing Machinery. Special Interest Group on Computer and Human Interaction; Canadian Information Processing Society. Canadian Man-Computer Communications Society; Human Factors Society. Amsterdam, Oxford: North-Holland (0713-5424), pp. 183–188.

Weßkamp, Vanessa; Seckelmann, Tatjana; Barthelmey, André; Kaiser, Marcus; Lemmerz, Kai; Glogowski, Paul et al. (2019): Development of a sociotechnical planning system for human-robot interaction in assembly systems focusing on small and medium-sized enterprises. In *Procedia CIRP* 81, pp. 1284–1289. DOI: 10.1016/j.procir.2019.04.014.

WHO (2015): WHO Global Disability Action Plan 2014-2021. Available online at https://www.who.int/publications/i/item/who-global-disability-action-plan-2014-2021.

Wilimzig, Claudia; Tsuchiya, Naotsugu; Fahle, Manfred; Einhäuser, Wolfgang; Koch, Christof (2008): Spatial attention increases performance but not subjective confidence in a discrimination task. In *Journal of Vision* 8 (5), 7.1-10. DOI: 10.1167/8.5.7.

Wilkinson, Christopher R.; Angeli, Antonella de (2014): Applying user centred and participatory design approaches to commercial product development. In *Design Studies* 35 (6), pp. 614–631. DOI: 10.1016/j.destud.2014.06.001.

Williams, Tom; Hirshfield, Leanne; Tran, Nhan; Grant, Trevor; Woodward, Nicholas (2020): Using Augmented Reality to Better Study Human-Robot Interaction. In Jessie Y. C. Chen, Gino Fragomeni (Eds.): Virtual, Augmented and Mixed Reality. Design and Interaction, vol. 12190. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 643–654.

Williams, Tom; Szafir, Daniel; Chakraborti, Tathagata (2019): The Reality-Virtuality Interaction Cube: A Framework for Conceptualizing Mixed-Reality Interaction Design Elements for HRI. In : 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Daegu, Korea (South): IEEE, pp. 520–521.

World Health Organization (2011): World Report on Disability. Available online at https://www.who.int/disabilities/world_report/2011/report.pdf.

Wyart, Valentin; Tallon-Baudry, Catherine (2009): How ongoing fluctuations in human visual cortex predict perceptual awareness: baseline shift versus decision bias. In *J. Neurosci.* 29 (27), pp. 8715–8725. DOI: 10.1523/JNEUROSCI.0962-09.2009.

Yanco, H. A.; Drury, J. (2004): Classifying human-robot interaction: an updated taxonomy. In Peter Wieringa (Ed.): 2004 IEEE international conference on systems, man & cybernetics theme. Impacts of emerging cybernetics and human-machine systems : conference proceedings. 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583). The Hague, Netherlands, 10-13 Oct. 2004. Piscataway (N.J.): IEEE, pp. 2841–2846.

Yanco, Holly; Drury, Jill (2002): A taxonomy for human-robot interaction. In *Proceedings of the AAAI Fall Symposium on Human-Robot Interaction*, pp. 111–119.

Young, J.; Sharlin, E.; Igarashi, T. (2011): What Is Mixed Reality, Anyway? Considering the Boundaries of Mixed Reality in the Context of Robots. In Xiangyu Wang (Ed.): Mixed Reality and Human-Robot Interaction. Dordrecht: Springer Netherlands, pp. 1–11.

Zizlsperger, Leopold; Sauvigny, Thomas; Haarmeier, Thomas (2012): Selective attention increases choice certainty in human decision making. In *PloS one* 7 (7), e41136. DOI: 10.1371/journal.pone.0041136.

7. APPENDIX

7.1 Paper I. Understanding Human-Robot Collaboration for People with Mobility Impairments at the Workplace, a Thematic Analysis.

The following pages contain the submitted version to the 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). The final published version can be found at https://ieeexplore.ieee.org/abstract/document/9223489.

Arévalo Arboleda, Stephanie; Pascher, Max; Lakhnati, Younes; and Gerken, Jens; 2020. Understanding Human-Robot Collaboration for People with Mobility Impairments at the Workplace, a Thematic Analysis. In 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE, 561–566, 2020. DOI=10.1109/RO-MAN47096.2020.9223489.

Understanding Human-Robot Collaboration for People with Mobility Impairments at the Workplace, a Thematic Analysis

Stephanie Arévalo Arboleda[1], Max Pascher[1], Younes Lakhnati[1], Jens Gerken[1]

***Abstract*— Assistive technologies such as human-robot collaboration, have the potential to ease the life of people with physical mobility impairments in social and economic activities. Currently, this group of people has lower rates of economic participation, due to the lack of adequate environments adapted to their capabilities. We take a closer look at the needs and preferences of people with physical mobility impairments in a human-robot cooperative environment at the workplace. Specifically, we aim to design how to control a robotic arm in manufacturing tasks for people with physical mobility impairments. We present a case study of a sheltered-workshop as a prototype for an institution that employs people with disabilities in manufacturing jobs. Here, we collected data of potential end-users with physical mobility impairments, social workers, and supervisors using a participatory design technique (Future-Workshop). These stakeholders were divided into two groups, primary (end-users) and secondary users (social workers, supervisors), which were run across two separate sessions. The gathered information was analyzed using thematic analysis to reveal underlying themes across stakeholders. We identified concepts that highlight underlying concerns related to the robot fitting in the social and organizational structure, human-robot synergy, and human-robot problem management. In this paper, we present our findings and discuss the implications of each theme when shaping an inclusive human-robot cooperative workstation for people with physical mobility impairments.**

I. INTRODUCTION

According to the World Report on Disability, collected data from 69 countries showed that 20% of people live with severe or extreme difficulties in mobility [1]. People with physical mobility impairments (PWPMI) have limitations in their activities of daily living and work life, causing lower economic participation. However, if adequate environments are settled with the help of assistive technologies, PWPMI could be one step closer to participate in economic activities. Different employers' perspectives about hiring people with disabilities across industries have been debated in terms of acceptance, hiring probability and co-workers' attitudes [2]. An interesting fact presented by the U. S. Bureau of Statistics shows that from the employed population with disabilities in that country, almost the same percentage of people with and without disabilities would be likely to work in manufacturing and leisure/hospitality industries [3]. Manufacturing industries could provide opportunities for PWPMI, wherein human-robot cooperative tasks are already becoming a common synergy [4], [5], [6]. For instance, PWPMI could perform manipulation and grasping tasks by controlling robotic arms that allow interaction in a safe shared space, e.g., Kuka LBR iiwa [7], Kinova [8], Universal Robots [9].

As a use case, we considered a sheltered-workshop. Sheltered-workshops are institutions that provide a safe environment for people with disabilities to develop work-related skills and offer employment opportunities[10]. Our particular choice employs people with different types of disabilities and in variety of manufacturing tasks. Here, we carried out a set of observations at different already existing workstations and had special interest in manufacturing (assembly line) workstations. Therein, we took a closer look at employees with upper and/or lower limb disability to understand their needs, expectations, and preferences. In order to achieve this, we chose a participatory design approach through a technique known as Future-Workshop (FW). Our goal was to collect data about potential hands-free modalities and human intervention techniques, but we later discovered that there were underlying concerns that the FW brought to light and will be discussed Section 4.

When analyzing collected data, qualitative methods have been suggested to be appropriate when doing research with vulnerable population [11]. Thematic analysis is a common approach in health journals [12], and its versatility allows researchers to analyze data collected from different resources, including focus groups [13]. That is why we opted for thematic analysis to gain a better understanding of PWPMI in a human-robot cooperative workstation.

The purpose of this paper is to present the views of PWPMI to design a human-robot cooperative workstation. We organized the information as follows. Section 2 presents an overview of Human-Robot Collaboration (HRC) in manufacturing and for people with disabilities. Section 3 presents our method, participants, and results. These include a design-workshop (FW), the steps followed for our thematic analysis, a description of participants, and our outcomes. We further interpret and discuss our results in Section 4 to finally present our limitations and conclusions in Sections 5 and 6.

II. RELATED WORK

A. Human-Robot Collaboration in manufacturing and for people with disabilities

The goal of HRC within manufacturing is to create an environment which counters the inherent limitations of both humans and robots through working collectively [14]. The main difference between industrial and collaborative robots

*This work is funded by the German Federal Ministry of Education and Research (BMBF)

[1]The authors are with the Human-Computer Interaction Group, Westphalian University of Applied Sciences, Gelsenkirchen, Germany {stephanie.arevalo, max.pascher, younes.lakhnati, jens.gerken} @w-hs.de

regards the possibility of working in close contact and coordination with humans. Here, safety has been found to be the most important enabling factor [14], while trust determines the achievement of successful human-robot cooperation [15]. Kolbeinsson et al. highlight that cooperation and collaboration are often used interchangeably in HRC; and describe cooperation as sequential actions towards a shared goal, conversely to collaboration, described as shared sequential actions towards a shared goal [16].

There have been few research efforts within HRC that consider people with disabilities in manufacturing environments. An example is presented by Stöhr et al. who introduce a work HRC model for users with disabilities considering a pool of robot capabilities and the user's types of disabilities [17]. Kyrarini et al. investigate possible human-robot cooperative scenarios that include people with disabilities performing manipulation tasks [18]. Both works show the potential of HRC for PWPMI in manufacturing.

A number of published studies have focused on technology testing with people with disabilities, however, there have been fewer studies that focus on capturing the needs of this group of people. There is only a handful body of research that consider co-designing with people with disabilities (i.e. [19], [20]). Also, human factors need to be considered when implementing industrial HRC such as, operator participation in the implementation, visible senior management and commitment, and empowerment of the workforce [21]. The following research addresses this issue as it approaches HRC from a participatory design perspective and considers the social arrangement of primary users (PWPMI) and secondary users (caregivers, social workers, supervisors) to design a human-robot cooperative workstation.

III. METHOD

A. Procedure

1) Future-Workshop: This technique focuses on understanding particular situations, helps to identify common problems, and aims at developing a vision of the future [22]. Having participants in a group has the advantage of participants interacting with each other, i.e. challenging or building upon other's ideas; conversely to conducting individual interviews where information from one single participant is collected at a time [23].

We carried out a FW at the sheltered-workshop with two distinct groups of stakeholders, PWPMI, as primary users and social workers and supervising personnel, as secondary users. As a possible scenario, we presented an already existing manufacturing workstation of the sheltered-workshop. At this workstation, different metal pieces are currently assembled by hand. As part of the participatory design approach taken, there were previous sessions where we introduced the technology that could potentially be used in the workstation to gather experiences and impressions. Therefore, participants were already familiar with the broad research theme as well as the researchers from an ongoing research cooperation. The FW was held on a single day but divided into two sessions, one for each group of stakeholders.

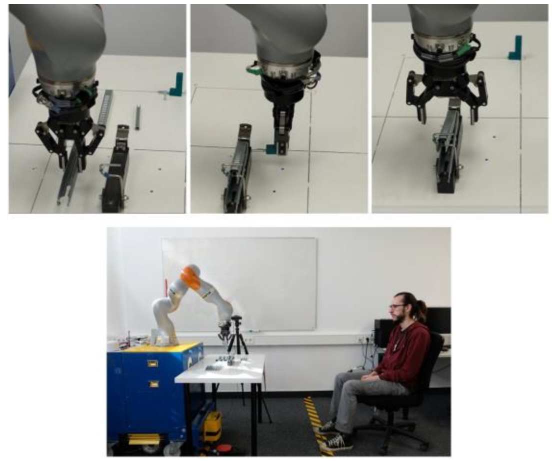


Fig. 1. Top: Screenshots from the video shown during the preparation phase. Bottom: Representation of a human-robot cooperative workstation.

Two researchers conducted the sessions, one taking notes of ideas shared and the other was encouraging ideation, discussion, and feedback through a set of open questions.

A FW is divided into different phases: preparation, critique, fantasy, and implementation [24]. During the preparation phase, we explained our goals for the workshop, handed a written consent form, explained some terminology that would be used, and created an inviting atmosphere that could promote sharing ideas. As a baseline, we presented two videos of the manufacturing scenario. On the first video the metal pieces were assembled by hand and on the second video they were built by the robot (Fig.1). The assembled piece is used in adjustable beds and it is one of the products that is already being produced in the manufacturing workstation. The second video showed how the robot could move (speed, distance) and the way in which it could manipulate objects. It also represented a human-robot cooperative workstation and therefore allowed participants to ponder about factors such as the distance between the robot and the human, as seen in Fig. 1. We then moved on to the critique phase were we mainly encouraged participants to think about problems or concerns through questions like: "What problems do you see in performing this task with a robot?" or "What worries you about working together with the robot?"

For the fantasy phase, we divided it into two general topics: hands-free interaction and human intervention techniques. Our goal was to discover preferences in different hands-free modes of interaction, e.g. eye-gaze, head-gaze, or speech to control the robot. For this topic, we asked a set of questions together with first-person-perspective videos to present a task wherein an industrial robot executes a set of actions, e.g., moving, turning, picking and placing objects, which are common in manufacturing and assembly tasks at the sheltered-workshop. Some of the questions related to modalities were: "How do you communicate with the robot?" "How can you control the robot and command certain movements?" "As an example: controlling the robot could be like controlling your wheelchair; if you want to go forward you would tilt your joystick as long as you reach the position where you want to be." The second topic, human intervention, infers that a robot operator steps in during an autonomous robot action to stop or modify such action. To represent intervention techniques, we showed a pick- and-place task, wherein the robot fails to place an object inside a box due to an unexpected movement of that box. Then, we asked questions such as: "How can we make the robot stop?" "When would you consider necessary to make a stop and when not?" "The system can also warn you about things that are happening around you. How would you like to get this alert?" Finally, in the implementation phase, we summarized the ideas of the group and shared our thoughts on the feasibility of the shared ideas.

TABLE I
LIST OF PRIMARY USERS

Participant	Diagnosis	Functional Ability
EU1	Duchene Muscular Dystrophy	Some head mobility, limited hand and finger mobility, help needed to position arms, no functional legs mobility
EU2	Duchene Muscular Dystrophy, learning disabilities, restrictive lung disease	Limited head mobility, limited finger mobility, help needed to position arms, ventilation, no functional legs mobility
EU3	Achondroplasia, learning disabilities, incapacity to walk, breathing difficulties	Some head mobility, some hand and arm mobility, no functional legs mobility

2) *Participants:* We organized two separate sessions (Fig. 2), both with the same general structure and topics. On the first session, we had 6 participants (P1 social pedagogist, P2, P3, P5, P6 social workers, P4 mechanical engineer) among which P1, P4, P6 were supervisors. The second session was held with our potential end-users (PWPMI), see Table 1. Our sessions lasted 90 and 75 minutes with the primary and secondary users respectively. The time invested in these sessions was considered part of their working hours.

B. Thematic Analysis

The data collected from the FW was analyzed using thematic analysis. This method was chosen because of its flexibility in finding patterns that are not necessarily theoretically bounded [13]. Our main goal was to recognize overarching conceptions of a human-robot cooperative workstation which might go beyond the specific line of questions of the FW.

First, we organized the data collected (detailed notes, summaries) and sorted them into the phases and topics from the FW. However, only the ideas from the critique and fantasy phase were analyzed, since the preparation and implementation phases were of explanatory nature without participants engaging in discussions. Then, the data was translated into English to be analyzed by two different researchers, a facilitator from the FW and a second researcher

Fig. 2. Top: Workshop with secondary users. Bottom: Workshop with primary users .

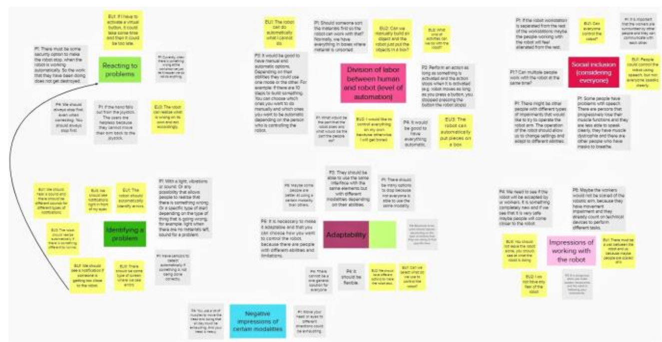


Fig. 3. Thematic map from coder 1: different shades of green and pink indicate themes that are related. Primary users' comments are shown in yellow and secondary users' comments in gray.

who had not been involved in this particular research. The reason behind that was to have a neutral perspective that could strengthen the reliability of the results. To provide context, the second researcher was briefed on the research goals and context of the workshop. Both researchers followed the same procedure but performed it separately.

We considered Braun & Clarke's six phases of thematic analysis and carried it out with an inductive approach [13]. First, the available textual information was read several times to gain familiarity. Second, comments describing emotions or impressions were identified to generate initial codes. Third, similar comments were grouped and descriptive keywords assigned. In order to have a visual representation of the information, we developed thematic maps. One of the thematic maps is shown in Fig. 3. Some comments were found to be related to multiple themes; however, they were placed on the best fitting theme. Fourth, themes were revised and evaluated to understand how they can fit together and were presented in a developed visual thematic map. Fifth, both coders analyzed their maps and together developed a final thematic map that encompasses all the findings. Similar and complementing themes were found by both coders, revealing a certain pattern

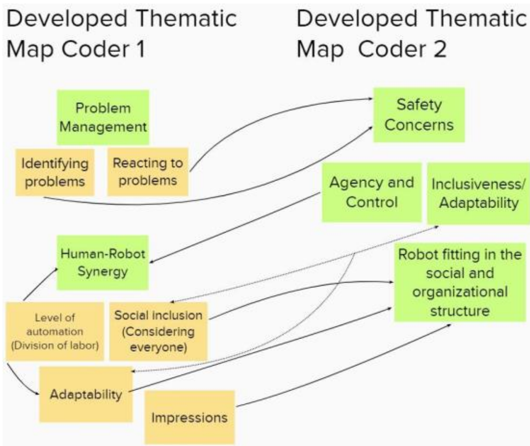


Fig. 4. Developed thematic maps of both coders. Overarching themes are presented in green and sub-themes are shown in orange.

in the data. The final thematic map is presented in Fig. 4 with respective sub-themes and connections between them. Both researchers looked for connections between the maps and produced one that encompasses the resulting themes and sub-themes that will be later used in this paper. Once the thematic analysis was performed the resulting thematic map was revised by a third researcher, who was the second facilitator at the FW. The third researcher evaluated if the comments matched the proposed themes, and looked for misconceptions on the researchers' side, e.g., check if comments were accidentally taken out of context. Sixth, considering these comments a summary of results was elaborated.

C. Results

The consolidated map, produced by both coders, identified patterns suggesting themes we labeled as: **the robot fitting in the social and organizational structure**, **human-robot synergy**, and **human-robot problem management**. Along with these themes, we identified sub-themes that contribute to each overarching theme and are described as follows.

1) The robot fitting in the social and organizational structure: This theme comprises the views of different stakeholders about the way that a robot can affect their pre-existing social and organizational structure. Both groups of stakeholders highlighted concerns with respect to the robotic arm disrupting social interactions. This was explicitly mentioned by P1, "It is important that the workers are surrounded by other people and they can communicate with each other. If the robot workstation is separated from the rest of the workstations maybe the people working with the robot will feel alienated from the rest." Adding to that, primary users were mostly worried about **being included**, e.g. EU1, "Can everyone control the robot?" Later the same person mentioned that "People could control the robot using speech, but not everyone speaks clearly." P4 mentioned the social acceptability of the robot, "We need to see if the robot will be accepted by our workers." Related with the concerns of being included, they mentioned the need for **Adaptability** on a social and organizational level. Secondary users mentioned that the robot should not create a sense of division between the workers. P1 elaborated "There are people with different types of impairments that would like to try to operate the robot arm. The operation of the robot should allow us to change settings and adapt to different abilities." Besides, we could identify that primary users had **contradicting impressions of working with a robotic arm**. Some expressed reservations about working with it. EU1 pointed out, "There must be a wall between the robot and us, because maybe people are scared of it." Conversely, EU2 said, "I do not have any fear of the robot." Additionally, P6 mentioned, "Maybe the workers would not be scared of the robotic arm because they have movement impairment, and they already count on technical devices to perform different tasks."

2) Human-robot synergy: This theme integrates the stakeholders' concerns and impressions about how PWPMI can work with a robot. The first evidence regards the **division of labor**. P1 asked "What would be the part that the robot does and what would be the part the people do?" EU2 also asked "What kind of activities can we do with the robot?" to which EU1 added, "The robot can do automatically what I cannot do." This brings the next concern regarding the **level of automation** that primary users would have in manufacturing tasks. EU2 asked, "Can we manually build an object and the robot just puts the objects in a box?" and P3 mentioned, "It would be good to have manual and automatic options. Depending on their abilities they could use one mode or the other." In this context, the need for **adaptability** was also mentioned, P3 said, "They should be able to use the same interface with the same elements but with different modalities depending on their abilities." P4 also expressed the need to have a flexible interface, "Elements to be used should appear depending on the type of actions that they are doing at that specific time." Additionally, primary users mentioned a sense of agency when controlling the robotic arm. EU3 said "You should not leave the robot alone, you should look at what the robot is doing," followed by EU2, "I would like to control everything on my own because otherwise, I will get bored." Conversely, a secondary user (P4) stated, "It would be good to have everything automatic."

3) Human-robot problem management: Safety is a prominent theme when working with PWPMI, who might not be able to react and protect themselves when there is a safety hazard. Therefore, secondary users wondered how unexpected situations may be handled. P1 mentioned, "If many people are working with the robot at the same time maybe they can make mistakes and accidentally harm people who are around." Here, two aspects were determined, identifying a problem and reacting to it. For **identifying problems**, secondary users expected that **the robot is responsible** for this task, to what P1 said, "The robot should have sensors to detect automatically if something is not being done correctly." This was also expressed by the primary users, EU1, "The robot should automatically identify errors." And EU3 added, "The robot should realize automatically if there is something different to normal." Similarly, primary and secondary users also expected the robot to be in charge of **reacting to problems**. EU3 demanded, “The robot can realize what is wrong on its own and act accordingly.” The reasoning behind it can be found in the limitations that PWPMI have, to what EU1 said, “If I have to activate a virtual button, it could take some time, and then it could be too late.” P1 also mentioned problems that some PWPMI with finger mobility have when controlling their wheelchair with a joystick, “If the hand falls out from the joystick, our workers are helpless because they cannot move their arm back to the joystick.” Adaptability was also mentioned when the users needed to stop the robot because of a failure, P1 mentioned, “There should be many options to stop because not everyone is able to use the same modality.”

IV. DISCUSSION

The first identified theme was labeled as the robot fitting in the social and organizational structure. Adding a robot to the regular workflow will certainly alter social behaviors. Primary and secondary users coincided in the viewpoint about a space where people with different abilities can be included. They raised concerns regarding the robot disrupting the sense of community that exists at the sheltered-workshop. Here, working close to each other and in an open area with different workstations was emphasized. Hence, the robot should adapt to the current space without drastically altering the social interactions that take place. Further, these concerns led us to consider an industrial robot as a social factor, since it can enable the inclusion of people with different conditions to take part in a certain type of work (manufacturing) that has been often excluded for PWPMI.

Adaptability in an industrial human-robot cooperative environment infers that a robot can adapt to perform different tasks. However, the analysis showed that for PWPMI adaptability goes beyond task execution and relates to being part of the community. Within the same line of thinking, industrial robots could potentially act as social facilitators for PWPMI, a topic that to our knowledge has been scarcely researched. In fact, social facilitation has been primarily analyzed with social robots [25], and has had promising results with children with autism [26].

In human-robot industrial environments, adaptability is closely tied to acceptability, since it could increase the willingness of people to cooperate with robots [27] and this aligns with the comments about adaptability from both primary and secondary users. Finally, different reactions are invoked when working with robots – some primary users explicitly expressed their fear of the robotic arm, while others mentioned the lack of it. Secondary users mentioned the familiarity of PWPMI with technological devices as a factor that could reduce the fear of the robotic arm. Thus, paying close attention to adaptability could minimize opposed impressions of a human-robot cooperative workstation.

The second theme concerns human-robot synergy. Primary and secondary users were not only interested in the division of labor between people and the robot, but they at the same time wanted to determine how much the robot can autonomously achieve. Additionally, there were controversial views about levels of automation within each group. It is noteworthy that both groups were interested in automated options. Previously, we hypothesized that primary users would be eager to use a robotic arm as an extension of their arms, i.e., teleoperate it to perform manipulation tasks. However, the workstation in a manufacturing environment, the use of an industrial robotic arm, and the diversity in impairment conditions might have contributed to driving people away from this line of thinking. Primary users did not only consider the robotic arm as an external entity but as a tool used at work, deriving in the desire to increase efficiency and thereby also consider automation. David & Nielsen, consider that robots have capabilities that can be autonomous in a certain context and need human intervention in others [28]. Building upon this, we shall consider an approach that combines different levels of automation through supervisory control of robots (semi-autonomous robotic arms), as a potential workstation for PWPMI. This could also provide them with the possibility of observing and judging the tasks that the robot is performing at the workstation, creating a sense of agency.

The third finding relates to the human-robot problem management theme. Note that we only considered problematic situations deriving from interaction problems, which Steinbauer defines as problems coming from “uncertainties in the interaction with the environment, other agents and humans” [29]. We left software and hardware problems aside, since this topic was not mentioned in our FW, and we believe that our primary users should not be responsible for solving these type of problems.

In our findings primary and secondary users agreed that they would prefer and even expect the robot to be the one that detects problematic situations and reacts accordingly. This is also related to the robot being considered as a tool and expected to work as such with a fault-tolerant design. However, this would require a high level of automation and disregard the advantages of human cognition and HRC. A possible reason behind automation preferences is that some participants have cognitive disabilities added to their physical impairments, which affects greatly their ability to handle problems. In fact, Honig & Oron-Gilad mention that interacting with a robot when there has been a failure, is an information processing task, and cognitive factors influence the ability to perceive and act to problems [30]. Another rationale points to the fact that PWPMI cannot physically react when a problem is found, added to the fact that some of them need and are used to the caregivers’ assistance. We believe that this phenomenon could be particularly different in a group of people without disabilities, who might want to investigate and solve problems either themselves or together with the robot. However, the motivation to solve human-robot problems is an aspect that to our knowledge has not yet been subject of study in HRC.

In general, viewpoints from primary and secondary users coincided in most aspects within the three presented themes. However, there were a a few opposing postures within the primary users group and almost none among secondary users. A possible explanation is the heterogeneity of PWPMI, who despite having a similar type of condition have other factors that influence greatly their needs and attitudes, e.g., the amount of time living with a certain condition, degrees of severity, and other disabilities (mental or cognitive).

V. Limitations

We acknowledge that due to the size of our sample our results mainly serve as basis for exploration. Our results provide an underrepresented voice to research and shows attitudes and reactions in an HRC workplace context, which is a topic that has not received much attention.

VI. Conclusions

The analysis performed unveiled three underlying concerns from primary and secondary users of a human-robot cooperative workstation. First, **the robot fitting in the social and organizational structure** should be carefully evaluated, especially in places that employ people with disabilities. Second, allowing for adaptability can be the key for successful **human-robot synergy**. As such, it is an essential part that allows for inclusion in a human-robot cooperative workstation. Third, viewpoints about **human-robot problem management** lean towards relying on the robot to detect or assist human operators to detect and solve human-robot problems deriving from interaction.

In this paper, we identified themes that can be used to influence design decisions for HRC at the workplace for PWPMI. This could serve to motivate further research in the area of designing human-robot cooperative workspaces for people with disabilities.

References

[1] WHO, "World report on disability," 2011. [Online]. Available: https://www.who.int/disabilities/world_report/2011/report/en/

[2] A. Houtenville and V. Kalargyrou, "Employers' perspectives about employing people with disabilities," *Cornell Hospitality Quarterly*, vol. 56, no. 2, pp. 168–179, 2015.

[3] U.S. BUREAU OF LABOR STATISTICS, "Employment of workers with a disability in 2017," 2017. [Online]. Available: https://www.bls.gov/opub/ted/2018/employment-of-workers-with-a-disability-in-2017.htm

[4] T. B. Sheridan, "Humanrobot interaction: Status and challenges," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 58, no. 4, pp. 525–532, 2016.

[5] I. Maurtua, A. Ibarguren, J. Kildal, L. Susperregi, and B. Sierra, "Human–robot collaboration in industrial applications," *International Journal of Advanced Robotic Systems*, vol. 14, no. 4, p. 172988141771601, 2017.

[6] J. T. C. Tan, F. Duan, Y. Zhang, K. Watanabe, R. Kato, and T. Arai, "Human-robot collaboration in cellular manufacturing: Design and development," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Piscataway, NJ: IEEE, 2009, pp. 29–34.

[7] KUKA, "Lbr iiwa — kuka ag," 21/04/2020. [Online]. Available: https://www.kuka.com/en-de/products/robot-systems/industrial-robots/lbr-iiwa

[8] Kinova, "Professional robots for industry," 23/04/2020. [Online]. Available: https://www.kinovarobotics.com/en/robots-industry-solutions

[9] Universal Robots, "Collaborative robotic automation — cobots from universal robots," 21/04/2020. [Online]. Available: https://www.universal-robots.com/

[10] A. J. Armstrong, "Sheltered employment," in *Encyclopedia of clinical neuropsychology*, J. S. Kreutzer, B. Caplan, and J. DeLuca, Eds. New York and London: Springer, 2011, p. 2286.

[11] S. Hartley and M. Muhit, "Using qualitative research methods for disability research in majority world countries," *Asia pacific disability rehabilitation journal*, vol. 14, no. 2, pp. 103–114, 2003.

[12] J. Green and N. Thorogood, *Qualitative methods for health research*, ser. Introducing qualitative methods. London: SAGE, 2004.

[13] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[14] G. Michalos, S. Makris, P. Tsarouchi, T. Guasch, D. Kontovrakis, and G. Chryssolouris, "Design considerations for safe human-robot collaborative workplaces," *Procedia CIRP*, vol. 37, pp. 248–253, 2015.

[15] G. Charalambous, S. Fletcher, and P. Webb, "The development of a scale to evaluate trust in industrial human-robot collaboration," *International Journal of Social Robotics*, vol. 8, no. 2, pp. 193–209, 2016.

[16] A. Kolbeinsson, E. Lagerstedt, and J. Lindblom, "Foundation for a classification of collaboration levels for human-robot cooperation in manufacturing," *Production & Manufacturing Research*, vol. 7, no. 1, pp. 448–471, 2019.

[17] M. Stohr, M. Schneider, and C. Henkel, "Adaptive work instructions for people with disabilities in the context of human robot collaboration," in *2018 IEEE 16th International Conference on Industrial Informatics (INDIN)*. [S.l.]: IEEE, 7/18/2018 - 7/20/2018, pp. 301–308.

[18] M. Kyrarini, A. Leu, D. Ristić-Durrant, A. Gräser, A. Jackowski, M. Gebhard, J. Nelles, C. Bröhl, C. Brandl, A. Mertens, and C. M. Schlick, "Human-robot synergy for cooperative robots," *Facta Universitatis, Series: Automatic Control and Robotics*, vol. 15, no. 3, p. 187, 2016.

[19] M. Ferati, A. Babar, K. Carine, A. Hamidi, and C. Mörtberg, "Participatory design approach to internet of things: Co-designing a smart shower for and with people with disabilities," in *Universal access in human-computer interaction. Part II*, ser. LNCS sublibrary. SL 3, Information systems and applications, incl. Internet/Web, and HCI, M. Antona and C. Stephanidis, Eds. Cham: Springer, 2018, vol. 10908, pp. 246–261.

[20] P. Dorrington, C. Wilkinson, L. Tasker, and A. Walters, "User-centered design method for the design of assistive switch devices to improve user experience, accessibility, and independence," *Journal of Usability Studies*, vol. 11, no. 2, pp. 66–82, 2016.

[21] G. Charalambous, S. Fletcher, and P. Webb, "Identifying the key organisational human factors for introducing human-robot collaboration in industry: an exploratory study," *The International Journal of Advanced Manufacturing Technology*, vol. 81, no. 9-12, pp. 2143–2155, 2015.

[22] I. Skoglind-Ohman and H. Shahnavaz, "Assessment of future workshop's usefulness as an ergonomics tool," *International journal of occupational safety and ergonomics : JOSE*, vol. 10, no. 2, pp. 119–128, 2004.

[23] J. Kitzinger, "The methodology of focus groups: the importance of interaction between research participants," *Sociology of Health and Illness*, vol. 16, no. 1, pp. 103–121, 1994.

[24] R. Jungk and N. R. Müllert, *Future workshops: How to create desirable futures*. London: Institute for Social Inventions, 1996.

[25] N. Riether, F. Hegel, B. Wrede, and G. Horstmann, "Social facilitation with social robots?" in *Proceedings of the seventh annual ACMIEEE international conference on Human-Robot Interaction*, ser. ACM Digital Library, H. Yanco, Ed. New York, NY: ACM, 2012, p. 41.

[26] I. Iacono, H. Lehmann, P. Marti, B. Robins, and K. Dautenhahn, "Robots as social mediators for children with autism - a preliminary analysis comparing two different robotic platforms," in *2011 IEEE International Conference on Development and Learning (ICDL 2011)*. Piscataway NJ: IEEE, 2011, pp. 1–6.

[27] C. Strassmair, N. Taylor, and R. Aylett, "Human robot collaboration in production environments," 2014.

[28] R. David and P. Nielsen, "Defense science board summer study on autonomy," 2016.

[29] G. Steinbauer, "A survey about faults of robots used in robocup," in *Robocup 2012*, ser. LNCS sublibrary: SL 7 - Artificial intelligence, X. Chen, Ed. Heidelberg: Springer, 2013, vol. 7500, pp. 344–355.

[30] 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.

7.2 Paper II. Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects.

Arévalo Arboleda Stephanie; Pascher, Max; Baumeister, Annalies; Klein, Barbara; and Gerken, Jens. 2021. Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects. In The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA 2021). Association for Computing Machinery, New York, NY, USA, 147–155. DOI: 10.1145/3453892.3458044.

Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects

Stephanie Arévalo Arboleda
stephanie.arevalo@w-hs.de
Westphalian University of Applied Sciences
Gelsenkirchen, NRW, Germany

Max Pascher
max.pascher@w-hs.de
Westphalian University of Applied Sciences
Gelsenkirchen, NRW, Germany

Annalies Baumeister
annalies.baumeister@fb4.fra-uas.de
Frankfurt University of Applied Sciences
Frankfurt am Main, Hessen, Germany

Barbara Klein
bklein@fb4.fra-uas.de
Frankfurt University of Applied Sciences
Frankfurt am Main, Hessen, Germany

Jens Gerken
jens.gerken@w-hs.de
Westphalian University of Applied Sciences
Gelsenkirchen, NRW, Germany

(a) (b) (c)

Figure 1: Different use cases of Human-Robot Collaboration. (a) Workplace (b) Home (c) StreetX

ABSTRACT

Human-robot technology has the potential to positively impact the lives of people with motor disabilities. However, current efforts have mostly been oriented towards technology (sensors, devices, modalities, interaction techniques), thus relegating the user and their valuable input to the wayside. In this paper, we aim to present a holistic perspective of the role of participatory design in Human-Robot Collaboration (HRC) for People with Motor Disabilities (PWMD). We have been involved in several multiyear projects related to HRC for PWMD, where we encountered different challenges related to planning and participation, preferences of stakeholders, using certain participatory design techniques, technology exposure, as well as ethical, legal, and social implications. These challenges helped us provide five lessons learned that could serve as a guideline to researchers when using participatory design with vulnerable groups. In particular, young researchers who are starting to explore HRC research for people with disabilities.

CCS CONCEPTS

• **Human-centered computing → Participatory design**; **Accessibility design and evaluation methods**.

KEYWORDS

human-robot collaboration, people with motor disabilities, participatory design, challenges and lessons learned

ACM Reference Format:
Stephanie Arévalo Arboleda, Max Pascher, Annalies Baumeister, Barbara Klein, and Jens Gerken. 2021. Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects. In *The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA 2021), June 29-July 2, 2021, Corfu, Greece.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3453892.3458044

PETRA 2021, June 29-July 2, 2021, Corfu, Greece

ACM ISBN 978-1-4503-8792-7/21/06…$15.00
https://doi.org/10.1145/3453892.3458044

1 INTRODUCTION

The WHO collected data from 69 countries and found that 15% of the world's population lives with some sort of disability [53]. Among these, 20% report a severe form of disability with a further 10.3% found to have an extreme form, and 9.7% have difficulties in mobility [52]. Motor disabilities affect upper and/or lower limbs due to injuries, diseases, or congenital problems [16]. These limit physical dexterity and different life aspects of People With Motor Disabilities' (PWMD).

Human-robot technologies have the potential to enhance the abilities of PWMD. PWMD frequently require the assistance of others to perform activities of daily living including reaching and manipulating objects around them. This can affect their sense of autonomy and self-efficacy [8]. PWMD emphasize the desire to live independently, i.e. being able to be on their own for several hours without the need of a care assistant [41]. We consider that robot designers can benefit greatly from the adoption of a Participatory Design (PD) approach in their practices to achieve a truly inclusive human-robot collaborative environment.

PD has been considered a tool for sharing expertise among stakeholders and designers, but most of all for inspiring change [30]. PD is used to better engage with all stakeholders and involve potential users as full partners in the design process, which leads to empowerment of stakeholders [28]. This provides an additional benefit as it leads to greater democratization between designers and stakeholders [34]. Additionally, it results in greater inclusion, as PWMD have been previously largely overlooked in the design process [51]. Further, this involvement has shown to positively influence their attitudes and behavior [26].

During the last three years, we have been involved in several multiyear projects related to Human-Robot Collaboration (HRC) for PWMD. The projects we include in this paper cover different use-cases: a home-support scenario for PWMD (eating and drinking with a robotic arm, HOME project, 2017-2020 [39], a workplace environment (working in collaboration with a robotic arm in a sheltered-workshop to perform manufacturing tasks, WORKPLACE project, 2017-2021) [40], and an urban mobility issue (crossing streets in a sensor enhanced wheelchair, STREETX project 2018-2019) [7], see Figure 1. This variety allowed us to explore and apply a multitude of methodological approaches and therefore achieve a better understanding of individual and overarching challenges. The outcomes of each project are not intended to be part of this paper, as they deserve their own analysis in separate publications.

Managing the different aspects of having a PD approach with PWMD can be a challenge for early-career researchers with limited knowledge of user-centered design. Research that has been carried out with PWMD has long been a matter of ethical concern in social sciences [45], [18]. Also, methodological challenges have been evaluated and compared to carrying out research with mainstream participants [37]. We build upon these works and align our experiences with those previously reported. Our paper takes a user-centered design perspective for HRC and highlights the particularities of using PD designing for PWMD.

Holone & Herstad [23] point out that "the ideals of Participatory Design must be approached with pragmatism." This leads us to present our experiences and thus offer a guide from the lessons we have learned. Our contribution is twofold. First, our related work section presents an approach to HRC from a PD perspective which emphasizes the role of PWMD in this context. Second, we present a short overview of our projects and discuss the challenges we experienced. Based on this, we provide a set of practical lessons learned when designing a human-robot collaborative environment for PWMD. Through this, we aim to generate a dialogue about the advantages of PD in HRC for PWMD. This paper is directed to people who are starting to explore research areas that use robots to assist people with disabilities, often coming from a more technical background and less experience in user-centered design.

2 RELATED WORK

People with disabilities have been of interest among the HRC research community. Previous work has focused on exploring the use of different modalities [14], interaction [33], and graphical user-interfaces [44]. However, designing technology that has the potential to increase self-sufficiency calls for special attention to the social aspect and the relevance of evaluating social situations when designing assistive technologies [46]. The upcoming section presents a review of PD and HRC in design for people with disabilities.

2.1 Human-Robot Collaboration and People with Motor Disabilities

HRC addresses research where humans and robots are considered partners in carrying out joint goals [6]. In recent years, HRC has become widespread in manufacturing lines [13]. However, HRC has also been of interest in assistive environments for the aging population [20] and disability care [29] to collectively overcome obstacles.

In a space where humans and robots coexist, there are a number of factors requiring consideration. For instance, the role of safety in the prevention of collisions by optimizing timing and intervention [36]. Situation awareness is a pertinent component of safety in HRC. It involves determining when it is appropriate to take an action, and what information needs to be communicated to the human counterpart [50]. HRC must consider the socio-technical arrangement of primary (end-users) as well as the one of secondary users (e.g. caregivers). These perspectives can differ from each other and need to be appropriately reflected in the designed technology. Here, a previous study showed that primary users stressed autonomy, privacy, and choice, whereas caregivers perceived safety as important as autonomy [27]. Additionally, work-load is another aspect that should be addressed. Tan et al. [49] found a correlation between mental workload with robot motion speed and distance, i.e., people experienced an increase in mental workload when the robot moved faster and in closer proximity.

2.2 Participatory Design in Human-Robot Interaction for People with Disabilities

Participatory research according to Balcazar et al. [4] has the power to influence the design and research agenda. It addresses problems that people with disabilities experience by providing them an active role in the decision making process. PD can therefore be described as a "verbal exchange of ideas" [32], which can be used when working together with PWMD. Focus groups and workshops are common techniques implemented within PD. These have also been used when designing for different types of disabilities in the context of accessible design, e.g., people with visual impairments [35], or people with lower-limb impairments to design a wheelchair [15]. Interviews are a technique that is commonly used, wherein semi-structured interviews can be effective to collect individual experiential knowledge [32] together with participatory observations. The latter is particularly useful when participants are not able to talk in detail about specific tasks that they perform with the help of their caregivers.

PD techniques have proven to show better results when combined. For instance, Azenkot et al. [3] used interviews, workshops, and individual Wizard-of-Oz session to collect design insights about a robot tasked to help people with visual impairments to navigate indoors. Ferati et al. [19] took a closer look at different PD techniques (cartographic mappings, future-workshop, cultural probes). These techniques were used to determine the main concerns of the target user group regarding their daily living situation and thus design solutions tailored to specific needs.

PD can also affect expectations in users. Bratteteig et al. [9] explore expectation management and idea selection to encourage reflection in pursuit of feasible and realistic outcomes, thus increasing end user acceptance. This is particularly relevant among PWMD. Many PWMD are used to customized designs, and designers should avoid discarding their ideas without providing the necessary argumentation. Further, Reich-Stiebert [43] showed how involving users at the earliest stages of design can positively influence their attitudes towards robots. It is important not to underestimate the designers' behavior and skills, since these are crucial to achieving a true interchange of ideas always respecting the users' knowledge [10].

In her Disability Interaction Manifesto, Holloway points out the need for a new body of knowledge acquired by co-creation with people with disabilities [22]. Thus, opening discussions in Human-Robot Interaction (HRI) about PWMD and encouraging research efforts towards extending people's capabilities.

2.3 Empowering Users through Participatory Design

PD has the ability to connect different stakeholders with designers with the common goal to share knowledge. Lee et al. [31] showed how carrying out PD improved the understanding of their target demographic, and showed a path to the creation of new knowledge in the development of better social robots. Another way of empowerment occurs when users recognize their ideas in the end solution. For instance, Bratteteig & Wagner [10] used nurses' ideas and translated them directly into a user interface.

The feeling of self-efficacy assists in the well-being of people with motor impairments [21] since it provides a feeling of achievement by one's efforts. Hence, involving people with disabilities in the design process not only boosts the feeling of ownership of a solution but also adds a feeling of self-efficacy and empowerment derived from their co-creator's role. Further, Barbareschi et al. [5] used an emotional design model to evaluate the design and production of a wheelchair. Their results showed that a user-centered approach, when producing a customized solution, is perceived as an "empowering and meaningful form of self-expression".

3 CONTEXT AND METHODS

This section will provide the foundation to understanding our research carried out over the last years. Further, it puts the challenges and lessons learned we present in § 4 and § 5 in perspective. In user-centered design approaches and PD in particular, involving all stakeholders during the design process is essential. Table 1 presents a summary of the participants with motor disabilities who partook in our projects.

An overview of the methods used is presented in Table 1. One essential approach we relied upon was contextual inquiry and subsequent interviews and participatory observations. This allowed us to observe participants in-situ and gain further context through interaction with participants. With regards to the evaluation methods, user-testing with the group of PWMD focused on qualitative analysis of systematic exposure to demonstrators, factoring in trust, well-being, and placing quantitative measures in second place. Our prototypes, see Figure 2, were evaluated with participants without disabilities prior to the evaluation with PWMD due to (1) the high individual variability of PWMD; (2) the potential of physical harm. High individual variability makes interpretation difficult. Thus, establishing a base-line with people without physical constraints can help to put results in perspective and it avoids physical harm when participants cannot physically move.

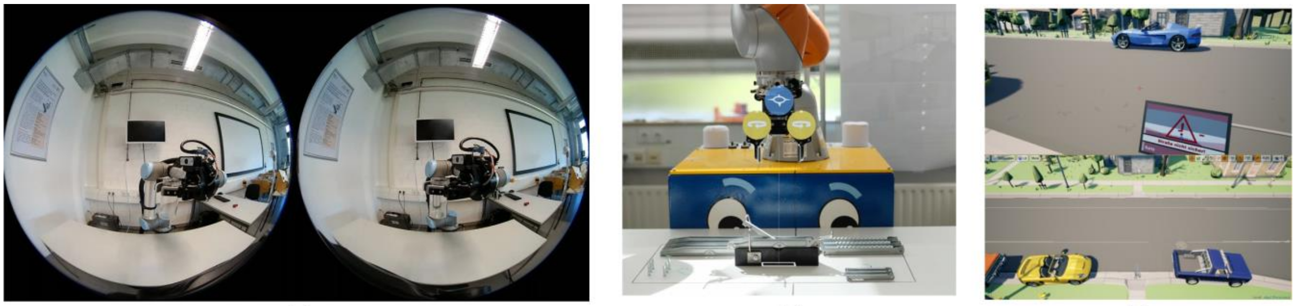
3.1 Home Project (HOME)

The goal of the HOME Project is to explore the opportunities and limitations of a robotic arm to assist PWMD with food consumption. To that end, we investigated the personal living and care spaces of our user group. Technology-wise, we focused on autonomous behaviors of a robotic arm, where PWMD can recognize the robot intentions and develop intervention strategies when/if needed. Our solutions aim to provide visual feedback about the robot intentions and its status to keep PWMD informed at all times. We highlighted safety and trust since the robotic arm would come into close proximity to the users' face.

3.1.1 Participants. In this project, we worked together with a hospital that put us in contact with 15 people with quadriplegia (and their caregivers). They were scattered across the country and were individually visited at home. Subsequently, they took part in a series of workshops that addressed different issues of ethical, legal, and social implications. In addition to this selected participants' group, we reached out to an anonymous, but larger user base through online surveys.

3.1.2 Method. We carried out interviews and participatory observations at the participants' homes along with prepared video-props to illustrate a potential HRC scenario. We documented audio and video materials together with pictures of the physical environment.

Based on the difficulty of getting access to larger populations of the end-user group, we also conducted surveys via Amazon Mechanical Turk [1] and similar platforms including SurveyCircle [48]. We used this approach to collect feedback on a variety of robotic setups regarding its position, speed, and proximity to the user. These



(a) (b) (c)

Figure 2: (a) HOME. View of the videos wearing Google Cardboard. (b) WORKPLACE. View of the workstation wearing a Microsoft HoloLens. (c) STREETX. VR simulation of street crossing

Table 1: Description of participants and methods per project

Project (Number of Participants)	Diagnosis	Technique	Materials
WORKPLACE (3)	Duchene Muscular Dystrophia, learning disabilities, restrictive lung disease, achondroplasia, incapacity to walk, breathing difficulties, multiple sclerosis	Participatory observations, future workshop, videos, technology testing, prototype testing	AR prototypes (Microsoft HoloLens)
HOME (15)	Locked-in Syndrome, inclusion body myositis , multiple sclerosis, physical trauma, arthrogryposis multiplex congenita	Participatory observations, surveys, interviews, videos	Google Cardboard, pictures, videos, prototype testing with VR simulations
STREETX (4)	Spastic quadriplegia, paralyzed limbs, spina bifida, quadriplegia	Interviews, participatory observations, prototype testing	VR simulations

aspects could not be verified quickly in a physical environment but could be easily represented through videos within the survey structure.

3.1.3 Materials. As mentioned above, we relied on videos documenting a first person perspective of the view on a HRC scenario. For our interviews at the participants' homes, we used stereoscopic videos shown through Google Cardboard. Additionally, we showed pictures of different robotic settings, which are currently on the market, e. g. iEat [11], Obi [38], JACO [25], and iArm [12]. In our surveys, we used monoscopic videos visualized through a standard display rather than VR or Cardboard headset. In addition to videos, we developed interactive 3D environments to simulate HRC scenarios and presented those to potential users.

3.2 Workplace Project (WORKPLACE)

The main goal of the WORKPLACE project is the development of an interaction concept that allows PWMD to execute manipulation and grasping tasks with a robotic arm in a work environment. We aim to simplify existing complex robot control. Especially for people without prior experience and who are still mostly able to move their heads, eyes, and are able to speak. Hence, we focused mostly on Augmented Reality based approaches to overlay the visual space with basic robot controls and use visual cues to ease understanding the relationship between the robot and the workspace.

3.2.1 Participants. Due to the nature of this project we chose to collaborate with a sheltered-workshop. Sheltered workshops employ people with a variety of physical and cognitive disabilities. Here, a "lead" user group of three PWMD was established with the goal to accompany the project. This lead user group was involved in various stages of the project including design workshops, technology demonstrations, and further interviews. The lead user group was selected based on their availability and interest to participate. In addition to that, we included secondary users such as caregiving personnel, engineers, social workers, social pedagogists, and workstations' supervisors.

3.2.2 Method. We used contextual inquiry to gain an overview of the tasks, behaviors, and social protocols at various workstations. Thus, evaluating the possibility of adding a collaborative robot to the current working environment.

In a flipped exploration/observation setting, we invited our lead users from the WORKPLACE project to visit our laboratory. They were able to see our robots and prototypes, and experience firsthand the technologies that would be used. For instance, during these visits they used devices such as the Microsoft HoloLens 1 (head-mounted augmented reality display), see Figure 3, or a customized sensor-enhanced headband that allows operating a robot through head movements.

Also, we applied a "future-workshop" as an envisioning method. Primary (PWMD) and secondary users (staff from the sheltered-workshop such as social workers and supervisors) ideated about problems and envisioned novel design approaches collaboratively with researchers [2].

3.2.3 Materials. In our future-workshop, we worked with first-person videos of different interaction scenarios to help participants envision such an environment. Additionally, we used the Microsoft HoloLens 1 to present augmented reality prototypes that enhance the workstation to provide relevant task information.

3.3 Crossing Streets (STREETX)

The aim of the STREETX project is the development of an assistive system that allows wheelchair users to safely cross streets. We analyzed their habits and concerns about independently crossing streets to improve road safety. In order to achieve this, we aimed to modify an electric wheelchair to simulate a robotic assistant that provides information about the current traffic. This is performed in a way similar to the intersection assistants used in cars.

3.3.1 Participants. To avoid common limitations of access to the target group, we explored a different and less formal approach. We identified a Reddit [42] community and analysed existing posts and discussions to extract frequently mentioned needs and problem areas for PWMD. Among other trends, we identified the problem of street crossing. Based on this, we conducted a user research similar to the one from the WORKPLACE project and cooperated with a local sheltered-workshop. The potential users that were interested in cooperating in the project were again contacted for prototype testing and feedback.

3.3.2 Method. For our Reddit analysis, we conducted text-based thematic content analysis to identify main issues hidden in the casual conversations and threads. Additional information was collected through interviews with PWMD from sheltered workshops. Based on this, we developed a prototype and carried out usability testing.

3.3.3 Materials. The collected information helped us to create 3D simulations that were displayed on a computer screen and in VR. This setup was chosen as a situation like street crossing would be potentially dangerous for participants in reality.

4 CHALLENGES

To the best of our knowledge, much of the challenges from previous literature leave aside the pragmatic use of PD in HRC for PWMD. Thereby, we grouped the challenges we found into different topics. Concerning participants, we found challenges in planning & participation and preferences of stakeholders. Regarding the methods used, we identified the advantages of using several PD techniques. With respect to the materials used, we realized the importance of technology exposure. Finally, we highlight that ethical, legal, and social implications go beyond consent.

Figure 3: Primary users testing an Augmented Reality prototype on the Microsoft HoloLens.

4.1 Planning & Participation

Vines et al. [51] have previously highlighted the benefits of having end-users as project partners, although, it might not always be possible. Hence, it is crucial to consider different means to approach a target group that is not easily accessible. That is why we contacted and recruited participants in manifold ways. We had a sheltered-workshop as a project partner (WORKPLACE & STREETX project), we analyzed a Reddit community for PWMD (STREETX project), and we were directed to PWMD through hospital medical staff and support groups (HOME project).

When working together with PWMD, planning takes in particular an important role. For example, when recruiting participants in a sheltered workshop for the WORKPLACE project, in despite of the willingness of users to participate, caregivers expressed their concerns about disrupting participants daily routine. Caregives expressed that this can add severe stress to some participants. Therefore, together with the social workers and caregivers, we decided to establish a lead user group approach, wherein we carefully selected participants that would be able to handle this involvement. Then, we carefully planned the research sessions to avoid unnecessary disturbances and focused on inclusion into regular routines of participants.

In the case of interviews and observations, participants might be located in geographically distant locations which hinders the access to them. For example, the HOME project had participants scattered across the country, making it difficult for the researchers to visit every participant. In addition, a caregiver presence was needed in case of emergency. During the future-workshop, caregivers mentioned the importance of choosing an adequate time of day and duration for the workshop. Appointments should not interfere with lunchtime and restroom schedule, as this could cause an extra burden to PWMD. Also, we needed to arrange transportation (a van that can transport wheelchairs), since not all participants were located at the same building where the workshop took place. Finally, we considered as well the fact that some PWMD might need assistance during the workshop, hence, at least one caregiver was in an adjacent room.

Further, to enable user testing with PWMD, we found it very demanding (size, connectivity, and also safety) to move an industrial robot to the location of PWMD. The alternative, have PWMD visit our premises, involves also a lot of planning and added resources, e.g. for secondary users.

4.2 Preferences of Stakeholders

Holloway highlights the importance of "creating a new body of knowledge with disabled people through the exchange of ideas around domain-specific knowledge" [22]. We capitalize on that and included not only researchers from different fields but worked with stakeholders from different domains as well, e.g., WORKPLACE, social workers, civil-engineers, social pedagogists. Yet, we sometimes found difficulties in handling opposing viewpoints of different stakeholders.

An important matter when involving different stakeholders comes in terms of the diversity of interests and concerns that they may have and how these would be considered in the design process. Laitano [30] already identified that finding consensus among stakeholders is a challenging task. Additionally, Spiel et al. [47] mentioned the importance of negotiating multiple agendas through mediation and withdrawing. The involvement of each stakeholder provides valuable input from different perspectives, some bound to their respective professional backgrounds. Even in organizations wherein it is assumed that everyone shares the same goal, nuances in perspectives can be identified. For instance, some supervisors at the sheltered workshop prioritize production, while social workers highlight well-being. This suggests that different perspectives are clearly visible and can be controversial, even among small groups of people that share a common organizational goal.

From our experiences, while PWMD encompasses people with similar diagnoses, there are huge differences on the resulting type of impairment and the degree of severity. Related variables such as age or the duration a person has lived with a disability also contribute to very different individual characteristics. These aspects add complexity when deriving knowledge. For instance, we involved participants with quadriplegia, multiple sclerosis, locked-in syndrome, duchene muscular dystrophia, and cerebral palsy. Some of these participants had other cognitive and learning disabilities in addition to motor impairments. This high degree of individual differences leads to variability when measuring quantitative data and makes interpreting the results difficult, e.g. task times with a certain input modality.

Qualitative assessments can also be biased due to these particular differences. For example, a participant might prefer the use of certain modality. This preference is related to a part of their body that is the least affected or one that they maintain control despite the progression of a disability. Hence, qualitative approaches allow deepening into the reasoning behind certain choices and opinions. Forming thus together a collective understanding of the population. Luck [32] calls for attention to the users' difficulties in articulating their preferences. These can be unrelated to the type of disability and suggest different understanding of certain terms used. Thereby, it is important to clarify the terminology and share a common understanding of the terms used.

4.3 Using Various PD Techniques

Using different PD techniques allows to better capture different views on HRC. However, every PD technique involves different types of challenges. For instance, interviews often require a one-to-one interaction. Such personal interactions were difficult in the case of PWMD. Caregivers need to be present in case of unexpected emergencies or to act as a spokesperson when physical communication abilities were limited. We observed that participants tend to be more open when their caregivers are not in direct proximity but located in a different (adjacent) room.

Idealizing participation versus having realistic expectations has been previously discussed in the literature [10]. In our experiences we realized that this is particularly the case with PWMD, some with added cognitive disabilities, e.g. they might lose focus easily, get distracted, or experience physical discomfort. Hence, participation and engagement of PWMD depend more prominently on their current physical and mental state, when compared to people without disabilities. For example, one participant who is typically active and talkative was suddenly very quiet during our future-workshop due to only recently recovering from a sickness.

Secondary users also perceive design sessions differently. While it might raise curiosity in some attendants it can be considered as a burden and loss of time to others. Sometimes, stakeholders have the feeling that they lack expertise in certain topics, e.g., during our design workshop some social workers felt unease when approaching topics related to technology.

Another methodological aspect concerns the researchers in the design team. In the HOME project, our team consisted of social scientists and computer engineers, which allowed interviews to both cover social and technological aspects. Also, using different PD Techniques and materials allows researchers to be prepared for unexpected situations. For instance, showing videos to participants was still possible to be carried out remotely. It further allowed us to continue our research during the SARS-CoV-2 pandemic and subsequent restrictions.

4.4 Technology Exposure

Holzinger et al. [24] identified a positive correlation between technology exposure and acceptance. Further, they mention the importance of further exploring ways to expose people to technology in the design process.

Designing HRI is not abstract and virtual like designing a new mobile app. It involves interacting with a robot and other sensors and actuators, as well as with other feedback devices, e.g. head-mounted displays. This results in a complex and rather large technical setup. This compromises the desired technology exposure in the real environment. Hence, we recurred to other manners of presenting prototypes, e.g., through videos and VR simulations (HOME, STREETX projects)

As previously shown [17], familiarization is crucial in HRI since it can lead to users feeling more comfortable and trusting during interaction with robots. In the WORKPLACE project, carrying an industrial robot to the sheltered workshop is not something that can be done regularly due to specific spatial and connectivity configurations. Therefore, we looked to gather first impressions of the robot by having end-users visit the research laboratory. All of the end-users had never seen an industrial robot, thus, it served us to collect impressions. It also enabled end-users to become familiar with the human-robot affordances. However, we encountered several difficulties regarding the devices used. For instance, putting on the Microsoft HoloLens when wearing an oxygen mask was not possible, certain eye-shapes did not allow for calibrating an eye-tracker or for proper visualization of the imagery displayed in a head-mounted display. Additionally, some participants' neck could not hold a head-mounted display and not everyone was able to execute head-movements (yaw, pitch, and roll) to the same extent.

4.5 Ethical, Legal and Social Implications (ELSI) in the research process

The consideration of ethical, legal and social implications is a funding requirement of European and other international publicly funded projects. These aspects affect the research process as well as the technology to be developed. Here we focus on the research process and the challenges observed. Doing research with people with disabilities usually requires an ethics board. In particular, in technology research, finding an ethics board might not be straight forward—most ethics boards we are aware of, focus on topics related to medical or social research leaving aside aspects related to technology development/research.

In order to comply with legal data protection, security requirements, and international standards, an informed consent is necessary. For interviews, observations, and workshops, the standard procedure is signing a consent form. However, some PWMD in our projects could not physically sign by themselves. As long as the user does not have cognitive disabilities, a video record of a verbal agreement is generally a legally accepted form of consent. In the case of cognitive disabilities, an authorized representative needs to be on-site to physically sign the forms. In all cases, it is necessary to closely monitor the desires and physical state of the person in the current situation and be prepared to stop the study at any point.

Spiel et al. [47] invite to go beyond managing ethics as a checklist approach in PD. Our experiences align with this statement. For instance, to obtain reliable statements, it was important to consider the experience and knowledge of the technology to be developed. Pictures, video-clips, mock-ups, or prototypes can be means to inform participants and allow for the same level of knowledge. As PWMD and other stakeholders had different ideas and fears with respect to technology. In the HOME project, a participant believed that the robotic arm would be literally attached to the body and requires the amputation of their own arm.

In general, we align with the findings of Bratteteig et al. [9] regarding expectation management. We stress the need for transparency about the goals and a realistic outcome of the research. PWMD might not directly benefit in the immediate future or even long-term from the research outcomes. Consequently, it is important to communicate this to primary and secondary users.

5 LESSONS LEARNED

We present five significant lessons learned that could guide other researchers and practitioners when designing HRC for people with disabilities. These lessons consider Holloway's action plan for disability interaction, "learn from what has been done" [22] and are based on our own challenges.

(1) Approach participants through different channels and allow for multidisciplinarity in the research team. Having hospitals, sheltered-workshops, support groups, or similar institutions as partners can ease participant recruitment. Moreover, getting involved in online communities can help to get in touch with people who are willing to participate but not in close physical proximity. Also, a design team with members from different disciplines allows for in-depth analysis with different viewpoints and facilitates solution development.

(2) Consider the social dependencies in the selection of a PD technique. Designers should consider inter-dependencies between primary and secondary users when gathering information. Interviews gather individual experiences, however, the trustworthiness of the information can be affected by the presence of secondary users, e.g., caregivers. Focus groups or workshops foster discussions and co-design but cause additional stress for PWMD, compromising the discussions. Having a lead user-group can reduce stress in PWMD and ease their involvement along the different design stages. Therefore, it is crucial to set a detailed meeting plan together with secondary users since they have important information about daily routines, requirements, and organizational factors.

(3) Plan for early exposure to robots and other technology. It is vital that participants experience or familiarize themselves to some extent with the technology to be used at an early phase. This could contribute to better evaluate the use of different devices/possibilities; especially, when a certain device/idea has been found difficult to use by the end-users. Further, there is high variability on what type of device works for an individual, especially in the case of PWMD. Also, research technology is often not "ready-to-use" which makes it difficult for participants to experience it at the same level as commercial products, affecting their evaluation. Designers should be prepared for alternative ways of emulating interaction with a robot. It might not always be possible to transport a certain type of industrial robot so that stakeholders have first-hand experience. Hence, showing videos and using immersive 3D technology, can help to create an experience of interacting with a robot.

(4) Take into account all opinions in design sessions. Managing design sessions separately is decisive when stakeholders have potential divergent approaches. This contributes to gain specific perspectives, favor the flow of ideas, and avoid bottleneck discussions. Stakeholders can be grouped by similarities in their backgrounds and goals, and researchers should carry out separate design sessions for each group of stakeholders. For example, session 1 for end-users; session 2 for social workers, caregivers, and medical staff; and session 3 for supervisors and managers. Designers should share the gathered insights from different sessions among all stakeholders to favor mutual information discoveries and allow future collaborations. This allows sensitizing different stakeholders to diverse perspectives, as well as, getting insights about the connectivity of arguments and possible hindrances for technology/solution acceptance.

(5) Acknowledge that ethical implications go beyond consent. Information on the technology to be developed should be transparent, precise, and carefully communicated in order to avoid misunderstandings and manage expectations. Designers should take time to ensure that the purpose and outcomes of the research are fully understood. Also, participants' physical and mental well-being should be carefully analyzed along the design process, and this could be better achieved with the help of social workers, care-givers, and medical staff (secondary users).

Our last two lessons learned align with those of Spiel et al. [47] in terms of prioritizing topics, negotiation of needs, and commitment to participants. Their work reflected on micro-ethical decisions for PD with marginalized children. This hints the relevance of publicizing this type of work when using PD with vulnerable groups and the particular insights that can be derived from each group.

6 LIMITATIONS

We acknowledge that some of the challenges we encountered align with those previously found by people who carry out research with people with disabilities. However, we consider it relevant to present them in a human-robot collaborative context. To the best of our knowledge presenting insights of PD for PWMD in a HRC context is a novel approach. Our goal is that this type of research invites discussions.

Another limitation comes in terms of the techniques used along the research process. Other techniques could provide additional challenges and thereby lessons learned. However, we consider that the aforementioned techniques are among the most frequently used in PD.

Additionally, the use of robotic arms and our specific choice of technology certainly limits the applicability of the proposed lessons learned. Still, many of our observations and insights are not tightly bound to certain technology but should be applicable to other HRC technologies as well.

7 CONCLUSION

Participatory Design has the potential to create solutions that help researchers empathize with the needs of the end-users, which is essential when designing with and for people with motor disabilities. However, in particular in the context of Human-Robot Collaboration, technology still takes a dominant role and makes the application of Participatory Design difficult. Thereby, designers of a human-robot collaborative environment for people with disabilities might encounter similar crossroads to the ones we found in our research.

In this work, we presented 5 lessons learned, summarized as follows:

- Approach participants through different channels and allow for multidisciplinarity in the research team.
- Consider the relationship between social dependencies in the selection of a PD technique.
- Plan for early exposure to robots and other technology.
- Take into account all opinions in design sessions.
- Acknowledge that ethical implications go beyond consent.

We highlight that Human-Robot Collaboration calls for special attention to two main aspects: technology and ELSI. Technology exposure plays a important role since the novelty and complexity factors in the resulting solution to participants. In addition, ELSI is per se a complex and vast aspect that requires careful evaluation in the design process which does not only involve consent but expectation management.

ACKNOWLEDGMENTS

This research is supported by the German Federal Ministry of Education and Research (BMBF, FKZ: 13FH011IX6, 16SV7866K, and 16SV7868).

We would also like to thank all our participants and the personnel at Büngern Technik for opening to us and supporting our research.

REFERENCES

[1] Amazon. 14/01/2021. Amazon Mechanical Turk. https://www.mturk.com/
[2] Stephanie Arevalo Arboleda, Max Pascher, Younes Lakhnati, and Jens Gerken. 8/31/2020 - 9/4/2020. Understanding Human-Robot Collaboration for People with Mobility Impairments at the Workplace, a Thematic Analysis. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 561–566. https://doi.org/10.1109/RO-MAN47096.2020.9223489
[3] Shiri Azenkot, Catherine Feng, and Maya Cakmak. 2016. Enabling building service robots to guide blind people a participatory design approach. In *HRI'16*, HRI (Ed.). IEEE, Piscataway, NJ, 3–10. https://doi.org/10.1109/HRI.2016.7451727
[4] Fabricio E. Balcazar, Christopher B. Keys, Daniel L. Kaplan, and Yolanda Suarez-Balcazar. 1998. Participatory action research and people with disabilities: Principles and challenges. *Canadian Journal of rehabilitation* 12 (1998), 105–112.
[5] Giulia Barbareschi, Sibylle Daymond, Jake Honeywill, Aneesha Singh, Dominic Noble, Nancy N. Mbugua, Ian Harris, Victoria Austin, and Catherine Holloway. 2020. Value beyond function: analyzing the perception of wheelchair innovations in Kenya. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ACM Digital Library)*, Tiago Guerreiro (Ed.). Association for Computing Machinery, New York,NY,United States, 1–14. https://doi.org/10.1145/3373625.3417017
[6] Andrea Bauer, Dirk Wollherr, and Martin Buss. 2008. HUMAN–ROBOT COLLABORATION: A Survey. *International Journal of Humanoid Robotics* 05, 01 (2008), 47–66. https://doi.org/10.1142/S0219843608001303
[7] Marvin Becker, Florian Borsum, Maike Ihmig, and Clara Terbeck. [n.d.]. Wheelchair Traffic Assistant - Ein Konzept für mehr Sicherheit durch Usability. https://doi.org/10.18420/MUC2019-WS-379
[8] Roger Bemelmans, Gert Jan Gelderblom, Pieter Jonker, and Luc de Witte. 2012. Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *Journal of the American Medical Directors Association* 13, 2 (2012), 114–120.e1. https://doi.org/10.1016/j.jamda.2010.10.002
[9] Tone Bratteteig, Ole Kristian Rolstad, and Ina Wagner. 2016. The Life and Death of Design Ideas. In *COOP 2016*, Antonella de Angeli, Liam Bannon, Patrizia Marti, and Silvia Bordin (Eds.). Vol. 8. Springer, Switzerland, 259–275. https://doi.org/10.1007/978-3-319-33464-6{_}16
[10] Tone Bratteteig and Ina Wagner. 2016. Unpacking the Notion of Participation in Participatory Design. *Computer Supported Cooperative Work (CSCW)* 25, 6 (2016), 425–475. https://doi.org/10.1007/s10606-016-9259-4
[11] Assistive Innovations bv. 17/01/2021. https://www.assistive-innovations.com/eatingdevices/ieat-robot
[12] Assistive Innovations bv. 17/01/2021. https://www.assistive-innovations.com/robotic-arms/iarm
[13] Andrea Cherubini, Robin Passama, André Crosnier, Antoine Lasnier, and Philippe Fraisse. 2016. Collaborative manufacturing with physical human–robot interaction. *Robotics and Computer-Integrated Manufacturing* 40 (2016), 1–13. https://doi.org/10.1016/j.rcim.2015.12.007
[14] Craig A. Chin, Armando Barreto, J. Gualberto Cremades, and Malek Adjouadi. 2008. Integrated electromyogram and eye-gaze tracking cursor control system for computer users with motor disabilities. *Journal of rehabilitation research and development* 45, 1 (2008), 161–174. https://doi.org/10.1682/jrrd.2007.03.0050
[15] Brandon Daveler, Benjamin Salatin, Garrett G. Grindle, Jorge Candiotti, Hongwu Wang, and Rory A. Cooper. 2015. Participatory design and validation of mobility enhancement robotic wheelchair. *Journal of rehabilitation research and development* 52, 6 (2015), 739–750. https://doi.org/10.1682/JRRD.2014.11.0278
[16] Disabled World. 2019. Disability Statistics: Information, Charts, Graphs and Tables. https://www.disabled-world.com/disability/statistics/
[17] Anca Dragan and Siddhartha Srinivasa. 2014. Familiarization to robot motion. In *HRI '14*, Gerhard Sagerer, Michita Imai, Tony Belpaeme, and Andrea Thomaz (Eds.). ACM, New York, 366–373. https://doi.org/10.1145/2559636.2559674
[18] Elizabeth Eckhardt and Jeane Anastas. 2006. Research methods with disabled populations. *Journal of social work in disability & rehabilitation* 6, 1-2 (2006), 233–249. https://doi.org/10.1300/j198v06n01{_}13
[19] Mexhid Ferati, Ayesha Babar, Kanani Carine, Ali Hamidi, and Christina Mörtberg. 2018. Participatory Design Approach to Internet of Things: Co-designing a Smart Shower for and with People with Disabilities. In *Universal access in human-computer interaction. Part II*, Margherita Antona and Constantine Stephanidis (Eds.). LNCS sublibrary. SL 3, Information systems and applications, incl. Internet/Web, and HCI, Vol. 10908. Springer, Cham, 246–261. https:

//doi.org/10.1007/978-3-319-92052-8{_}19
[20] Susanne Frennert, Hedvig Aminoff, and Britt Östlund. 2020. Technological Frames and Care Robots in Eldercare. *International Journal of Social Robotics* 27, 11–12 (2020), 2473. https://doi.org/10.1007/s12369-020-00641-0
[21] Natalie Friedman and Alex Cabral. 2018. Using a Telepresence Robot to Improve Self-Efficacy of People with Developmental Disabilities. In *ASSETS '18*, A. Special Interest Group on AccessibleC.M. Computing and Faustina Hwang (Eds.). Association for Computing Machinery, New York, New York, 489–491. https://doi.org/10.1145/3234695.3240985
[22] Catherine Holloway. 2019. Disability interaction (DIX). *interactions* 26, 2 (2019), 44–49. https://doi.org/10.1145/3310322
[23] Harald Holone and Jo Herstad. 2013. Three tensions in participatory design for inclusion. In *CHI 2013*, Susanne Bødker, D. Steven Brewster, Patrick Baudisch, Michel Beaudouin-Lafon, and Wendy Mackay (Eds.). ACM, New York, 2903–2906. https://doi.org/10.1145/2470654.2481401
[24] Andreas Holzinger, Gig Searle, and Michaela Wernbacher. 2011. The effect of previous exposure to technology on acceptance and its importance in usability and accessibility engineering. *Universal Access in the Information Society* 10, 3 (2011), 245–260. https://doi.org/10.1007/s10209-010-0212-x
[25] Kinova Inc. 17/01/2021. https://www.kinovarobotics.com/en/assistive-technologies/column-a1/kinova-assistive-robotic-arm
[26] Sriram Kalyanaraman and S. Shyam Sundar. 2006. The Psychological Appeal of Personalized Content in Web Portals: Does Customization Affect Attitudes and Behavior? *Journal of Communication* 56, 1 (2006), 110–132. https://doi.org/10.1111/j.1460-2466.2006.00006.x
[27] Barbara Klein and Inga Schlömer. 2018. Ein robotisches Duschsystem : Akzeptanz und ethische Fragen. *Zeitschrift fur Gerontologie und Geriatrie* 51, 1 (2018), 25–31. https://doi.org/10.1007/s00391-017-1345-9
[28] Morten Kyng and Lars Mathiassen. 1998. *Computers and design in context*. MIT, Cambridge, Mass. and London.
[29] Maria Kyrarini, Adrian Leu, Danijela Ristić-Durrant, Axel Gräser, Anja Jackowski, Marion Gebhard, Jochen Nelles, Christina Bröhl, Christopher Brandl, Alexander Mertens, and Christopher M. Schlick. 2016. HUMAN-ROBOT SYNERGY FOR COOPERATIVE ROBOTS. *Facta Universitatis, Series: Automatic Control and Robotics* 15, 3 (2016), 187. https://doi.org/10.22190/FUACR1603187K
[30] María Inés Laitano. 2017. Developing a Participatory Approach to Accessible Design. *International Journal of Sociotechnology and Knowledge Development* 9, 4 (2017), 1–11. https://doi.org/10.4018/IJSKD.2017100101
[31] Hee Rin Lee, Selma Šabanović, Wan-Ling Chang, Shinichi Nagata, Jennifer Piatt, Casey Bennett, and David Hakken. 2017. Steps Toward Participatory Design of Social Robots. In *HRI'17*, HRI (Ed.). IEEE, Piscataway, NJ, 244–253. https://doi.org/10.1145/2909824.3020237
[32] Rachael Luck. 2003. Dialogue in participatory design. *Design Studies* 24, 6 (2003), 523–535. https://doi.org/10.1016/S0142-694X(03)00040-1
[33] Saifuddin Mahmud, Xiangxu Lin, and Jong-Hoon Kim. 2020. Interface for Human Machine Interaction for assistant devices: A Review. In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, Satyajit Chakrabarti (Ed.). [IEEE], [Piscataway, New Jersey], 0768–0773. https://doi.org/10.1109/CCWC47524.2020.9031244
[34] Sebastian Merkel and Alexander Kucharski. 2019. Participatory Design in Gerontechnology: A Systematic Literature Review. *The Gerontologist* 59, 1 (2019), e16–e25. https://doi.org/10.1093/geront/gny034
[35] Oussama Metatla, Nick Bryan-Kinns, Tony Stockman, and Fiore Martin. 2015. Designing with and for people living with visual impairments: audio-tactile mock-ups, audio diaries and participatory prototyping. *CoDesign* 11, 1 (2015), 35–48. https://doi.org/10.1080/15710882.2015.1007877
[36] George Michalos, Sotiris Makris, Panagiota Tsarouchi, Toni Guasch, Dimitris Kontovrakis, and George Chryssolouris. 2015. Design Considerations for Safe Human-robot Collaborative Workplaces. *Procedia CIRP* 37 (2015), 248–253. https://doi.org/10.1016/j.procir.2015.08.014
[37] Melanie Nind. 2008. Conducting qualitative research with people with learning, communication and other disabilities: Methodological challenges.
[38] Obi. 17/01/2021. https://meetobi.com/
[39] Westphalian University of Applied Sciences. 17/01/2021. https://hci.w-hs.de/research/projects/mobile/
[40] Westphalian University of Applied Sciences. 17/01/2021. https://hci.w-hs.de/research/projects/mia/
[41] Max Pascher, Annalies Baumeister, Barbara Klein, Stefan Schneegass, and Jens Gerken. 2019. Little Helper: A Multi-Robot System in Home Health Care Environments. In *1st International Workshop on Human-Drone Interaction*. Glasgow, United Kingdom. https://hal.archives-ouvertes.fr/hal-02128382
[42] Reddit. 14/01/2021. Reddit. https://www.reddit.com/
[43] Natalia Reich-Stiebert, Friederike Eyssel, and Charlotte Hohnemann. 2019. Involve the user! Changing attitudes toward robots by user participation in a robot prototyping process. *Computers in Human Behavior* 91 (2019), 290–296. https://doi.org/10.1016/j.chb.2018.09.041
[44] Nina Rudigkeit, Marion Gebhard, and Axel Gräser. 2016. A Novel Interface for Intuitive Control of Assistive Robots Based on Inertial Measurement Units. In *Ambient Assisted Living*, Reiner Wichert and Helmut Klausing (Eds.). Advanced Technologies and Societal Change, 2191-6853, Vol. 12. Springer International Publishing Imprint : Springer, Cham, 137–146. https://doi.org/10.1007/978-3-319-26345-8{_}12
[45] Marta Sałkowska. 2018. Carrying Out Research Among Persons with Disabilities and their Relatives – Selected Ethical Issues. *Zoon Politikon* Special Issue (2018), 200–217. https://doi.org/10.4467/2543408XZOP.18.010.10066
[46] Kristen Shinohara. 2012. A new approach for the design of assistive technologies. *ACM SIGACCESS Accessibility and Computing* 102 (2012), 45–48. https://doi.org/10.1145/2140446.2140456
[47] Katta Spiel, Emeline Brulé, Christopher Frauenberger, Gilles Bailly, and Geraldine Fitzpatrick. 2018. Micro-ethics for participatory design with marginalised children. In *Proceedings of the 15th Participatory Design Conference Full Papers - Volume 1*, Liesbeth Huybrechts, Maurizio Teli, Ann Light, Yanki Lee, Julia Garde, John Vines, Eva Brandt, Anne Marie Kanstrup, and Keld Bødker (Eds.). ACM, [Place of publication not identified], 1–12. https://doi.org/10.1145/3210586.3210603
[48] SurveyCircle. 14/01/2021. SurveyCircle. https://www.surveycircle.com
[49] Jeffrey Too Chuan Tan, Feng Duan, Ye Zhang, Kei Watanabe, Ryu Kato, and Tamio Arai. 2009. Human-robot collaboration in cellular manufacturing: Design and development. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Piscataway, NJ, 29–34. https://doi.org/10.1109/IROS.2009.5354155
[50] Vaibhav V. Unhelkar, Shen Li, and Julie A. Shah. 03092020. Decision-Making for Bidirectional Communication in Sequential Human-Robot Collaborative Tasks. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, Tony Belpaeme, James Young, Hatice Gunes, and Laurel Riek (Eds.). ACM, New York, NY, USA, 329–341. https://doi.org/10.1145/3319502.3374779
[51] John Vines, Rachel Clarke, Peter Wright, John McCarthy, and Patrick Olivier. 2013. Configuring participation. In *CHI 2013*, Susanne Bødker, D. Steven Brewster, Patrick Baudisch, Michel Beaudouin-Lafon, and Wendy Mackay (Eds.). ACM, New York, 429. https://doi.org/10.1145/2470654.2470716
[52] World Health Organization. 2011. World Report on Disability. https://www.who.int/disabilities/world_report/2011/report.pdf
[53] World Health Organization. 2018. Disability and Health. https://www.who.int/en/news-room/fact-sheets/detail/disability-and-health

7.3 Paper III. What's behind a choice? Understanding Modality Choices under Changing Environmental Conditions.

Arévalo Arboleda Stephanie; Miller, Stanislaw; Janka, Martha; and Gerken, Jens. 2019. What's behind a choice? Understanding Modality Choices under Changing Environmental Conditions. In 2019 International Conference on Multimodal Interaction (ICMI '19). Association for Computing Machinery, New York, NY, USA, 291–301. DOI: 10.1145/3340555.3353717.

What's behind a choice? Understanding Modality Choices under Changing Environmental Conditions

Stephanie Arévalo Arboleda
Human-Computer Interaction Group
Westphalian University of Applied Sciences, Gelsenkirchen, Germany, stephanie.arevalo@w-hs.de

Stanislaw Miller
Human-Computer Interaction Group
Westphalian University of Applied Sciences, Gelsenkirchen, Germany, stanislaw.miller@w-hs.de

Martha Janka
Human-Computer Interaction Group
Westphalian University of Applied Sciences, Gelsenkirchen, Germany, martha.janka@w-hs.de

Jens Gerken
Human-Computer Interaction Group
Westphalian University of Applied Sciences, Gelsenkirchen, Germany, jens.gerken@w-hs.de

ABSTRACT

Interacting with the physical and digital environment multimodally enhances user flexibility and adaptability to different scenarios. A body of research has focused on comparing the efficiency and effectiveness of different interaction modalities in digital environments. However, little is known about user behavior in an environment that provides freedom to choose from a range of modalities. That is why, we take a closer look at the factors that influence input modality choices. Building on the work by Jameson & Kristensson, our goal is to understand howdifferent factors influence user choices. In this paper, we present a study that aims to explore modality choices in a hands-free interaction environment, wherein participants can choose and combine freely three hands-free modalities (Gaze, Head movements, Speech) to executepoint and select actions in a 2D interface. On the one hand, our results show that users avoid switching modalities more often than we expected, particularly, under conditions that should prompt modality switching. On the other hand, when users make a modality switch,user characteristics and consequences of the experienced interaction have a higher impact in the choice, than the changes in environmental conditions. Further, when users switch between modalities, we identified different types of switching behaviors. Users who deliberately try to find and choose an optimal modality (single switcher), userswho try to find optimal combinations of modalities (multiple switcher), and a switching behavior triggered by error occurrence (error biased switcher). We believe that these results help to further understand when and how to design for multimodal interaction in real-world systems.

ICMI '19, October 14-18, 2019, Suzhou, China

ACM ISBN 978-1-4503-6860-5/19/10...$15.00
https://doi.org/10.1145/3340555.3353717

CCS CONCEPTS

• **Human-centered computing~User studies**; Pointing; User centered design

KEYWORDS

Experiment, Multimodality, Interaction, Point and Select, Modality Choices, Input Modalities, Hands-Free Interaction

ACM Reference format:
Stephanie Arévalo Arboleda, Stanislaw Miller, Martha Janka and Jens Gerken. 2019. What's behind a choice? Understanding Modality Choices under Changing Environmental Conditions. In *Proceedings of ICMI '19: INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '19), October 14-18, 2019, Suzhou, China.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3340555.3353717

1 Introduction

In a general sense, freedom of choice allows choosing an option that suits best for a particular scenario at a moment in time. However, when people need to make a choice and determine what is "best", people do not always reflect and make choices wisely. Thereby, we decided to take a closer look at the way choices are made and investigate them in terms of input modalities. For that, we base our research on a model by Jameson & Kristensson [8] that contains a set of variables that can influence modality choices. In particular, we aim to get a better understanding of how the user characteristics relate to the features of the situation in terms of their impact on modality choice. As a scenario of use, we decided to look at system control in a workplace setting, where a user operates a remote system, e.g., a robotic system in manufacturing. Such workplace is prone to changing environmental conditions, as different types of distractors, e.g., visual, auditory, levels of task complexity, might affect task execution, and these changes in the environment might interfere with certain input modalities more than with others.

One of the biggest advantage of multimodal interaction is its flexibility and adaptability, in scenarios where users have different abilities, environmental conditions change, or when multitasking is needed [13]. Different multimodal input channels could aid people to adapt to such changing conditions. However, letting users dynamically choosefrom a set of modalities raises a set of questions, mainly,about the different factors that lead a user to prioritize one modality over another. While a specific environmental condition might favor certain interaction modalities, personal preferences, skills, fear, and other user's particularities can also have an influence and might overshadow external factors. Added to that, the frequency and occurrence of the need for multimodal interaction can be intermittent, making it difficult for users to formtrained patterns of interaction. Altogether, this leads us toquestion whether people are actually able to determinethe influence of such environmental conditions during orin between tasks and change input modalities accordingly. In order to explore this question, we focus on hands- free interaction modalities. We implemented gazeinteraction (Gaze) [11], head movements [3] (Head) andspeech interaction [2] (Speech) to control an interface. Ascontrol parameters, our work focuses on a set of basicinteraction tasks in graphical user interfaces, e.g., point(or positioning) and select. Hands-free interaction modalities can support system control at the workplace, as operators might be in a temporary hands-busy situation.For example, in a human-robot collaboration scenario, operators might need their hands for part of the task,hands-free interaction then, still allows operating or finecontrolling the robotic system. For people with a morepermanent physical impairment, e.g., tetraplegia, hands-free interaction modalities can further be a path to an inclusive environment, both in the workplace and at home.

This paper is structured as follows. First, we explore in more detail the existing work related to modality choices. Then, we present our experiment, exploring modality choice under changing environmental conditions. Next, we discuss the implications of our results and the limitations of our experiment before concluding the paper.

2 Related Work

There has been limited effort trying to understand modality choices. Several studies have focused on output modalities [14], e.g., efficiency and safety across output modalities [1] or managing output modalities conflicts [9]. However, input modalities have been less analyzed, especially in terms of modality choices, rather than modalities evaluation and comparison.

It is important to understand how people make choices from a general level. Jameson [7], presents a general model (ASPECT Model) to conceptualize human choice and decision-making process in HCI. He presents six basic choice patterns, based on psychological research: attributes, evaluation depending on particular attributes; social influence, do what others recommend or expect; policies, apply a general policy for a choice; experience, dowhat has been previously best for a user; consequence,anticipate and evaluate consequences of options; trial and error, search what works well for a user [8].

Understanding multimodal interaction requires differentiation of two concepts as stated by Jameson & Kristensson [8]: Modality Combination, and Modality Choice. The concept of Modality Combination refers to one or multiple input modalities that are used simultaneously to perform a task. It can also be aligned with Wickens' Multiple Resource Theory [19], i.e., multimodal stimuli are processed by different structures in the brain in modality specific codes. Cross-modal when the stimuli are processed by different sensorial input channels, and intra-model when the same sensorial input channel processes the stimuli. Added to that, each modality processes information in a different way, depending on the human sense that it is related to [12]. Modality Choice, on the other hand, is concerned with understanding and identifying variables that allow or prevent users from (un)consciously selecting a specific interaction modality in a particular situation. Relevant to this, Wechsung [18] also presented a taxonomy to evaluate multimodal interfaces through a careful consideration of a set of aspects: influencing factors (context, user and system characteristics), interaction performance and interaction quality aspects (joy of use,usability, and ease of use).

Further, Jameson & Kristensson considered a set of previous studies that evaluated different variables separately in terms of modality choices. Then, they presented a model to depict on a high level, the relationships among such choice-relevant variables that users might consider when choosing a modality. They identified a set of variables and group them into three groups. A group defined by the user characteristics: global attributes (age, gender, etc.), evaluation criteria for interaction (novelty, cognitive load, general liking, etc.)and skills or abilities. Another group for features of the situation, which comprise nature of the multimodal task, demands of competing tasks, social and physical environment, and system properties. These two groups of aspects influence consequences for interaction, the third group, in terms of objective aspects of performance (speed, errors, success, physical harm), impact on other tasks, subjective responses to interaction (enjoyment, unfamiliarity, stress). When determining a modality combination, users might or might not consider user characteristics and features of the situation. Added to that, features of the situation, consequences for interaction, and user characteristics could influence interaction evaluation. That is why we designed an experiment to explore theweight of these groups when choosing, or combining modalities.

3 Study

To explore users' modality choice behavior in pointing and selecting tasks under various environmental and physical conditions, we conducted a controlled study. Referring to Jameson & Kristensson's model of modality choice, as discussed previously, we can distinguish three main areas that could influence users to make a deliberate choice: features of the situation, user characteristics, these two in turn influence the third area, consequences for interaction. As features of the situation, this study implemented a total of six of different external conditions (Noise, Visual, Physical, Time, Interaction Complexity, and None; as will be detailed in the Experimental Design Section) with the goal of observing and understanding ifand how they impact users' choice behavior. In order to enable this evaluation, users were able to choose and switch between Gaze, Head and Speech interaction.

Research Questions

In order to better understand people's modality choices, we designed the study with the following research questions in mind:

Q1: We wanted to explore if people recognize that certain modalities are by design better or worse suited for specific conditions, which relates directly to the design of these conditions (discussed in the following section). Each condition favored or hindered specific modalities. Furthermore, we wanted to understand how people actually behave when conditions suddenly change, and if certain usage patterns can be identified among our participants.

Q2: We were also interested in gaining a better understanding of the influence of individual performance and interaction perception on modality choices.

Q3: We assumed that our modalities would lead to differences in cognitive load, e.g., physical demand. Our goal was to understand, how this together with personal preference could affect the modality choice behavior.

Based on these questions, the study is explorative in nature and specifically, it investigates users' modality choice behavior –it does not examine or compare in detail the effectiveness or efficiency of the three interaction modalities (Gaze, Head, Speech) and was not designed to allow that.

Experimental Design & Metrics

The study followed a within-group design with external conditions as the independent variable. A total of six- factor levels were included: Noise, Visual, Physical, Time, Interaction Complexity, and None. For counter-balancing the order of appearance, a Latin-Square design wasapplied, and the basic task remained identical for all conditions. In order to enable replicability of the experiment and transparency, we added more information in a Mendeley repository (http://tiny.cc/roa95y).

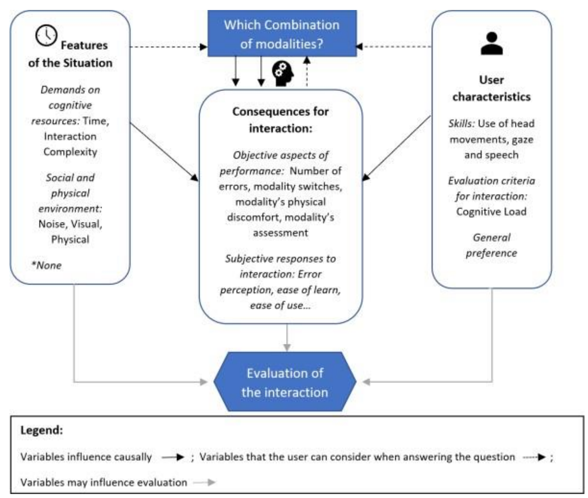


Figure 1: Variables applied in the study and their relationships taken into account to analyze modalitychoices, applying the Jameson & Kristensson's model [8].

To illustrate our variables we present them in relation to the model by Jameson and Kristensson, as shown in Figure 1. The "Features of the Situation" are represented by the six different external conditions, which are basedon the scenario presented in the introduction (system control in a workplace setting). These conditions not onlyto simulate possible changes in the environment, but were designed with their effect on certain modalities in mind, and they are explained as follows:

- The Noise condition introduced a loudly playing conversation track (conversation about fixing a car, including engine sounds), representing a constant sound of other machines or people talking in the surroundings. Speakers were placed in front of the participant next to the screen, see Figure 2C. It potentially has a negative interference with Speech interaction, but not with Head or Gaze.

- The Visual condition introduced a visual distractor that could appear while performing atask. Humans use eye-gaze to explore the environment, thus, when a new element appears on the environment people tend to direct their gaze to it. In our study, a set of growing squares appeared and disappeared from different random positions on the screen. The Visual condition, therefore, should mostly interfere with Gaze.
- The Physical condition required participants to hold a desk phone (see Figure 2C) between neck and head, thereby, limiting their head movements. It represents a temporary or permanent neck mobility limitation that some people might have due to an accident or aging factors. This condition was designed to represent an "easy to relate to" scenario. It should interfere with Head modality, as it limits the possibility to freely move the head, but should not interfere with Gaze or Speech.
- The Time condition introduced a time limit of 30 seconds for a task. It represents cognitive demand typical in hectic environments, where a task needs to be finished in time to not hold the further production steps. In the study, it requires users to more carefully choose the modality which they consider to be most efficient with in order to stay within the time limit.
- The Complexity condition increased the difficulty of the task by requiring more fine detailedpointing and selecting. It presents a second type of cognitive demand, referring to a task that needs a higher level of precision. Hansen et al. [4] showed that Head might be more precise compared to Gaze.
- The None condition was specified by the absence of any added constraint.

The dependent variables are presented in Figure 1 as the "Consequences for interaction", and they were basedon the three evaluated modalities (Gaze, Head and Speech). Behavioral data from the study were analyzed by measuring the following aspects: number of choices, number of trials a modality was used in, number of switches of modalities within trials and conditions, and number of errors. The definition of an error had to bespecified for each modality. In Gaze and Head, we considered an error when a dwelling operation wasinadvertently aborted. In Speech, we defined an error was committed when a command was not recognized. In both cases, we did not distinguish between user error (command mispronounced) or a system error (unrecognized command). A satisfaction questionnaire and qualitative feedback interview were used to assess modality discomfort and overall modality assessment.

Further variables were measured to investigate the potential influence of user characteristics. The level of skill with respect to each interaction modality was measured through task times at the end of the training phase. Likewise, Cognitive Load was measured by a short version of the N.A.S.A. task load index (TLX) [5], also known as Raw TLX (RTLX), and General preference was assessed ina Pre-Test questionnaire.

There were 24 trials per participant (6 conditions x 4 trials), totaling in 288 trials. Besides, each participant conducted a training phase with each modality, consisting of at least 3 trials per modality. The study applied a mixed-method design, including both qualitative and quantitative data collection and analysis.

Tasks

Figure 2A presents the different options on the Menu (Select, Go, Rotate, Grasp). "Select" activates the gripper. "Go" activates the option of moving the gripper to a different location on the screen. "Rotate" activates two new items to start rotating the gripper left or right, see Figure 2B. "Grasp" closes the gripper to grasp an object, when an object has already been grasped the label on that item changes to "Release", which opens the gripper and leaves the grasped object in the gripper's position.

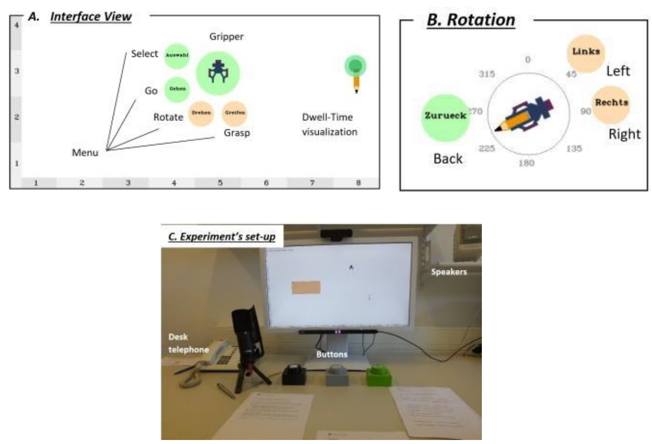


Figure 2: A. The Interface View presents the menu around the gripper and the dwell visualization on the right. B. Rotation: Presents the interaction options during rotation of the gripper, left or right and back to return to the menu. C. The experiment's set up with the different devices used to represent changes in the environment.

All modalities functioned in the same manner through the Menu, but with the inherent particularities of each modality for selecting and pointing. In Gaze and Head, eye-dwell and head-dwell are used correspondingly, to activate the options on the Menu, and a position to move the gripper to is also selected dwell. For rotation, the gripper will rotate as long as the option “Right” or “Left” isdwelled, and stop immediately when dwells is directed to another position, rotation can be resumed if the option is dwelled again within 640 ms. In Speech, users had to say the label on the Menu’s option to activate it, preceded by the keyword “Function”, e.g., “Function Go”. Participants were given a list of the available speech commands asreference. Pointing a position is not a straightforward taskin Speech. Our approach to tackle this problem applied a coordinate system on the 2D interface. For pointing the user had to name target coordinates, e.g., “Go to 6 [comma] 5”, with 6 being the value on the x-Axis and 5 on the y-Axis. To improve usability, we also showed the pointer representing the Head position, which had the current coordinates printed next to it, allowing the user to identify the target coordinates. For rotation, the action of rotating could be stopped by saying the keyword “Stop”.

Users first learned the interaction modalities individually and got familiar with the interface with a training task. For the training task, users first had to activate the gripper. Then, move the pointer towards the item to be grasped (pencil). Next, grasp the item and rotate it.

The experimental task was in principle similar to the training task, but with additional steps: After grabbing an item (book), the user had to move it to the table, rotate it upside-down (roughly 180°) and release it. Additionally,for the Complexity condition, users had to precisely rotatethe book to match a displayed visual template. For theTime condition, users were given a time limit of 30seconds. We presented a visual timer on a mobile phoneand placed it next to the microphone. Besides that, the moderator informed the participant every 10 seconds about the remaining time for the completion of the task. If the participants were not able to finish the task, they weregiven the option to finish it or continue with the next trial.The task was designed in this way to represent: first, a basic object manipulation scenario for operating a remotesystem (in this case robot arm); second, to includedifferent point and select actions, which might take advantage of the strengths of each modality in different circumstances.

Procedure

The procedure of the experiment is shown in Figure 3. After a standardized introduction and consent form, a background questionnaire was presented where we asked users about their anticipated preference of the modalities by ranking them from 1 to 3. Then, a training phase was performed aiming to overcome initial learning obstacles and reach an intermediate to high level of proficiency for all users (with at least three trials per condition). Participants could continue with additional trials, if they did not feel confident yet using a particular interaction modality. In addition, the training phase also allowed us to have a controlled set of modality usage. We capitalized on this by measuring task times for the last trial of thetraining phase and measuring cognitive load for each modality with the NASA RTLX.

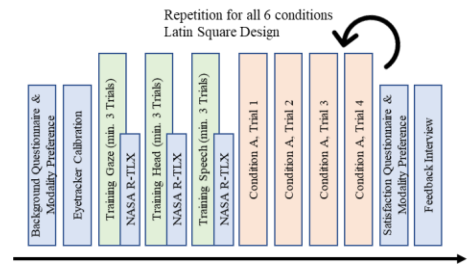


Figure 3: Procedure of the Experiment

After the training phase, participants were instructed, that they could for the upcoming tasks: use, switch and combine the modalities as desired at any moment while performing the task. For switching modalities, they had to press the modality’s corresponding button located in front of them. Participants went through the experimental trialsin a Latin Square design for all conditions (24 trials in total). After executing the tasks, participants were given a satisfaction questionnaire based on Five-point Likert items (Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree), where they were asked about their level of exhaustion, ease of use, accuracy, ease of learning,effectiveness, and speed as well as their ranking of the modalities, followed by a short interview. Each sessiontook on average 60-90 minutes to complete (training time M=15.8 minutes, task execution M=25.1 minutes and therest of the time for an interview, questionnaires, and calibration of the tracking device).

Participants

We recruited 14 participants, through analog and digital advertising at the University. However, the data of 2 of them were discarded, due to calibration problems with the eye tracker, which made difficult for the participants to control the interface using Gaze or Head. Of the resulting 12 participants, all were students or staff from theUniversity, 3 female, 9 male, average age M = 28.9, SD = 7.2 (range 20-49 years old). The pre-test questionnaire revealed that only 2 were familiar with eye trackers and gaze interaction, but 10 had experiences using speech control –Siri or Alexa. Five participants wore eye-glasses for correcting their vision. Participants received 7€ for their participation.

Apparatus

The experiment was conducted in a Desktop-PC setting with participants sitting approx. 60cm in front of a 24” screen (see Figure 2C). The study-application whichintegrated the user task (see Figure 2A) was written inOpenCV 3.4.1 [6] and showed a simple 2D game-like interface displaying a gripper as used on robotic arms and additional objects related to the task (pencil, book, andtable). Interaction commands were presented in a menuaugmenting the gripper upon activation (select, move, rotate, grab). Users’ gaze and head orientation wasrepresented with a circular pointer (black for Gaze and purple for Head) – both only in case the respective input modality was activated.

A Rode NT-USB microphone [16] was used for speech recognition in combination with the Microsoft Speech API 5.1 [10], which is available directly in Microsoft Windows OS (Version 10). The recognition mechanism was trainedto improve recognition rates. Furthermore, speechcommands were chosen to achieve high robustness andnot interfere with Windows system commands. For both gaze and head tracking, we used a Tobii Eye tracker 4C [17] and a Tobii Streaming Engine v1.1.1.1741. The Tobii API provides synchronized callback functions to detectchanges of head position and rotation, allowing apragmatic reading of head parameters. Tobii determinesthe head position as an array of three floats (x,y,z) for theposition of the head, and another three floats to track itsrotation. Depending on the automatically queried resolution of the test computer, the angles are mapped to the monitor and illustrate a direct reproduction of the head rotation according to the center of the display. Anindicator on the screen illustrates the current position ofthe head alignment and works as user feedback. In ourexperiment, only the yaw and pitch information weretaken, covering the 2D screen area. Further processing ofthe data was done to adapt the normalized raw data to thescreen resolution. As Tobii already does some basic outlierfiltering in the “raw” data, no further filtering was applied.Selection for Gaze and Head was done via dwelling. The dwell timer was triggered after staying above aninteraction object for 640ms and required the user to staythere until 1600ms to trigger activation. We opted for arather long dwell time compared to related work to avoidfalse positives and decided on a dwell time based on ourpilot testing of the application. As a feedback mechanism,a circle around the pointer was smoothly colored duringthis time span and activated objects changed their color (Figure 2A).

Switching between modalities was achieved via three physical buttons placed in front of the participant (Gaze: green button, Head: black button, and Speech: gray button), see Figure 2C. This mechanism was chosen so that modality choice and switching does not interfere with an actual interaction modality. The technical switch between modalities was however conducted by an invisible (to the participant) human operator, similar to Wizard-of-Oz studies, who observed the scene, and also logged certain events for the analysis. After each switch, the user received on-screen feedback to indicate when a new modality was active. This made sure that users always know that a modality switch succeeded before they try to use a new modality.

Results

Behavioral Examination . We analyzed how the changes in the environment affect the use of a modality. From this, we found that in 55% of cases participants did not switchto a different modality when the condition changed, andin 45% of cases, they actually switched to a different modality. A χ^2 test confirms that this is not significant different from a balanced distribution χ^2 (1, N = 60) = 0.6,p >.05.

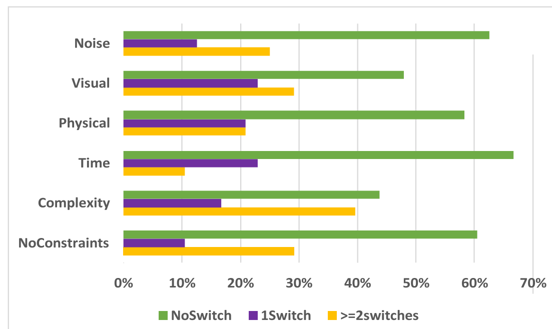


Figure 4: Percentages of modality switches per condition

Figure 4 depicts the switching distribution among conditions, which shows a pattern that is significantly different from a balanced one (χ^2 (10, N = 288) = 79.40, p <.01). Looking in detail, it reveals that for the Time condition, in 66.67% of the trials, participants avoided switching. The Complexity condition seems to invite users to switch modalities, as in 39,58% of trials participants switched 2 or more times. Also, this condition has thehighest percentage of 2 or more switches among the restof the modalities. A possible explanation could be thatparticipants were exploring with different modality combinations to find the most adequate.

Table 1: No. of Switches per Trial

Participant	No Switch	1 Switch	>=2 Switches
P1	12	3	9
P2	13	2	9
P3	6	0	18
P4	15	3	6
P5	13	10	1
P6	14	4	6
P7	18	2	4
P8	13	6	5
P9	16	2	6
P10	14	7	3
P11	13	7	4
P12	16	5	3
Total	163	51	74

Nevertheless, a more thorough analysis of the switching pattern was needed. Table 1 considers switches within trials from a trial by trial analysis. This analysis shows that in most trials, participants did not switch modalities. In general terms, in 56.60% of the trials, participants did not switch modalities, in 17.7% they switched once, and in 25.69% they switched twice or more. This trend was kept regardless of the condition, and participants stated that sometimes they thought to be quicker without modality switches (P4, "If I use only one modality I will be quicker"), or found it cognitive demanding to think about switching (P7, "It is hard to make the switch in your mind from one modality to another"). Examining the cases when participants actually did switch modalities on an individual levelin detail, we recognized a certain pattern and identify two groups of participants based on the switching occurrence. One group who made one switch and the other group that made at least 2 switches, Table 1 depicts that behavior. The group that switched mostly once (P5, P10, P11, P12) is highlighted in orange, and a second group that mostly switched twice or more (P1, P2, P3, P4, P6, P9) is highlighted in green, with only P8 matching both behaviors. Following the concept of Jameson & Kristensson, the "once switcher" seem to apply a modality choice strategy and the "multiple switcher" seem to try a modality combination strategy. For instance, P1 used the sequence combination of Gaze-Speech-Gaze during the four trials on the Visual condition. He used Gaze for pointing at an object and moving it, then, used Speech for rotating it, and finally wentback to Gaze for moving it to the final position. Also, he stated,"I am looking for the most efficient combination". When weanalyzed the modality use within a trial, we found that within the multiple switcher group, 72,97% of the times the same modality was used at the beginning and the end of a trial, which we consider a "back-and-forth" switching strategy.

Further, we evaluated if the growing familiarity with the interface and modalities over the course of the study, affects the switching behavior. For this, we grouped the trials in groups of 4. From 1st-4th trials there were 58switches in total, 5th-8th had 49 switches, 9th-12th had 30 switches, 13th-16th had 23 switches, being the one withthe least switches, 17th-20th had 34 switches, and finally the 21st-24th had 36 switches. The lowest number of switches happened within the 13th-16th trials, i.e., the participants used one modality most of the time. In general terms, we could say that the participants showed a more exploratory behavior during the first and last trials, and a more stable and constant behavior in between. This pattern seems similar to the Mere-exposure effect, which proposes that repeated exposure increases familiarity, and within the 10th and 20th presentation the Mere-exposure effect reaches its maximum [20].

Besides the switching behavior, we evaluated the number of errors that occurred when each modality was used – a usage being equivalent to a deliberate selection of this modality via pressing its corresponding button, no matter how long the user stayed with this modality. Table 2, shows the errors for each modality and the total usagein terms of modality selections. We find that when using Gaze in 30.49% of the cases an error occurred, when using Head in 39.22% and when using Speech in 68.22% of the cases.

Table 2: Total of Errors vs Total of Usage

Modality	Total usage	Total errors	Percentage
Gaze	246	75	30.49%
Head	51	20	39.22%
Speech	214	146	68.22%

It is equally important to consider that some modality switches might have been influenced by an error while using a certain modality. For that, we analyzed the number of switches based on errors. The experiment had a total of 241 errors and 125 trials wherein switches occurred. When taking a closer look at the moment when these switches happened, we found that in 47 out of the 241errors, participants performed a switch. This adds a new type of switching behavior, the "error biased switcher", since in 19,5% of the cases an error occurred, participants switched to a different modality.

Preferences and UsageI . n order to understand the preferences of the participants, we considered how they ranked them before and after the experiment. Before starting the experiment there was a clear preference for Speech, then Gaze and Head the least opted choice. One possible explanation for this preference could be that 10out of 12 participants were already familiar with speech- based interfaces (Siri, Alexa, etc.). After performing the experiment this ranking changed, and most participants preferred Gaze control as a first option, leaving Speech and Head in a tie as a second option.

We were interested to analyze if the preferences were reflected in how participants used the modalities, see Figure 5. The percentage of usage was calculated by the number of trials wherein a modality was used at least one time (multiple uses per trial still counted only once). We can see that Gaze was the most used modality during the

Time, Noise, and Complexity conditions. Surprisingly, Gaze was still used during the Visual condition, but it is the second least used among all the uses of Gaze. As expected, Speech was the most used modality in the Physical condition, since the design of this condition was intended to hinder the use of Head. This condition also seemed to hinder Gaze to some extent. In addition, Head was the least used modality overall, which was also reflected by statements in the interview (P10, "It did not feel so natural to move my head. Using speech or gazefeels more familiar"). Within this lack of use, the condition wherein participants used Head the most, was during the Visual condition.

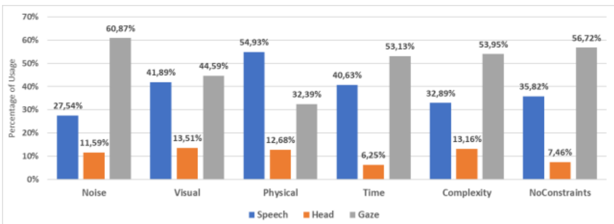


Figure 5: Percentage of modality Usage per condition

Aspects of Performance and Responses to InteractionT . o assess individual performance in terms of task times with each modality, we looked at the individual performance during the training phase. We only considered the tasktime (expressed in seconds) on the last trial of the training phase to exclude initial learning (M_{Gaze}=36.33, SD=10.46; M_{Hea}d=75.5, SD=27.57; M_{Speech}=63.17, SD=21.75). A repeated-measures ANOVA shows a significant main effect on interaction modality (F(1,11)=198,49, p<0.01) with Mauchly's Test of Sphericity indicating that the commands"). Head was considered the hardest to use and assumption of sphericity had not been violated, χ^2 (2) = 0.81, p = .667 and an inspection of residual histograms and Q-Q plots showed no violation of normality distribution. Post-hoc pairwise comparisons (Bonferroni adjusted) show that Gaze outperformed both Head (T(11)=-4,95, p<0.01) and Speech (T(11)=-4,01, p<0.01). We compared participant's preference vs training times and found that 9 out of 12 participants chose as their preferred modality, after executing the experiment, the one that they were the fastest with during the training phase. Whereas in 10 out of 12 participants, the most used modality was the one that they were the fastest with.

Also, participants were asked to evaluate the different modalities through a subjective questionnaire with six categories. A five-point Likert Item was used in order to determine their level of agreement with each criterion. Figure 6 shows that using Speech was perceived as the one with the least errors, conversely to what is shown on the actual error rate in Table 2. Speech is perceived as the easiest to learn, use (P5, "Speech is the easiest to use") and accurate (P4, "Speech is accurate if you know the mostly inaccurate. Gaze was voted as the fastest modality –confirmed by the fact that Gaze was the most used modality during the Time condition. Gaze was alsoconsidered the most effective and with high accuracy. Concerning ease of use and learn it rated high in both (P6, "Everything was positive in using Gaze"). All in all, Speech and Gaze got the highest scores in all the evaluated criteria. Moreover, physical comfort was also evaluated, but none of the modalities were perceived as causing major physical discomfort.

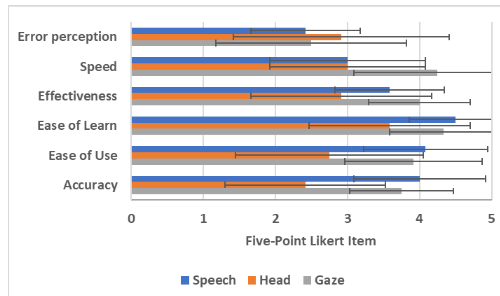


Figure 6: Average rating of modality assessment, the error bars represent the Standard Deviation.

Regarding cognitive load, the RTLX questionnaires showed differences about the cognitive load of each modality (see Figure 7). For statistical tests of significance, we applied RM-ANOVA (significance level 5%, inspection of residual histograms and Q-Q plots showed no violation of normality distribution). For physical demand, the results show a significant difference between the modalities, with Head being rated most demanding ($F_{Physical}$ (2,22)=19.12, p<0.001, η^2_p=0.64, Sphericity not violated with χ^2 (2) = 0.26, p = .88, M_{Speech} =19.17,

SD=16.76, M_{Head}=55.41, SD=17.38; M_{Gaze}=27.08, SD=21.05;). A similar result can be found for frustration ($F_{Frustration}$ (1.34,22)=3.79, Sphericity violated, Greenouse-Geisser corrected: p=0.063, η^2_p=0.26, M_{Speech}=42.08, SD=24.07, M_{Head}=62.08, SD=28.24; M_{Gaze}=39.58, SD=16.30).

A Bonferroni adjusted post-hoc pairwise comparison (over the six dimensions) shows that the differencebetween Head and Speech, as well as Head and Gaze, is statistically significant for the Physical Demand dimension (p<0.005, Bonferroni adjusted). Also, it is significant for Head and Gaze in the Frustration dimension (p<0.01), but not for Head and Speech. This trend of the use of head being rated worse than the other modalities can be seen for all dimensions of the RTLX. However, the results here are not statistically significant.

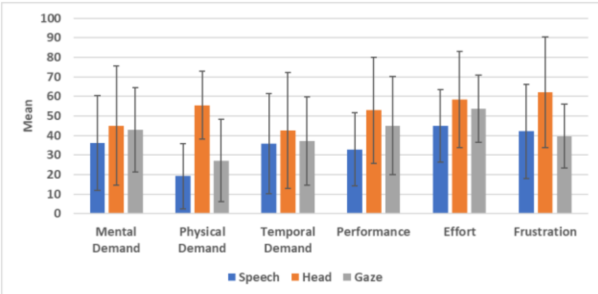


Figure 7: Average of response in RTLX questionnaires per modality. Error bars represent Standard Deviation.

Overall using Head was perceived as more frustrating (P6, "When you use your head and it does not respond you get frustrated faster") with major effort (P2, "Using head movements is just too exhausting") and physical demand (P1, "Rotating with the head is impossible"), and with performance difficulties (P3, "It is very difficult to reach the menu"). Gaze was perceived as the least frustrating and with good performance (P8, "I got used to gaze very fast and it is the easiest") despite the mental (P3, "When using gaze you have to concentrate very much") and physical (P1, "I actually stopped breathing for a few seconds when I was using gaze") demand that was expressed. Speech showed to be the one with the leastmental and physical demand, as well as the one with the least effort, but with medium performance (P6, "I wouldn't use speech for turning since I have problems stopping it"), which explains why frustration is a little bit higher when using Speech than when using Gaze. Moreover, Speech was mentioned as the most familiar modality, since mostof the participants have had experience using it.

4 Discussion

First (Q1), we wanted to explore if users can recognize that certain modalities are better suited than others under different conditions (features of the situation). Surprisingly, our results showed that other factors such as user's characteristics and interaction consequences, have a higher impact in modality choices than changes in the environmental conditions. Despite that, we would like to discuss the modality use under the different conditions:

- The use of Speech decreased under the Noise condition when compared to its use on the other conditions. However, users do not avoid completely using Speech. From our observations, users try to adapt to it, e.g., speaking louder or saying a command when there is a pause in the noise.
- The use of Gaze during the visual condition was amongst the lowest. Users do not avoid using gaze when there is a visual distractor. However, during the interviews, users commented that it had some impact only at the beginning of exposure.
- We could not prove if the usage of Headdecreases when there is a physical constraint on the neck (Physical condition). Since Head was the least used modality, we cannot really determine if users did not use it because of the constraint or simply because they were not willing to use it. Also, the Physical constraint hindered the use of Gaze, which was not expected. This explains why Gaze was used the least during this constraint when compared to its general use.
- We wanted to understand if users identify certain advantages of some modalities. We can say that most users recognized the speed of Gaze. Gaze was rated as the one with the highest speed, and the most used during the Time condition. Regarding error perception, we found surprising results: Speech, as the modality with the least error perception, was the one that had a higher actual error rate.

Second (Q1, Q2), we wanted to understand users' behavior and find interaction patterns, and for that, we analyzed the behavior of participants within trials. We discovered that in around the same percentage of trials, users either do not switch or switched multiple times when the conditions changed. An explanation to switching avoidance can be that switching implies a higher cognitive effort, since it implies a different problem-solving strategy than staying with a previously used modality [18]. It is interesting to note that when users switched modalities, we found two behavioral patterns: a "multiple switcher", who tries to find amodality combination strategy, and a "single switcher",who has a modality choice strategy. In addition, we found another switching behavior due to the occurrence of an error while using a deliberately chosen modality, "error biased switcher". An error occasionally led to a switch, since most participants kept using the chosen modality,despite the occurrence of an error. In fact, we were surprised to see some participants trying to combine modalities in such a way, as we would expect that to be cognitively more demanding. Still, as about half the users tried to focus just on finding the one "optimal" modality, designing for modality combination is difficult to achieve without excluding this group of users.

On those grounds, modality switches are based on user's personal experience (individual performance) when using a modality. For instance, participants mostly chose and used modalities that they were the fastest with, presented low physical discomfort and had a medium tolow number of errors. In consequence, these factors influenced interaction perception, since those same modalities were the highest ranked and had positive subjective responses in the modality assessment.

Third (Q3), regarding interaction evaluation in terms of cognitive load, we found that there are not many significant differences among the perceived cognitive load, except for physical demand and frustration. However, we see a general trend of rating Head worse. While higher physical demand is expected, we did not anticipate a low rating in the other dimensions. In terms of modality use, most participants used more frequently the modalities that they were the fastest with. Further, individual skills in the use of a modality seem to vary greatly based on actual performance –but seem not much affected by the different environmental or prior preferences.

5 Limitations

First, we need to consider the design of the conditions. During the experiment we realized that some conditions were not hindering the use of certain modalities as anticipated, e.g., the visual constraint was not hindering the use of Gaze and users were able to block such distractors out. We encourage researchers to look at different designs of the conditions in the future.

A second limitation concerns the choice of a 2D interface on a screen – as we know from working with Head-mounted Augmented Reality Glasses. Head-based pointing can feel more natural in a more immersive environment. Thus, it can change the preference for Head. Still, as our focus was on understanding modality choices rather than comparing specific modalities, this should not change our overall findings.

Lastly, Gaze and Head relied on an initial calibration, but the accuracy of an eye tracker is still dependent on the anatomy of the user's eyes. Hence, as Qvarfordt [15] points out, eye tracking can lead to limitations in terms of tracking error due to physical individual differences, e.g., glasses, small eyes, droopy eyelids. As discussed, this ledto the exclusion of 2 participants. Also, the interaction modalities of Head and Gaze were not adapted to behavioral individual differences, e.g., it was difficult for some participants to avoid drastic changes from their initial calibrating position. Based on the overall acceptance of the Gaze modality, we emphasize that these limitations did not lead to significant interferences withour results.

6 Conclusions

Our goal in this paper was to provide an analysis that helps to understand the factors that influence users' reasoning behind choosing a hands-free modality when the environment changes. To enable this, the actions of point and select were tested under different conditions that represented changes in the environment. In order to visualize better the variables that influence modality choices, we adapted the Jameson & Kristensson model to represent the variables in our experiment. Our results showed that some variables have a higher impact than others, in terms of interaction evaluation and modality choices. Specifically, users' characteristics and consequences for interaction have a higher impact than features of the situation, represented in our experiment as different changes in the environment. Among the users' characteristics that stand out in our experiment are personal preferences (skills and use), which are determined by the level of proficiency with a modality, and this impacts how a user selects a modality over another. Moreover, regarding the consequences for interaction, the resulting experience (actual and perceived speed, success rate, and error) of using a modality is also a determining choice factor, since it defines how a modality is evaluated and thereby influences its frequency of use. Further work is required to determine the nature of modality choices in more sophisticated tasks and conditions, however our results can shed some light about how and when to design a multimodal interaction.

Our future work will follow the lines of Jameson and Kristensson and combine the different strategies from the ARCADE model [7] to support people in making sensible modality choices. This could be achieved by evaluating how each strategy of the model impact choices under changing environmental conditions and determining which strategy suits better given a certain condition or group of conditions.

ACKNOWLEDGMENT

This research is supported by the German Federal Ministry of Education and Research (BMBF). We would further like to thank Tim Rittich for his support during the experiment.

REFERENCES

[1] Brumby, D. P., Davies, S. C.E., Janssen, C. P., and Grace, J. J. 2011. Fast or safe? In 29th Annual CHI conference on human factors in computing systems. Association for Computing Machinery, New York, 473. DOI=10.1145/1978942.1979009.

[2] Elepfandt, M. 2012. Pointing and speech. In Proceedings of the 7th Nordic Conference on Human-Computer Interaction Making Sense Through Design. ACM, New York, NY, 807. DOI=10.1145/2399016.2399158.

[3] Gizatdinova, Y., Špakov, O., Tuisku, O., Turk, M., and Surakka, V. 2018. Gaze and head pointing for hands-free text entry. In Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications - ETRA '18. ACM Press, New York, New York, USA,1–9. DOI=10.1145/3204493.3204539.

[4] Hansen, J. P., Rajanna, V., MacKenzie, I. S., and Bækgaard, P. 2018. A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In Proceedings of the Workshop on Communication by Gaze Interaction - COGAIN '18. ACM Press, New York, New York, USA, 1–5. DOI=10.1145/3206343.3206344.

[5] Hart, S. G. and Staveland, L. E. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In

Human mental workload, P. A. Hancock and N. Meshkati, Eds. Advances in Psychology 52. North-Holland, Amsterdam, Oxford, 139–183. DOI=10.1016/S0166-4115(08)62386-9.

[6] Intel Corporation. OpenCV Library. https://opencv.org/. Accessed 28 September 2018.

[7] Jameson, A. 2013. Choice Architecture for Human-Computer Interaction. Foundations and Trends, in Human-Computer Interaction 7, 1–2, 1–235.

[8] Jameson, A. and Kristensson, P. O. 2017. Understanding and supporting modality choices. In The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1, S. Oviatt, B. Schuller, P. R. Cohen, D. Sonntag, G. Potamianos and A. Krüger, Eds. ACM, 201–238. DOI=10.1145/3015783.3015790.

[9] Lebeck, K., Kohno, T., and Roesner, F. 2019. Enabling Multiple Applications to Simultaneously Augment Reality. In Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications - HotMobile '19. ACM Press, New York, New York, USA, 81–86. DOI=10.1145/3301293.3302362.

[10] Microsoft. Microsoft Speech SDK. https://www.microsoft.com/en-us/download/details.aspx?id=10121. Accessed 28 September 2018.

[11] Neggers, S. F. and Bekkering, H. 2001. Gaze anchoring to a pointing target is present during the entire pointing movement and is driven by a non-visual signal. Journal of neurophysiology 86, 2, 961–970.

[12] Norris, S. 2009. Analyzing multimodal interaction. A methodological framework. Routledge, New York, NY.

[13] Oviatt, S. L. and Cohen, P. R. 2015. The paradigm shift to multimodality in contemporary computer interfaces. Synthesis lectures on human-centered informatics 30. Morgan & Claypool Publishers, San Rafael.

[14] Prewett, M. S., Yang, L., Stilson, F. R. B., Gray, A. A., Coovert, M. d., Burke, J., Redden, E., and Elliot, L. R. 2006. The benefits of multimodal information. In ICMI '06. Eighth International Conference on Multimodal Interfaces : conference proceedings, November 2-4, 2006, the Fairmont Banff Springs Hotel, Banff, Alberta, Canada / SIGCHI, special interest group on computer human interaction. Association for Computing Machinery, New York, N.Y., 333. DOI=10.1145/1180995.1181057.

[15] Qvarfordt, P. 2017. Gaze-informed multimodal interaction. In The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1, S. Oviatt, B. Schuller, P. R. Cohen, D. Sonntag, G. Potamianos and A. Krüger, Eds. ACM, 365–402. DOI=10.1145/3015783.3015794.

[16] Rode. Rode NT Microphone. http://www.rode.com/microphones/nt-usb. Accessed 11 September 2018.

[17] Tobii. Tobii Eye Tracker 4C. https://tobiigaming.com/product/tobii-eye-tracker-4c/. Accessed 11 September 2018.

[18] Wechsung, I. 2014. An evaluation framework for multimodal interaction. Determining quality aspects and modality choice. T-labs series in telecommunication services. Springer, Hedielberg.

[19] Wickens, C. D. 2008. Multiple resources and mental workload. Human Factors: The Journal of the Human Factors and Ergonomics Society 50, 3, 449–455.

[20] Zajonc, R. B. 1968. Attitudinal effects of mere exposure. Journal of Personality and Social Psychology 9, 2, Pt.2, 1–27.

7.4 Paper IV. Opportunities and Challenges in Mixed-Reality for an Inclusive Human-Robot Collaboration Environment.

Arévalo Arboleda, Stephanie; Pascher, Max; and Gerken, Jens. 2018. Opportunities and Challenges in Mixed-Reality for an Inclusive Human-Robot Collaboration Environment. In 1st. International Workshop on Virtual, Augmented, and Mixed Reality for Human-Robot Interaction (VAM-HRI), held in conjunction with ACM/IEEE Human-Robot Interaction (HRI 2018).

Opportunities and Challenges in Mixed-Reality for an Inclusive Human-Robot Collaboration Environment*

Extended Abstract†

Stephanie Arévalo Arboleda
Westphalian University of Applied Sciences
Germany
stephanie.arevalo@w-hs.de

Max Pascher
Westphalian University of Applied Sciences
Germany
max.pascher@w-hs.de

Jens Gerken
Westphalian University of Applied Sciences
Germany
jens.gerken@ w-hs.de

ABSTRACT

This paper presents an approach to enhance robot control using Mixed-Reality. It highlights the opportunities and challenges in the interaction design to achieve a Human-Robot Collaborative environment. Human-Robot Collaboration could be an appropriate space that allows for social inclusion. It enables people with severe physical impairments, to interact with the environment by providing them movement control of an external robotic arm. Now, when discussing about robot control it is important to reduce the visual-split that different input and output modalities carry. Therefore, Mixed-Reality is could ease communication between humans and robotic systems.

KEYWORDS

Mixed-Reality, Robot Control, Human-Robot Collaboration, Severe Motor Impaired.

1 INTRODUCTION

Science fiction has encouraged people to imagine a world where technology enhances human capabilities. Technological advances in multiple areas, e.g. sensor technologies, artificial intelligence and display technologies, allow researchers and practitioners alike to find new ways to amplify and augment human perception, physiology and cognition [1]. The field of Human-Robot Interaction (HRI) is on the forefront of these developments, as it is concerned with all these aspects.

We venture into HRI through a research project that aims to empower people with severe motor impairments, e.g., tetraplegics, to participate in activities of work and daily life in a self-determined way (see Fig. 1). Due to the limb movement constraint from this user group, we have limited human input modalities available in such a HRI scenario. Further, our users are unable to quickly move themselves around in physical space. That is why we consider robot technology that provides semi-autonomous functionalities. It may require the user to sometimes intervene in autonomous tasks, e.g., in case of operational failures coming from the robot's side. Also, it may require robot operators to completely overtake control for certain tasks.

In this context, our research explores the continuum between teleoperation and peer-to-peer collaboration [2]. We find Mixed-Reality (MR) to be of particular interest, as prior research has often focused on enhancing the robot's perception with sensor technologies [4, 5], but left the human with traditional input and output devices. MR has the capability to provide completely new ways of how humans perceive the interaction with the robot, particular in form of head-mounted Augmented-Reality (AR) glasses such as Microsoft HoloLens.

One of the main challenges we see in HRI, from a human perspective, is the general need of a mediatory interface. While our daily interaction with technology has gotten more direct in the last years to the extent that often input and output interaction space have merged, e.g. in the context of touch screens, this is not the case of HRI. Here, such mediatory interfaces often are disruptive in a sense that it is not possible for the user to keep both the interface and the robot conducting the task in focus. Commercial robot systems normally come with an external touch screen as operating device, requiring the user to shift attention between the robot and the device [6]. From our observations, this cannot only limit efficiency and effectiveness while performing tasks but also impose security

*Produces the permission block, and copyright information
†The full version of the author's guide is available as acmart.pdf document

VAM-HRI, March 2018, Chicago, Illinois USA

risks in the context of human-robot collaborative work environments.

In this paper, we would like to present our understanding of the design opportunities and challenges for MR in Human-Robot Interaction. In doing so, we will keep a focus on our specific research scenario of supporting people with severe motor impairments.

Figure 1: Tetraplegic user operating a Universal Robots UR5 robot arm with MARG head-mounted sensors [3].

2 RELATED WORK

One of the first applications of AR in Human-Robot control relates to teleoperation. Milgram et al. [7] emphasize the use of AR for "achieving spatial human-robot communication". This provides the following advantages: expanding the visualization of the environment, serving as a mediator between humans and a robotic system, enabling action optimization, and warning the user of operational failures.

The use of MR in a HRI context has become more popular during the last years. One of the most explored areas within mixed and augmented reality is remote robot control. For instance, enabling decision making based on the robot's affordances by providing spatial information through a virtual environment augmented with real video information [8].

Kobashi et al. [9] consider the use of MR in the development of a humanoid robot that plans its actions in a virtual environment and interacts with the real one. It provides the person who is controlling the robot an idea of possible results, facilitating then the controller's decision-making process. In addition, trying to achieve better information exchange when using robots for object manipulation. Frank et al. [10] propose mixed reality interfaces to enhance communication of spatial information with robots for object manipulation.

A different approach that could be discussed in the context of robot control is the use of drone technology. Drones have allowed people to explore areas that are hard to reach or present dangers to humans. In a way, it has enabled humans to see through a robotic system's eyes. Artemiadis [11] goes beyond the fact of just seeing the world and proposes the control of multiple devices, by the use of Brain-Machine Interfaces (BMIs), in cases where the use of joysticks is not feasible. This leaves room to explore new interaction interfaces to control robotic systems.

2 DESIGN OPPORTUNITIES IN USING MIXED-REALITY FOR HUMAN-ROBOT INTERACTION

As we briefly discussed in the introduction, we see one of the main opportunities in using MR for HRI in the possibility to bridge the gap between input and output interaction spaces. To be more precise, it foremost allows us to have one unified output space, meaning the intermediary interface is merged with the real-world view of the robot and the task. In a MR setting we are able to augment the real-world view of the environment, including the robot and task. We can provide feedback and interaction robot controls, thereby providing digital visual information in the operator's line of sight. Further, the input space is linked closely to this newly unified output space – for example typical hand gestures when using a Microsoft HoloLens must be performed in the line of sight. In our scenario, hand gestures are not applicable. Therefore, we are considering gaze-based or head movement interaction with an AR head mounted display. Those input displays are closely linked to the output space and provide a direct interaction style. Further, we consider that by bringing input and output space closer together, awareness of actions will be improved, decision taking should be facilitated and cognitive load caused by divided attention can be reduced.

A second opportunity arises from a problem we came across in our prior research in this field, which is concerned with decision making. People have a hard time understanding and correctly predicting the movement of industrial robot arms, as those do not suffer the same limitations as humans' joints. In a supervisory control setup, this may lead to situations in which the operator decides to take action although the robot is following a certain task correctly. In a MR setting, we think that such situations could be reduced by providing visual movement predictions in real time. For instance, displaying superimposed images showing the trajectory to reach an object. Further, it is crucial that such information is presented in the line of sight of the operator, as they can thereby judge if

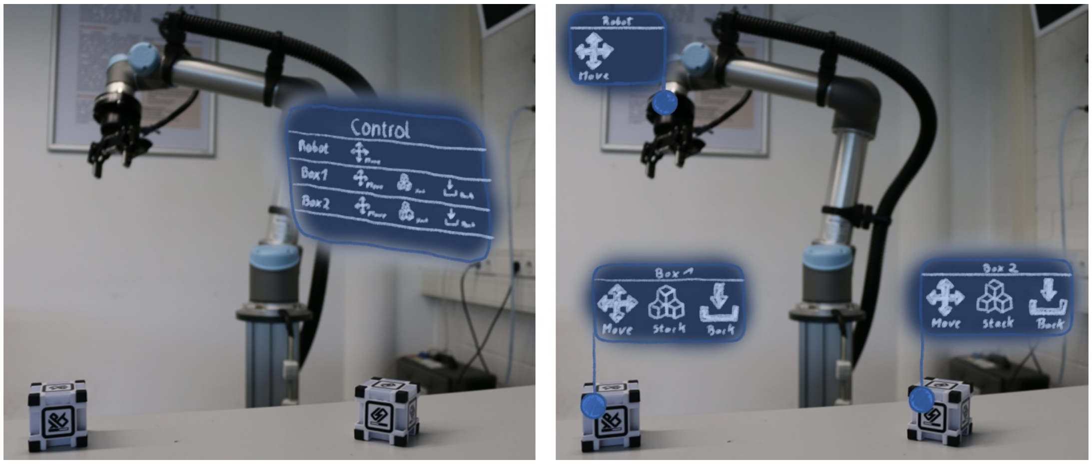


Figure 2: Variations of the interface (design mock-up). Left: Centralized interface as a holographic window next to the objects. Right: Distributed interface with an option panel over the objects

the information is correct and the chances of missing a target object because of a split attention is minimized.

3 DESIGN CHALLENGES IN USING MIXED-REALITY FOR HRI

For all these opportunities, we also identified several design challenges which arise by using MR in the context of robot control.

A first important challenge is the specific design of the user interface in a MR approach. As already mentioned, MR enables to present information and possible actions in the line of sight of the user. However, how these are organized and laid out and how that impacts for example the usability is not yet clear. We see at least three possible approaches for a User Interface (UI) design: (1) a centralized "flat" interface could provide feedback information and controls, similar to current external mediatory interfaces, (2) by making use of the depth tracking capabilities of modern AR glasses such as the Microsoft HoloLens, interface controls and feedback components could be augmenting the real-world objects in 3D space. We call this a distributed interface (See Fig. 2), (3) another option would be to take the "drone control" approach and putting the user in the perspective of the robot arm. Here, an MR approach closer to the VR spectrum of the continuum might be necessary. We call this an egocentric interface.

The centralized interface is closest to the kind of interfaces people are used to from interacting with today's technologies but is limited in the way it capitalizes on the possibilities of MR. The distributed interface makes especially use of these, while raising the challenge of efficiency as actions and feedback might be spread out and outside of our current attention focus. The egocentric interface allows a direct way of interaction as it provides an egocentric view of the environment out of the "robots' eyes." This approach presents some challenges and could not always be effective, e.g., an observer perspective provides more contextual information and creates a sense of agency.

Another challenge, especially in our specific scenario with severe motor impaired users, is the integration of input techniques. As already mentioned, hand gestures are not applicable. Thus, we have explored three approaches: (1) MARG or IMU sensors which make use of the head position and thereby allow head movements as input commands [3], and the Microsoft HoloLens which already has this type of input integrated, (2) speech interaction as an input modality, which has reached new levels with the success of speech assistants such as Apple, Siri, or Google Assistant [12, 13], (3) gaze-based interaction, making use of eye-gaze movements and fixations. The latter has been to allow people with motor impairments to interact with computers. However, it cannot be generalized that the use of one modality suits all use cases. That is why, we consider the use of different modalities and the combination of these by using multi-modal approaches.

When combining input and output as presented, another challenge arises. Should we only interact with digital UI elements or can some interaction be done implicitly by tracking the "interaction" with the real-world? The reason behind this lays on the Degrees of Freedom (DoF) of a robotic arm, e.g., the UR5 has 7 DoF including the grip. With head movements, we are only able to address 3 DoF directly and the challenges increase when consider multiple robotic arms in more complex manufacturing scenarios. This then means that even for just moving the robot arm around in a remote- control scenario, we would have to switch between different operating modes. That is why we find relevant to explore what we call implicit mode switching, i.e., using gaze to point at one specific robot arm and start the interaction with no need to explicitly switch between robot arms via a centralized UI.

We considered two types of gaze-based interactions as input modalities: fixation and smooth-pursuit [14]. Using fixation, the user can interact and control elements by dwelling at an object for a specific time. While this is generalizable to most standard UI controls, it requires a high precision tracking and asks the user to explicitly "hold still", which is not a natural viewing behavior. Through smooth-pursuit the user instead follows a small target with their gaze in order to invoke certain controls. Each such control must have their own target or stimuli element with a unique movement pattern. To identify the desired action, the system will correlate the user's gaze with the trajectory of potential stimuli, with no need for precise and calibrated gaze tracking.

3 CONCLUSIONS

In summary, in our research we explore the opportunities and challenges of using MR in human-robot collaboration. Our focus is to bring input and output modalities closer together. We aim to present information and feedback in the user's line of sight and even context-specific by augmenting the real world environment in a 3D space.

We seek for interaction opportunities for people with severe motor impairments by using MR technology. We propose 3 different types of UIs (centralized, distributed and egocentric) that merge input and output modalities which can be used in different scenarios. This also uncovers some challenges regarding the input modalities. Here, we consider integrating gaze, speech and MARG sensors to achieve a wider spectrum of use cases.

This research will serve as a base for future empirical studies and gather feedback about our design criteria. Further, we think that many of our observations might be generalizable to the general context of MR and HRI.

ACKNOWLEDGMENTS

This research is supported by the German Federal Ministry of Education and Research (BMBF).

REFERENCES

[1] A. Schmidt, S. Schneegass, K. Kunze, J. Rekimoto, and W. Woo, "Workshop on Amplification and Augmentation of Human Perception," 2017, pp. 668–673.

[2] M. A. Goodrich and A. C. Schultz, "Human-Robot Interaction: A Survey," Found. Trends Hum.-Comput. Interact., vol. 1, no. 3, pp. 203–275, 2007.

[3] N. Rudigkeit, M. Gebhard, and A. Gräser, "Evaluation of control modes for head motion-based control with motion sensors," in 2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings, 2015, pp. 135–140.

[4] T. Matsuda, Robot Vision : New Research. New York: Nova Science Publishers, Inc, 2009.

[5] T. Peynot, S. Monteiro, A. Kelly, and M. Devy, "Editorial: Special issue on Alternative Sensing Techniques for Robot Perception," J. Field Robot., vol. 32, no. 1, pp. 1–2, Jan. 2015.

[6] "KUKA smartPAD," KUKA AG. [Online]. Available: https://www.kuka.com/en-gb/products/robotics-systems/robot-controllers/smartpad. [Accessed: 07-Feb-2018].

[7] P. Milgram, S. Zhai, D. Drascic, and J. Grodski, "Applications of augmented reality for human-robot communication," 1993, vol. 3, pp. 1467–1472.

[8] C. W. Nielsen, M. A. Goodrich, and R. W. Ricks, "Ecological Interfaces for Improving Mobile Robot Teleoperation," IEEE Trans. Robot., vol. 23, no. 5, pp. 927–941, Oct. 2007.

[9] K. Kobayashi, K. Nishiwaki, S. Uchiyama, H. Yamamoto, S. Kagami, and T. Kanade, "Overlay What Humanoid Robot Perceives and Thinks to the Real-world by Mixed Reality System," in Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality,Washington, DC, USA, 2007, pp. 1–2.

[10] J. A. Frank, M. Moorhead, and V. Kapila, "Mobile Mixed-Reality Interfaces That Enhance Human–Robot Interaction in Shared Spaces," Front. Robot. AI, vol. 4, 2017.

[11] P. Artemiadis, "Brain-Swarm Control Interfaces: The Transition from Controlling One Robot to a Swarm of Robots," Adv. Robot. Autom., vol. 06, no. 01, 2017.

[12] "iOS - Siri," Apple. [Online]. Available: https://www.apple.com/ios/siri/. [Accessed: 08-Feb-2018].

[13] "Google Assistant - Your own personal Google," Google Assistant - Your own personal Google. [Online]. Available: https://assistant.google.com/. [Accessed: 08-Feb-2018].

[14] A. Esteves, D. Verweij, L. Suraiya, R. Islam, Y. Lee, and I. Oakley, "SmoothMoves: Smooth Pursuits Head Movements for Augmented Reality," 2017, pp. 167–178.

7.5 Paper V. There's More than Meets the Eye: Enhancing Robot Control through Augmented Visual Cues.

Arévalo Arboleda, Stephanie; Dierks, Tim; Rücker, Franziska; and Gerken, Jens. 2020. There's More than Meets the Eye: Enhancing Robot Control through Augmented Visual Cues. In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20). Association for Computing Machinery, New York, NY, USA, 104–106. DOI:10.1145/3371382.3378240.

There's More than Meets the Eye: Enhancing Robot Control through Augmented Visual Cues

Stephanie Arévalo Arboleda, Tim Dierks, Franziska Rücker, Jens Gerken
Westphalian University of Applied Sciences
Gelsenkirchen, Germany
{stephanie.arevalo,tim.dierks,franziska.ruecker,jens.gerken}@w-hs.de

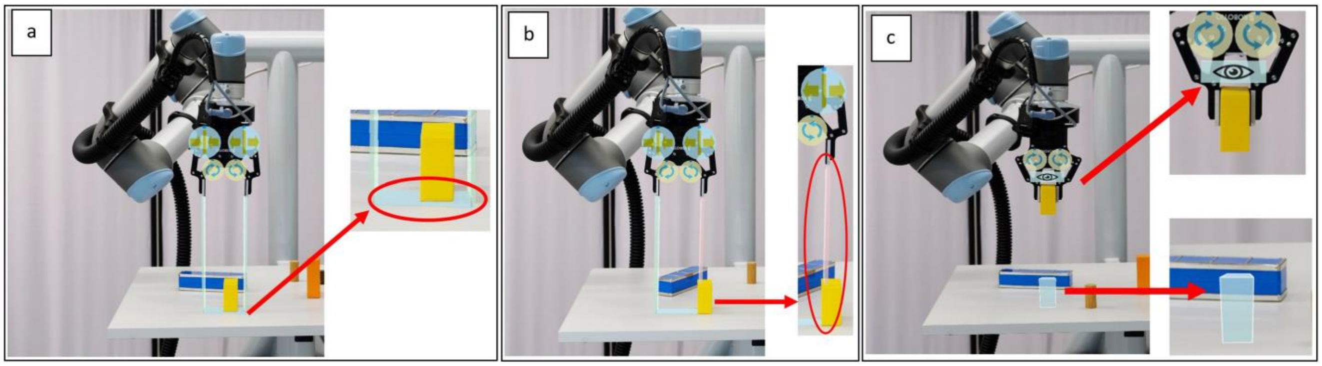


Figure 1: Manipulation and Grasping using augmented visual cues: a. Virtual extensions of the gripper. b. Collision of a grip- per's finger with an object. c. Projection of a grasped object

ABSTRACT

In this paper, we present the design of a visual feedback mechanism using Augmented Reality, which we call augmented visual cues, to assist pick-and-place tasks during robot control. We propose to augment the robot operator's visual space in order to avoid attention splitting and increase situational awareness (SA). In particular, we aim to improve on the SA concepts of perception, comprehension, and projection as well as the overall task performance. For that, we built upon the interaction design paradigm proposed by Walker et al.. On the one hand, our design augments the robot to support picking-tasks; and, on the other hand, we augment the environment to support placing-tasks. We evaluated our design in a first user study, and results point to specific design aspects that need improvement while showing promise for the overall approach, in particular regarding user satisfaction and certain SA concepts.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; **Visualization design and evaluation methods**.

HRI '20 Companion, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-7057-8/20/03.
https://doi.org/10.1145/3371382.3378240

KEYWORDS

Visualization, Mixed Reality User Interfaces, Robot Control, Human-Robot Interaction, Situation Awareness

ACM Reference Format:

Stephanie Arévalo Arboleda, Tim Dierks, Franziska Rücker, Jens Gerken. 2020. There's More than Meets the Eye: Enhancing Robot Control through Augmented Visual Cues. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20 Companion), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3371382.3378240

1 INTRODUCTION

Visual relationships are part of human-object interactions [8], where depth perception is the key to determine how far away objects are located [9]. When objects are located in a distant or remote area, understanding the space, distance, and object affordances can become difficult. In robot teleoperation, the robot operator can be detached from the scene where manipulation actions take place, affecting natural perception and Situation Awareness (SA)[2]. SA has long been a topic of concern [7] [14] in robot teleoperation and it can be described through the three Endsley's levels [3]: perception (size, location, colors, dynamics), comprehension — understanding the relationship between objects and its significance within the environment, and projection of future status — given by the ability to foresee future actions or states. We consider that these issues might not only be present when the robot and the operator are situated in two distant spaces, but in a co-located space as well. One possible way to address the aforementioned SA issues is by enriching the visual space. There has been an increasing body of research using Augmented Reality (AR) as a counterpart in robot control [13] [6].

AR in particular can facilitate also co-located robot control, bringing new opportunities for merging the virtual and physical interaction spaces. Examples from research include using AR for finding ways to visualize robot motion in trajectory planning [5], collision avoidance [1], or using visual cues to highlight important areas of a scene [11]. In our work, we take a closer look at pick-and-place scenarios, which are the foundation to many more complex operations, e.g., surgery scenarios or building pieces with high precision. We aim to explore the design space through the use of augmented visual cues. Contrary to [11], we do not aim to guide attention but improve the operator's perception, comprehension and projection of target objects and eventually increase task performance.

2 VISUAL CUES DESIGN

Walker et al. [12] propose a design framework that considers elements from AR for a user-centered HRI, which relates to the manner how additional information is presented to the user: "augmenting the environment", "augmenting the robot", and "augmenting the user interface". We place our design along the lines of Walker et al. framework by "augmenting the robot" in order to assist the operator in picking-tasks, and "augmenting the environment" to support during placing-tasks.

2.1 Augmenting the Robot for Picking

We aim to provide a better understanding of the position of the robot's gripper relative to the object to-be grasped. To enable this, we project virtually the gripper's fingers through a beam, until it reaches a contact surface. This allows the operator to determine if the object to be grasped is within reach, and adjust the grip region to the size of the object. Also, we add a reflection of the grip region presented through a green beam over the contact surface, this helps to evaluate if an object fits inside the grip region (Figure 1a). An additional visual cue changes the color of the beam to red (Figure 1b). This alerts the operator that one of the gripper's fingers is intersecting with some part of the object to be grasped – which is achieved by using object recognition. These visual cues are designed to help tackling the SA problems of perception and comprehension when grasping an object.

2.2 Augmenting the Environment for Placing

Our design aims to show the future status of an object when trying to place it on a specific location. To enable this, once an object has been grasped, we show a projection of the object (ghost object) in gray, which is displayed at the position where the operator is currently pointing (Figure 1c). This allows the operator not only to see where the object would be placed but evaluate its spatial placement in a determined area. We also add a color code, similar to the one in the picking-task. The ghost object turns green when in perfect alignment (orientation and position) with a target position, and it turns red when any part of the object is outside of that position. Through these visual cues that augment the environment, we aim to support the SA problem of projection.

3 USER STUDY

We developed a virtual co-located space with a robot arm to perform a pick-and-place task in Unity 2018.2.18f1 and used it on the Microsoft Hololens. We recruited 10 participants and compared the execution of a peg-in-hole task with three different shaped objects, each presenting an incremental level of difficulty (T1, T2, T3) using both augmented visual cues and without them. Users operated the robot through a combination of voice commands and head-gaze pointing as provided via the HoloLens. For example, to move the gripper to a certain location in space, users pointed to that and said "move". We wanted to analyze how our solution impacts SA problems, task performance and understand the users' perception of our design. For that, we measured the number of actions (number of voice commands for move, rotate, open/close griper) to execute each sub-task (picking and placing), the task success rate and applied the USE questionnaire [10] for subjective feedback. Also, we assessed SA problems (perception, comprehension and projection) through a SA Global Assessment Technique (SAGAT) [4].

3.1 Results

For picking sub-tasks, we did not find a difference regarding the average number of actions: all participants successfully grabbed objects both with and without the augmented visual cues. Similarly, SAGAT scores show no significant differences and were overall very high (above 80%). The USE questionnaire, conversely, revealed positive opinions, especially in terms of certainty. One reason for these results could be that the overall task complexity did not demand additional visual cues to successfully picking objects. In addition, the visual cues lead to some visual clutter that may have had negative effects on usefulness. For placing sub-tasks, we found a significant difference for T2 and T3 regarding the number of actions in favor of the "no cues" condition (T2: $p=0.01$, $t(9)=2.92$, M(Cues)=3.7 (SD=2.45) vs. M(noCues)=1.3 (SD=0.46); T3: $p=0.01$, $t(9)=3.07$, M(Cues)=5.7 (SD=2.1) vs M(noCues)=3.7 (SD=1.62)). However, the "visual cues" condition resulted in a higher success rate of 65% on average (SD=0.25) compared to 30% on average (SD=0.29), showing a significant difference ($p=0.01$, $t(9)=-2.97$). While we would have anticipated that visual cues might decrease the number of actions it seems that the increase is necessary to improve the success rate, i.e. users where able to make adjustments based on the richer feedback. Regarding SAGAT scores, we found a significant difference favoring the "visual cues" condition ($p=0.03$, $t(9)=2.44$) for the projection level, which reached the maximum score of 100% (SD=0) vs the "no cues" condition with 86% (SD=0.17). Results from the USE questionnaire showed mixed opinions for the different elements of our design, however 9/10 highlighted its usefulness. Therefore, we think that the approach does show it can be effective.

4 CONCLUSION AND FUTURE WORK

We conclude that in future work, we have to redesign and decrease the visual clutter to increase the effectiveness of the visual cues for picking. For placing, we will focus on the ghost object as main visual cue to examine further. We will also take into account more complex difficulty levels to understand how and when visual cues may be able to make a difference.

ACKNOWLEDGMENTS

This research is supported by the German Federal Ministry of Education and Research (BMBF).

REFERENCES

[1] J.W.S. Chong, S. K. Ong, A.Y.C. Nee, and K. Youcef-Youmi. 2009. Robot programming using augmented reality: An interactive method for planning collision-free paths. *Robotics and Computer-Integrated Manufacturing* 25, 3 (2009), 689–701. https://doi.org/10.1016/j.rcim.2008.05.002

[2] J. L. Drury, J. Scholtz, and H. A. Yanco. 2003. Awareness in human-robot interactions. In *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483)*, Vol. 1. IEEE, New York, New York, USA, 912–918 vol.1. https://doi.org/10.1109/ICSMC.2003.1243931

[3] Mica R. Endsley. 1995. Measurement of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1 (1995), 65–84. https://doi.org/10.1518/001872095779049499

[4] Mica R. Endsley. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1 (1995), 32–64. https://doi.org/10.1518/001872095779049543

[5] H. C. Fang, S. K. Ong, and A.Y.C. Nee. 2012. Interactive robot trajectory planning and simulation using Augmented Reality. *Robotics and Computer-Integrated Manufacturing* 28, 2 (2012), 227–237. https://doi.org/10.1016/j.rcim.2011.09.003

[6] Markus Funk, Andreas Bächler, Liane Bächler, Thomas Kosch, Thomas Heidenreich, and Albrecht Schmidt. 2017. Working with Augmented Reality?. In *PETRA 2017 (ICPS: ACM international conference proceeding series)*, Unknown (Ed.). ACM, New York, NY, USA, 222–229. https://doi.org/10.1145/3056540.3056548

[7] Yiannis Gatsoulis, Gurvinder S. Virk, and Abbas A. Dehghani-Sanij. 2010. On the Measurement of Situation Awareness for Effective Human-Robot Interaction in Teleoperated Systems. *Journal of Cognitive Engineering and Decision Making* 4, 1 (2010), 69–98. https://doi.org/10.1518/155534310X495591

[8] G. Gkioxari, R. Girshick, P. Dollár, and K. He. 2018. Detecting and Recognizing Human-Object Interactions. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, New York, New York, USA, 8359–8367. https://doi.org/10.1109/CVPR.2018.00872

[9] Ian P. Howard and Brian J. Rogers. 2012. *Perceiving in depth*. Oxford psychology series, Vol. 29. Oxford University Press, New York.

[10] Arnold Lund. 2001. Measuring usability with the use questionnaire. *Usability Interface* 8, 2 (2001), 3–6.

[11] Daniel J. Rea, Stela H. Seo, Neil Bruce, and James E. Young. 2017. Movers, Shakers, and Those Who Stand Still: Visual Attention-Grabbing Techniques in Robot Teleoperation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. Association for Computing Machinery, New York, NY, USA, 398–407. https://doi.org/10.1145/2909824.3020246

[12] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. 2018. Communicating Robot Motion Intent with Augmented Reality. In *HRI'18*, Takayuki Kanda, Selma Ŝabanović, Guy Hoffman, and Adriana Tapus (Eds.). Association for Computing Machinery, New York, New York, 316–324. https://doi.org/10.1145/3171221.3171253

[13] Rong Wen, Wei-Liang Tay, Binh P. Nguyen, Chin-Boon Chng, and Chee-Kong Chui. 2014. Hand gesture guided robot-assisted surgery based on a direct augmented reality interface. *Computer methods and programs in biomedicine* 116, 2 (2014), 68–80. https://doi.org/10.1016/j.cmpb.2013.12.018

[14] H. A. Yanco and J. Drury. op. 2004. "Where am i?" acquiring situation awareness using a remote robot platform. In *2004 IEEE international conference on systems, man & cybernetics theme*, Peter Wieringa (Ed.). IEEE, Piscataway (N.J.), 2835–2840. https://doi.org/10.1109/ICSMC.2004.1400762

7.6 Paper VI. 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.

Arévalo Arboleda, Stephanie; Dierks, Tim; Rücker, Franziska; and Gerken, Jens. 2021. 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. Human-Computer Interaction – INTERACT 2021. INTERACT 2021. Lecture Notes in Computer Science, vol 12932. Springer, Cham. DOI: 10.1007/978-3-030-85623-6_31.

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

Stephanie Arévalo Arboleda[0000–0001–5577–4407], Tim Dierks, Franziska Rücker, and Jens Gerken

Westphalian University of Applied Sciences, Gelsenkirchen, Germany
{stephanie.arevalo,tim.dierks, franziska.ruecker, jens.gerken}@w-hs.de

Abstract. Accurate depth perception in co-located teleoperation has the potential to improve task performance in manipulation and grasping tasks. We thus explore the operator's visual space and design visual cues using augmented reality. Our goal is to facilitate the positioning of the gripper above a target object before attempting to grasp it. The designs we propose include a virtual circle (Circle), virtual extensions (Extensions) from the gripper's fingers, and a color matching design using a real colormap with matching colored virtual circles (Colors). We conducted an experiment to evaluate these designs and the influence of distance from the operator to the workspace and the target object's pose. We report on time, success, and perceived certainty in a grasping task. Our results show that a shorter distance leads to higher success, faster grasping time, and higher certainty. Concerning the target object's pose, a clear pose leads to higher success and certainty but interestingly slower task times. Regarding the design of cues, our results reveal that the simplicity of the Circle cue leads to the highest success and outperforms the most complex cue Colors also for task time, while the level of certainty seems to be depending more on the distance than the type of cue. We consider that our results can serve as an initial analysis to further explore these factors both when designing to improve depth perception and within the context of co-located teleoperation.

Keywords: Human-robot interaction · depth perception · augmented reality · robot teleoperation · visual cues · certainty.

1 Introduction

One of the most common tasks in Human-Robot Interaction (HRI) is pick-and-place since it presents a basic interaction for teleoperation of robotic arms. These tasks are common in non-standardized assembly workspaces, where target objects often change in shape and position, which may require the human operator to fine-control (semi-autonomous) robotic arms to succeed in the tasks.

Picking an object can be further segmented into **1) (Coarse) Pointing**, which requires the user to identify the location of a target object and consequently move the gripper of the robotic arm roughly to that location; **2) Positioning**, which requires to align the gripper with the target object to successfully grasp it; and **3) Grasping** then requires the user to move the gripper to the point in space where the object can finally be acquired. While this, at first sight, might sound straightforward, it often requires many trial-and-error attempts from the operator due to misjudgments regarding the position in space of either the gripper or the object. This can lead to precarious situations when handling hazardous material.

The crucial point in this process takes place while positioning the gripper. It might not always be possible for the operator to accurately determine the gripper's position relative to the target object due to the visual perception of distance and object's pose, both factors related to depth perception. Even in co-located scenarios (robot and operator located in the same physical space) operators could have a fixed viewpoint due to physical mobility restrictions, e.g., people with disabilities, or due to environmental limitations, e.g., safety measures due to the manipulation of hazardous materials. As a result, spatial abilities are impeded, hindering the operator to correctly identify the shape, pose, and distance of a target object relative to the gripper and other objects within the workspace.

Augmented Reality (AR) creates opportunities to improve depth perception of real-world objects by enhancing the visual space to provide awareness of their position in the workspace. Some studies present evidence that AR counteracts visual feedback limitations in teleoperation [39], [53],[19]. Some of these limitations are related to the manner in which humans perceive distance and judge depth. Humans perceive visually the environment through different visual cues (monocular, binocular, dynamic) [14]. These cues can be enhanced through AR and thus refine the visual space. Here, it is important that these cues provide a message that can be clearly interpreted by the observer, avoiding unnecessary extra mental processing due to conflicts among cues [29].

In this paper, we contribute to the exploration of the design space to improve depth perception. Therefore, we first propose three designs of visual cues in AR that systematically build on each other. Each one considers different design elements that could foster depth perception in co-located robot teleoperation and provide a foundation for future HRI designs. Second, we present a systematic experiment investigating the relationship between the design of visual cues, the operator's distance, and the target object pose to evaluate the advantages and downsides of each visualization concept. Third, we include the operator's perceived certainty during picking as a novel measure to better understand if and how objective measures of performance such as success (certainty accuracy or effectiveness) and time (efficiency) correlate with the subjective certainty of the operator.

2 Background

We first review theoretical foundations of visual perception regarding distance and depth, the use of AR to improve depth perception in robot teleoperation, and the role of time, success, and certainty in visual perception.

2.1 Visual Perception of Depth

Visual perception can be defined as a complex representation of the visual world, where features of objects in the environment are collected visually to then be interpreted through computations in multiple areas of the brain [49]. In order to visually determine the depth of objects, factors such as size, shape, and distance are determinants to gain an accurate perception of objects in the environment [24]. Brenner & Smeets [9] present the relation between size and distance as straightforward, e.g., smaller sizes provide hints about how distant an object might be, while shape and distance present a more complex relation. Distance can compromise determining the shape of objects, e.g., problems in accurately determining the shape of far-away objects. Further, perceived objects' shape varies depending on the manner that they are positioned (pose). Here, having different perspectives acquired by motion assists with accurately determining an object's shape.

Within the visual space, it is relevant to explain the perceived location. This is understood as the perception of direction and distance from the observer's viewpoint (egocentric distance) or the distance between two external points (exocentric distance) [34]. Related to perceived location, the space layout of a person is determinant for evaluating distance and depth as farther distances compromise the manner how objects in the surroundings are perceived. Cutting& Vishton [12], divide the layout of space around a person into three: personal space (within 2m), typically a stationary working space; action space (from 2m to 30m), which is the moving and public action space; and vista space ($>$ 30m). Different authors agree that humans underestimate egocentric distances (distance relative to the observer) in the real-world [33],[34]. Distance can be evaluated through different measures, e.g., verbal reports, walking around a target, visually matching distances to a familiar one, and blind walking [47]. It can also be estimated through bisection or fractionation, which consists of determining the midpoint of a distance from the observer's perspective to a target, and specify that bisection does not provide absolute measures of egocentric distances [7].

Humans perceive and interpret distance through a set of visual cues: monocular (perceived using one eye), binocular (perceived using both eyes), and dynamic (perceived by movement) [14]. Nonetheless, interpreting all the cues perceived under some environmental conditions is not a straightforward process. Laramee & Ware [31] describe it as an ambiguity solving process, where different cues need to be primed over others to get an accurate picture of the environment (cue dominance). Further, Howards & Rogers [25] highlight that the reliability of cues is determined by cue average, cue dominance, cue specialization, range extension, and probabilistic models.

Rolland et al. [46] already found a relation between object size and distance in virtual objects displayed in ARHMD. In Mixed Reality (MR) environments, this problem can be not only inherited from the real-world but can be exacerbated [26], [40]. Further, El Jaimy & Marsh [14] present a survey on depth perception in head-mounted displays (HMDs) and highlight the importance of evaluating not only depth but also distance perception in virtual and augmented reality environments.

2.2 Designing to improve depth perception in AR Environments

We consider the work of Park & Ha [41] as a keystone to improving depth perception in virtual and mixed reality environments. They provide a classification of visual enhancements techniques that offer spatial information as follows: geometric scaling, understood as the variation of size to provide a distance relationship; symbolic enhancements, which are visual representations that allow creating associations, e.g., a grid surface or ground plane to transfer spatial information; visual cues, which are a combination of monocular cues to improve the perception of depth; a frame of reference that provides a mental model of the environment; and visual momentum, referring to providing perceptual landmarks to reduce visual inconsistency among different displays/scenes. Following this line of research, Cipiloglu et al. [11] grouped a series of methods for depth enhancement in 3D scenes focusing on perspectives, focus, shading and shadows, among others.

Heinrich et al. [21] provided a state-of-the-art overview of visualization techniques using AR for depth perception. They presented different visualization concepts that have been applied to improve depth perception with an emphasis on projective AR without using HMDs. Also, Diaz et al. [13] explored different factors that influence depth perception in AR such as aerial perspective, cast shadows, shading, billboarding, dimensionality, texture, and the interaction of cues. They presented two experiments where they evaluated the effect of these factors in participants' perception of virtual objects' depth relative to real targets. Their results showed that among all their designs, the use of cast shadows improved depth estimation the most. This aligns with other studies [56] that used casting shadows through AR as depth cues. They used pictorial cues (color encoded markers) to provide information about egocentric distance together with shadows and aerial perspective, showing similar results in distance perception.

When aiming to enhance the visual space through AR the work of Kruijff et al. [29] is of particular importance. They identified and classified a series of factors that affect the augmentations related to the environment, capturing, augmentations, display devices, and users. These provide a framework to identify the potential factors that influence depth perception.

2.3 AR for Depth perception in Robot Teleoperation

Previous studies present evidence that AR diminishes visual feedback limitations in teleoperation by providing additional information to the operator [39],

[52],[19]. Presenting cues through AR could enhance the perception of depth and distance in the real world. Choi et al. [10] present reinforcement of the user's cognitive abilities as the purpose of AR. Moreover, the use of AR for robot control has been found to reduce the mental load in robot programmers [48].

Depth perception is still an issue in MR and teleoperation [10] that invites further exploration. Casting shadows using AR has been used in the teleoperation of aerial robots [60], [53] since it has proven to support aerial navigation by improving the spatial relationships between the environment and the aerial robot. Zollman et al. [60] used different techniques that aim to maintain or replicate natural depth cues to design visual hints (waypoints and pathlists) that provide flight-relevant information. In pick-and-place tasks, AR has also been used to present a projection-based AR interface that provides task instructions [18]. Here, shadows have been used to highlight intended target positions and thus provide instructions to operators.

2.4 Certainty, Time, and Accuracy in Visual Perception

Certainty can be defined as a sense of conviction and is considered as a foundation of people's beliefs [42]. In order to achieve "good visual certainty", observers need to consider information that goes beyond monocular, binocular, or motion parallax cues and acknowledge sources of uncertainty, being thus able to predict an outcome [35]. In order to measure the degree of certainty, people need information derived from evidence and time, wherein evidence, in turn, contributes to accuracy [27].

Evidence can be acquired from experience, where certainty plays a pre-and-post-decision-confidence role. Pre-decision confidence relates to the current incoming information and post-decision confidence is the information derived from experience [20]. Also, evidence can be acquired through perceptual evaluation, wherein visual perception has a relevant role. Here, findings of [17] show that the amount of evidence has a positive correlation with accuracy.

Time plays a significant role in relation to attitude certainty. This relation has received special attention in neuroscience. Willis & Todorov [55] shown that a longer period of time allows to form greater impressions of certainty, yet it does not necessarily improve the impression of accuracy. In line with this view, Barden & Petty [4] provided evidence that greater thoughtfulness leads to greater certainty. However, there are controversial results about the influence of time on certainty. Some studies [3],[2] associate longer times with lower certainty. Further, Kiani et al. [27] showed that time alone cannot explain fluctuations in certainty but time together with difficulty have a critical role in certainty.

In visual perception, accuracy is bound to the ability to make a good judgment of the visual stimulus, while certainty relates to the ability to make a good judgment on the validity of a perceptual decision [35]. Accuracy has been suggested to be connected to the amount of evidence from the environment that can be collected [44]. Perceived visual certainty can be dissociated from accuracy [35], and adding to that, a line of research suggests that time can also be disassociated from accuracy [27]. Visual certainty is a topic with controversial findings that has several factors that influence it. Barthelme & Mamassian [5] found that visual uncertainty predicts objective uncertainty. Interestingly, Gardelle & Mamassian [16] showed that subjective uncertainty is abstract and task-independent. They compared identical and perceptual different task-trials in succession in a visual discrimination task, e.g. the orientation of a bar, and found no difference between conditions.

3 Exploration of Visual Cues for Co-located Teleoperation

Building on the research mentioned in Sections 2.2 and 2.3, we present three designs of visual cues to help operators to evaluate distance and depth which in turn could allow to better estimate the position of the gripper relative to a target object for successful grasping. In particular, our designs of visual cues capitalize on previous findings of the effectiveness of cast and drop shadows [13], symbolic enhancements [41], and matching physical and virtual landmarks [60] to provide better awareness of the position of the gripper on the workspace. While some related work has focused on improving depth perception of virtual objects and their interaction in the real-world, we focus our designs on using virtual elements to improve the depth perception of real-world objects. Consequently, all our designs augment the environment through AR using the Microsoft HoloLens. As can be depicted in Fig. 1, each of our designs builds upon the previous one. Thereby, the Extensions include the Circle visual cue and the Colors include both, the Extensions and Circle.

3.1 Virtual Circle using Cast Shadows (Circle)

In this design, we provide a virtual cue derived from the real-world (the gripper) which in turn acts on the physical world (workspace). We cast the real grip region as a virtual representation of it through a virtual circle. This virtual circle's diameter matches the width of the grip region and is shown right under the real gripper on the workspace, see Fig. 1a. We base our design on previous findings of the efficacy of cast shadows to improve depth perception [13]. Cast shadows can be described as the shadow of an object that is reflected on a different surface [59]. This first design of visual cue is minimalistic and intends to provide a simple and comprehensible hint of the location of the gripper on the workspace.

3.2 Virtual Extensions as a Symbolic Enhancement (Extensions)

In this design, in addition to the Circle, we virtually extend an object from the physical world (the gripper). Based on previous work about symbolic enhancements [41] and considering Walker et al.'s framework for AR in HRI [52], we augment the robot's end-effector, through a virtual elongation of the gripper's

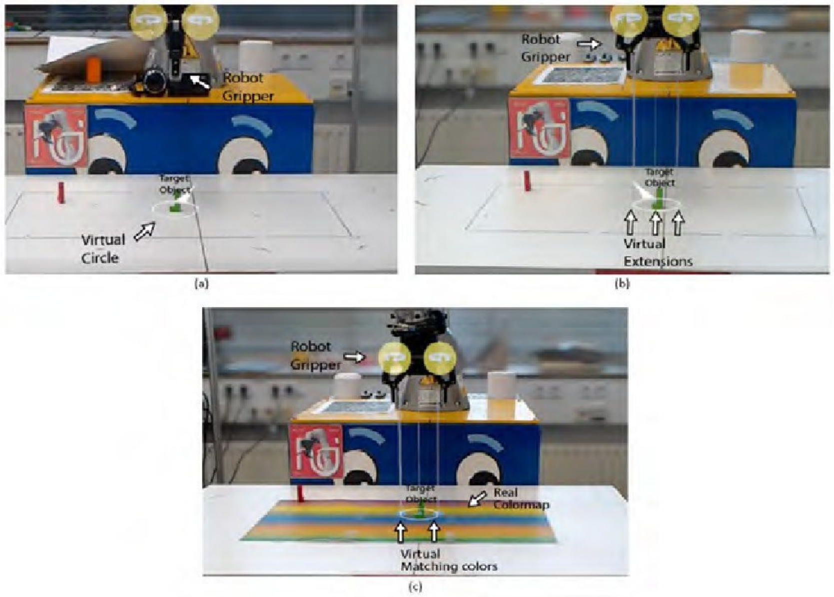


Fig. 1. (a) Circle. (b) Extensions. (c) Colors: Physical color map with matching virtual circles at the end of the Extensions.

fingers. We present three virtual elongations, one at the tip of each finger, and another one rendered at the center of the grip region, see Fig. 1b. These virtual extensions are designed to allow for a better visualization of the exact gripper position on the workspace. Through these visualizations, the operator can determine if the target object is within reach or if there are any potential collisions, e.g., between the fingers and some surface of the target object or other elements in the environment.

3.3 Mixed Reality Color Gradient using Physical Landmarks (Colors)

Building on the Circle and Extensions, the Colors cues aim to provide a connection to the real-world by matching physical landmarks with virtual ones. This could potentially improve the mapping and alignment between the augmentation and the real-world objects as it provides more information about the spatial arrangement of the environment. Here, we take into account the work of Zollman et al. [60], who explored this connection in flying aerial robots. We considered these findings and applied them to our scenario by providing a real physical landmark that connect the gripper and the workspace.

Our landmark is a physical colormap. It presents 5 different color gradients (6 cm per stripe) that cover the workspace, signaling incremental depth, see Fig. 1c. Our reasoning behind the use of a colormap comes from the problems that have been found of determining with precision a position in monochromatic surfaces, e.g. previous research showed that the human eye loses depth cues on uniform surfaces [29]. Further, the use of colors has proven to be effective to signal depth [56]. Through a colormap, the operator can identify a color in the workspace where a target object is located and then position the gripper with greater precision above it. To facilitate pointing at the target area of the workspace, the cursor also adopts the color of the area that is being pointed at. Additionally, we added small colored circles at the end of the virtual extensions, which adopt the color of the current area.

3.4 Interaction Design for Teleoperation

This research is part of a larger project that takes a closer look at hands-free multimodal interaction. Our interaction design consists of head orientation to point, head yaw for positioning, and voice commands to commit an action. Previous studies have shown that the use of speech and head movements are an intuitive interaction concept for human-robot collaboration in pick-and-place tasks [30]. Further, these modalities are natively supported by the HoloLens, and are common for MR environments. The initial and resting position of the robotic arm can be seen in Fig. 1 and the interaction is explained as follows:

Pointing. The operator sees a white head pointer which moves with the operator's head movements. Once an intended position has been determined, the operator gazes at that position and commits the action with the command "Move". After the command has been recognized, the head pointer turns green in a "pie timer" manner for one second. During this time, the operator can decide to keep (remain still) or modify (change head position) the pointer's position before the gripper starts moving to the selected position on the x,y plane. This helps to counter slight inaccuracies that can result from unintended head movements which can happen while uttering the voice command.

Positioning. Once the gripper is located over the desired position. We placed a set of virtual buttons anchored to the real gripper. These virtual buttons rotate the gripper to the left or right and allow the operator to ensure that the fingers' grasping points match the affordances of the object, see Fig. 1. In addition, the operator can activate a fine control mode through the command "Precision". This fine control is executed through the operator's head yaw, where the gripper moves slowly following the head's yaw. In this mode, we display (on the edges of the circle) an arrow to indicate the direction that the gripper is moving towards.

Grasping. When the operator has determined that the gripper is located at an adequate grasping position, the command "Pick" commits the action of grasping, i.e., the gripper moves with the fingers opened along the z-axis towards the object, stops at 0.5 cm above the tabletop, closes the fingers, and moves back up to the resting position.

4 Study

We present a study with 24 participants that aims at exploring the relationship between the different designs of cues (Circle, Extensions, Colors) and depth-related variables distance (2m and 3m) and object pose (clear, ambiguous) when teleoperating a co-located robotic arm. In particular, we looked at measures of effectiveness (success), efficiency (time), and perceived certainty. In the study, participants wore the Microsoft HoloLens to visualize our designs of visual cues while teleoperating a robotic arm.

4.1 Hypotheses

Distance perception of the bare eye has an imminent effect on depth perception. Based on this, we hypothesize that our designs of visual cues will reduce the impact of distance (H1). Similarly, objects' pose can reduce depth perception, we thus hypothesize that our visual cues will reduce the impact of pose (H2). Based on the fact that the individual designs of each visual cue builds systematically on each other, subsequently adding more depth information, we expect a step-wise rise between each of them. This means that the performance with Circle will be exceeded by Extensions and Extensions by Colors for the evaluated metrics (H3). Further, building upon visual perception and certainty, we designed our visual cues to provide more information (evidence) about the workspace, which will in turn show a correlation between these two metrics (H4). Specifically, we hypothesize that:

H1. Our designs of visual cues will lead to similar results in grasping time, success, and certainty at 2m and 3m.

H2. Our design of visual cues will lead to similar results in success and time for the clear and ambiguous pose, but with an incremental degree of certainty for the clear pose.

H3. Colors will lead to a higher success rate, shorter execution time, and higher perceived certainty compared to Extensions and Circle. Similarly, the Extensions would perform better on those metrics compared to the Circle alone.

H4. Our results will point to a positive correlation between certainty and success rate.

4.2 Participants

We recruited 24 participants among students and university staff with an average age of 28.67 (SD = 7.72). The pre-test questionnaire revealed that 6 participants had previous experience using the Microsoft HoloLens 1 and 11 participants had some experience with robots (not necessarily a robotic arm). One participant reported having red-green colorblindness but did not experience problems in distinguishing the colors used. Participants received 7 euros for their participation.

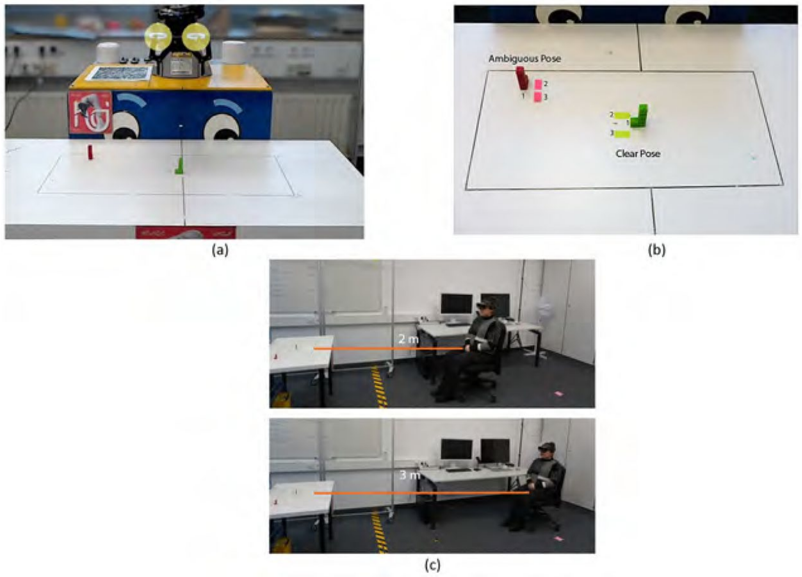


Fig. 2. (a) Experiment set-up for Circle and Extensions. (b) Upper view from the workspace with the target objects. (c) An operator located at different distances from the workspace.

4.3 Study Design

We designed a full factorial 2x2x3 within-subject experiment (2 distances x 2 poses x 3 designs of visual cues) where participants executed 3 trials per condition, yielding 36 trials per participant and 864 trials in total. The conditions were counter-balanced through a Latin-square design.

Our independent variables were the type of visual cue (Circle, Extensions, Colors), distance (2m, 3m), and pose (clear, ambiguous). For distance, we varied the egocentric space between the workspace and the stationary position of the operator (Fig. 2c). We chose 2m and 3m as distances to represent realistic co-located scenarios covering the egocentric personal (2m) and action space (3m) respectively. We did not vary the size of the object due to kinetic invariance–

Different poses were achieved by manipulating the position and orientation of an L-shaped target object. To differentiate the poses, we used two objects of the exact same size and shape but with different colors. A green object was deemed as the "clear" pose. It was located roughly in the center of the workspace and oriented in such a way that the shape was fully visible. A red object was deemed as the "ambiguous" pose. It was located in the left corner, almost at the border of the workspace and oriented in such a way that participants could only see one side of the object. We considered this pose ambiguous since it required participants to perform mental rotations (see Fig. 2).

We collected objective and subjective measures. As objective measures, we considered time, measured in seconds, and success, measured as a binary value. Our measure of time relates to the amount of time that participants took to position and align the gripper above the target object until invoking the picking operation—referring to positioning, the second step of the picking interaction. For each trial, we measured whether or not participants succeeded in grasping the target object (success). As a subjective measure, we prompted participants with the following question directly in the HoloLens whenever they invoked the picking operation: "In this position, how certain (in %) are you that you can grasp the object". The percentages were provided as choices on a 7-point scale. Tormala [50] argues for a 7-point scale to measure attitude certainty, as in certain circumstances extreme attitudes can be held with less certainty. Also, qualitative data were collected through a short interview at the end of the study.

We used a mixed-methods approach to evaluate our data. The data were analyzed using RM-ANOVA for time and certainty, as a normal distribution of data could be assumed through the means of Shapiro-Wilk test. For pairwise comparisons (through post-hoc Estimated Marginal Means), we applied Bonferroni corrections to control for Type I errors. For the analysis of the binary variable grasping success, we could not assume a normal distribution and therefore opted for GEE (Generalized Estimating Equations). GEE accounts for correlations within-participants in repeated measures designs and has been commonly used for binary outcome variables [57], [32]. To investigate potential associations between our dependent variables (certainty and success/time), we applied Spearman's partial correlation controlling for the effect of our independent variables.

4.4 Task

Our main task simulated a grasping task inspired from manufacturing workspaces. The task comprised of teleoperating the robotic arm to grasp an "L-shaped" object (4x1x2)cm placed at two different positions and with the different poses on the workspace. The operator performed this task at 2m and 3m from the workspace, see Fig. 2c. Participants were instructed to remain seated in a comfortable position and avoid movements to the left or right or tilt their heads to change their visual perspective. This instruction helped to evaluate the effects of distance and pose using our designs of visual cues on depth perception.

We performed a pilot test with 3 users to identify potential factors that may affect our experiment. We discovered learning effects due to the position invariance of the target objects during trials. Thus, we moved each object 1 cm apart from the last position for each trial for the experiment, see Fig. 2b.

4.5 Procedure

The experiment took approximately 60 minutes divided into (1) introduction, (2) calibration, (3) training (4) task, and (5) post-test questionnaires and interview.

(1) During the introduction participants were handed a standardized consent form and pre-test questionnaire. They were briefed about the devices, the objects that they will be interacting with, and the interaction modalities. Also, we showed an explanatory video depicting the interaction techniques, interface, and visual cues. (2) Next, participants calibrated the HoloLens, where the interpupillary distance was recorded for each participant. (3) Then, they proceeded to perform a training task, consisting of teleoperating an industrial robot arm to pick an L-shaped object, similar to the one to be used in the real task but bigger in size, at 1m from the robot. (4) Following, participants executed the experiment. Half of the participants started the task at 2m and after finishing all designs of cues and pose combinations repeated the same procedure at 3m. The other half started at 3m and then at 2m. For pose and designs of cues, we followed a Latin square distribution. (5) Finally, we carried out a short interview about their experiences.

4.6 Apparatus & Communication

We used a Kuka iiwa 7 R800 lightweight robotic arm with an attached Robotiq adaptive 2-Finger Gripper. Both are controlled via the Kuka iiwa's control unit. For our visual cues and the user interface, we used the Microsoft HoloLens 1, which is equipped with inside-out tracking, an HD video camera, and microphones allowing the use of head movement, speech, and gestures as input options [36]. The HoloLens has a field of view (FOV) of 30° × 17.5° with a resolution of 1268 × 720 px per eye. The application running on the HoloLens was programmed with Unity 2018.1.2 and the HoloToolkit Plugin [38].

The Kuka iiwa and the HoloLens communicate via the User Datagram Protocol in a local network. The control unit of the Kuka iiwa runs a specially designed back-end program that processes the received messages, moves the robot according to the receiving data, and returns its current status. To communicate position data between devices, we first converted the pose from Unity's left-handed coordinate system to the robot's right-handed one and used common length units (cm). Then, this cartesian position, rotation and velocity is sent to the control unit. This recognizes the command and allows the robot to plan the movement via internal inverse kinematics to then physically move the robotic arm.

A calibration process is important to achieve accuracy when using AR. The HoloLens adjust the hologram display according to the interpupillary distance. When it is not accurate, holograms may appear unstable or at an incorrect distance [37]. Thus, we ran the Microsoft calibration application for each user. Also, in order for the virtual tracking to be transferred to the real world, the position and alignment of the robot must be known in the virtual-world. For this, we use a 2D marker and the camera-based marker detection from Vuforia [45] when starting the application. Since the position of the 2D marker in relation to the robot is known, the position of the robot in the virtual'world can be determined and set as world anchors. The stability of world anchors has been previously evaluated and found a mean displacement error of 5.83±0.51mm [51]. This is precise enough to handle grasping tasks.

5 Results

5.1 Objective Measures

Time (Efficiency). Analyzing the time spent to perform the tasks, we opted for a RM-ANOVA with the design of cues, distance, and object pose, as independent variables and time as a dependent variable. For time, we used the average across the three trials for each condition. Testing the assumptions with Shapiro-Wilk, the distribution of some residuals showed a slight deviation from the normal distribution. The inspection of QQ-plots as well as skewness and kurtosis analysis, revealed that all residuals in question were skewed in the same positive direction and within an acceptable range [28]. In addition, the literature [6] suggests that a slight deviation from a normal distribution can be handled by ANOVA procedures, which is why we kept this approach. For pairwise comparisons, we refer to the estimated marginal means and report p-values and standard errors.

Results show that there were no significant three-way ($F(2, 46)=0.225$, $p= .8$) or two-way interaction effects for our independent variables (design of cues per distance $F(2, 46)=2.31$, $p= .11$; pose per distance $F(1, 23)=0.979$, $p= .333$; pose per design of cues, Greenhouse-Geiser corrected due to sphericity violation $F(1.562, 35.93)=2.036$, $p= .154$).

We found a significant main effect for each of our independent variables. For distance ($F(1, 23)=7.124$, $p = .014$, $\eta p^2= .236$), grasping objects at 2m (M=26.25s, SE=1.67) was significantly faster than at 3m (M=30.95s, SE=2.12), Fig. 3b. For pose ($F(1, 23)=7.145$, $p= .014$, $\eta p^2= .237$) participants completed the ambiguous pose (M=26.46s, SE=1.94) significantly faster than the clear pose (M=30.74s, SE=1.79), Fig. 3d. Finally, we also found significant differences in grasping times for the three design of cues ($F(2, 46)=13.029$, $p < .001$, $\eta p^2=.362$). Post-hoc pairwise comparisons (Bonferroni adjusted) showed significant differences for Colors (M=33.14s, SE=1.78) compared to Circle (M=26.12s, SE=2.04, p=.001) and Extensions (M=26.54s, SE=1.91, p= .003).

Success (Efficacy). For the dichotomous variable success, we applied a GEE model for which we used the GENLIN procedure in SPSS. As the working correlation matrix, we applied an exchangeable structure, and we used a binary logistic response model. For pairwise comparisons, we refer to the estimated marginal means and report p-values and standard errors.

Our results show no significant three-way or two-way interaction effects for our independent variables (design of cues per distance Wald $\chi^2(2, N=864) = 3.74$, $p= .154$; distance per pose Wald $\chi^2(1, N=864) = 3.59$, $p= .058$; pose per design of cues Wald $\chi^2(2, N=864) = 3.59$, $p= .166$).

Again, we found significant main effects for each of our independent variables. For distance (Wald $\chi^2(1, N=864) = 28.35$ $p < .001$), grasping objects at 2m (M=0.71, SD=0.46) was significantly more successful compared to 3m (M=0.43, SD=0.495). For pose (Wald $\chi^2(1, N=864)=11.526$, $p= .001$), the clear pose (M=0.66, SD=0.475) shows a significantly higher success compared to the ambiguous pose (M=0.48, SD=0.5). Finally, we also see a significant difference between the design of cues (Wald χ^2(2, N=864) = 42.96, p < .001). Post-hoc pairwise comparisons (Bonferroni adjusted) revealed that Circle (M=0.72, SE=0.03) showed a significantly higher success compared to both Extensions (M=0.58, SE=0.05, p= .004) and Colors (M=0.45, SE=0.04, p < .001). Also, there is a significant difference between Extensions and Colors (p= .02).

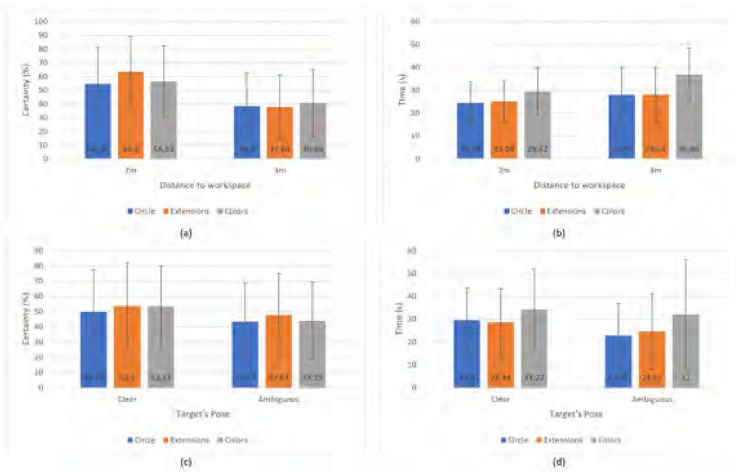


Fig. 3. Summary of M and SD with each design of visual cue (a) Shows certainty vs distance. (b) Shows time in seconds vs distance. (c) Shows certainty vs pose. (d) Shows time vs pose.

5.2 Subjective Measures

Certainty. We calculated the mean certainty rating across the three trials for each participant in each condition. Shapiro-Wilk tests and inspection of QQ plots showed that for the residuals normality could be assumed, we thus applied a RM-ANOVA. We found no significant three-way interaction (F(2, 46)=0.65, p= .94) but a significant two-way interaction for distance per design of cues (F(2, 46)=4.01, p= .025). Looking at the simple main effects for this interaction (post-hoc Estimated Marginal Means, Bonferroni adjusted), we found that for the 2m distance, the Extensions led to significantly higher certainty (M=63.30%, SE=4.27) compared to both Colors (M=56.53%, SE=4.11, p= .01) and Circle (M=54.58%, SE=4.83, p= .015). However, this is not the case for the 3m distance, where Extensions reached the lowest perceived certainty (Extensions M=37.64%, SE=4.06; Colors M=40.66%, SE=4.32; Circle M=38.40%, SE=4.01; differences not significant). We found a significant main effect both for pose (F(1, 23)=7.41, p= .012, ηp^2= .24) and distance (F(1, 23)=32.96, p < .01, ηp^2 = .589). For pose, the clear pose led to a higher certainty (M=52.13%, SD=27.42) compared to the ambiguous pose (M=44.91%, SD=26.02), Fig. 3c.

For distance, the closer distance of 2m led to a higher certainty (M=58.14%, SD=26.43) compared to 3m (M=38.90%, SD=23.89), Fig. 3a.

To understand the relationship between perceived certainty and measures of success and time, we calculated a Spearman Partial Correlation, and we controlled for the design of cues, distance, and pose. Results show that there is a significant positive partial correlation between certainty and success ($r_s(859)=0.144$, $p < .001$) but not between certainty and time ($r_s(859)=0.021$, $p= .53$).

6 Discussion

First **(H1)**, we hypothesized that our design of visual cues would perform similarly regardless of the different egocentric distances, improving hence depth perception. Our results could not support this hypothesis. A smaller distance prompted better results in terms of time, certainty, and success compared to 3m. This aligns with previous findings of depth perception decaying at greater distances in the real-world and in AR [43]. Further, Microsoft recommends a distance of 2m for an MR environment when using the HoloLens 1 as further distances can induce perceptual problems. Specifically, problems derived from capturing (flares and calibration) that can provoke scene distortion and problems in environment abstraction [29]. This might have influenced the results and the experience of using our visual cues at 3m. Additionally, all participants mentioned that performing the grasping task at 3m was harder than at 2m, e.g., P3, "The farther distance was exponentially more difficult." P19, "The farther away, the harder it is to understand depth independently of the virtual supporters." These results lead us to think that teleoperation in co-located spaces should be performed within the personal space (up to 2m) for better depth perception and thus performance. Despite that, we highlight the importance of further exploring distances beyond 2m to better understand the dynamics of teleoperation within the operator's action space.

Second **(H2)**, we hypothesized that our design of visual cues would reduce the effect of the target's pose, which would be reflected in similar results of success and time. However, we expected a higher degree of certainty for the clear pose compared to the ambiguous pose. Our results partially support this hypothesis. The effect of pose proved to be still present in spite of our designs of visual cues. Our results show that the clear pose prompts indeed higher certainty but also a significantly higher success with a compromise in terms of time (longer time) compared to the ambiguous pose. These results align with the findings of Barden & Petty [4] in terms of a direct relation between certainty and time. During our interview, we asked participants if they deemed one pose harder than the other for grasping, while most of them found the ambiguous pose harder, many also stated that they did not find differences among them, which could possibly be an effect of our visual cues. For instance, P7, "Both objects were equally hard," P9, "There was no difference between the 2 objects," "I did not find any object harder than another." Another factor that might have influenced these results is the type of task. Grasping tasks in a workspace with no other objects in vicinity that partially or completely occlude the target objects, may not highlight the potential benefits of our designs of cues.

Third **(H3)**, we hypothesized about potential differences between our design of cues with respect to time, success, and certainty. We assumed that Colors would perform best, followed by Extensions, and then Circle. Our results indeed showed that our designs of visual cues had different effects on our dependent variables, but the effects were different from what we hypothesized.

Regarding time, we found that our participants were faster using both the Circle and Extensions compared to using Colors. Additionally, no significant differences were found with respect to time between Circle and Extensions. When analyzing success, we found higher success rates when using Circle compared to using Extensions and Colors. Also, using Extensions showed a higher significant success compared to Colors.

Concerning certainty, we found a significant interaction effect for distance per design of cues. These point to differences at 2m, where Extensions prompted the highest level of certainty, which in turn was significantly higher compared to both Circle and Colors. Since our designs were incremental, Extensions, which included the Circle, seemed to provide the necessary information about depth without the visual complexity of Colors. The absence of this effect at 3m might be due to the fact that the effect of distance alone overshadowed any potential difference among our designs of cues. While Audley [2] remarked that the amount of information (evidence) that can be collected from the environment influences the amount of perceived certainty, e.g., the more evidence that is collected, the higher certainty that is evoked. Our results show that one must be careful when designing to provide more information about the environment. While the Colors certainly provided the most depth cues, our results point towards the fact that this information was not always usable to our participants. This is also reflected by our participants' comments, who mostly pointed out the simpler cues, Circle and Extensions, and deemed them as helpful: P1, "The visual extensions were good for the further distance." P10, "The most helpful thing was the circle. I used it the most to align the gripper." P11, "The circle was what helped the most." We attribute the preferences towards the Circle, to the fact that it did not fully occlude the real object, and when it did it was an indicator of misalignment of the gripper over the target object. Further, participants' subjective preferences for Extensions and Colors together with higher efficiency and effectiveness lead us to recommend a simpler design of visual cues.

In H3, we did not expect Colors to perform the worst across conditions. In consequence, we further explored the reasons behind the low scores in time and success by analyzing the participants' comments. During our interview, participants expressed confusion when using the Colors due to lack of color opponency. Specifically, they expressed problems with the color gradient on the physical colormap, e.g., P13, "The gradient of the colormap made it a little bit confusing. I could not tell if it was already yellow or still blue?" P15, "I would prefer stripes, not gradient colors, that would have made it easier to distinguish where I was on the tabletop." Added to that, participants expressed difficulties in distinguishing the real and physical colors, which is related to focal rivalry—the human visual system cannot focus on two elements at the same time, e.g., P14, "I had to concentrate to match the colors;" P15, "It was hard to see the real and the virtual colors because they were right above each other." Therefore, we believe that this lack of color opponency added to focal rivalry worsened depth perception. This accords to the observations of Ellis & Menges [15], who determined that physical surfaces influence depth perception misestimation.

Fourth **(H4)**, we hypothesized about finding a correlation between certainty and success as a possible pointer to better depth perception. Our results confirmed the correlation between certainty and success, even when controlling for our independent variables (distance, pose, and design of cues). We still found a positive and significant, albeit rather weak, correlation. This aligns with a line of research suggesting that subjective certainty correlates closely with objective success [8], [58]. We consider that all our design cues provided additional evidence about the position of the gripper in the workspace, which in turn influenced certainty. A note of caution is due here since, as mentioned in Section 2.4, there have been controversial findings related to the influence of time and success in certainty. We acknowledge that other factors can influence certainty and were not considered in our experiment such as fatigue and changes in attention, and these have proven to influence perceived certainty [20]. Further, when evaluating depth perception in 3D environments, it is necessary to separate "the amount of depth that an object is seen to have (mind independent property) and the realism of the experience (mind dependent property)" [23]. In fact, we consider that this construction can also be applied to AR environments and is related to certainty. This in consequence might have influenced the perceived certainty during our experiment.

7 Limitations

A limitation of our work relates to the multimodal interaction technique used. While a joystick or a control pad are commonly used in robot teleoperation, we aim to explore hands-free multimodal interaction. This type of interaction has raised interest in the research community, especially in HRI. Our previous experiences have shown that using speech and head movements is simpler to learn than using a control pad or even a joystick. Additionally, these modalities are natively supported by the technology used (Microsoft HoloLens 1). We further stress that the modalities were kept stable among conditions to avoid them causing a major influence on the evaluation of the other factors in our experiment.

This work may be also limited to the technology used. For instance, our design of the Colors cue, which combines a real-world colormap with virtual colors displayed over it, could have contributed to focal rivalry problems. However, this problem is present not only in the Microsoft HoloLens 1 but in the current generation of MR headsets [31]. Additionally, current ARHMDs have a limited field of view which do not cover human's peripheral vision. This presents a disadvantage when using AR, as mentioned by Williams et al. [54].

Our experimental design also presents certain limitations. We did not consider a condition without cues since a previous study [1] already evaluated the use of visual cues versus the absence of them, suggesting potential improvements in certainty. This study thus builds upon those findings and further explores the influence of distance and pose. We did not consider grasping accuracy, defined by the distance to an ideal grasping position, since we realized that the effect of the displacement of the virtual visual cues at 3m influences greatly this measure.

Our main goal is to capture first experiences with certain designs of visual cues that can provide direction for a better design that improves depth perception. Furthermore, people can perceive distance and depth differently due to individual differences in visual acuity, eye dominance, color vision, and spatial abilities [29]. We considered different color visions for the colors used in the colormap but left aside the other factors which might have influenced on how each participant experienced not only the types of cues but depth perception.

8 Conclusions

In this paper, we evaluated how egocentric distances and target objects' pose affect co-located teleoperation when using certain designs of visual cues presented through AR. To this end, we performed an experiment with 24 participants. Our results align with previous studies about depth perception within the observer's personal space. Teleoperating a robot at 2m with our designs of visual cues leads to a higher success rate, shorter time, and a higher degree of certainty compared to 3m. A clearer pose leads to higher success and certainty but requires longer time. As time and certainty are closely tied, i.e., greater thoughtfulness is required to achieve higher certainty and success. We also found a positive correlation between success and certainty, but we are careful with these findings since other factors, e.g., attention and fatigue, have an effect on certainty and were not considered in our experiment. Additionally, we found differences of our designs of visual cues. The Circle and Extensions cues prompted shorter times and higher success compared to Colors, wherein Circle showed the highest success. These findings suggest that simplicity of design leads to higher efficiency and effectiveness.

We consider that our findings are thought-provoking and present a detailed analysis of distance and target pose in co-located teleoperation. These, further contemplate the role of certainty as a factor that can shed some light about depth perception. Besides, our designs of visual cues suggest advantages and downsides of using cast shadows, symbolic enhancements, and combining real and virtual landmarks to enhance the visual space when teleoperating a robotic arm in co-located spaces. Although our study focuses on evaluating specific factors related to depth perception, our findings may well have a bearing on designing cues using AR to improve perception of real objects and facilitate teleoperation of robotic arms.

Acknowledgements

This research is supported by the German Federal Ministry of Education and Research (BMBF, FKZ: 13FH011IX6).

References

1. Arévalo Arboleda, S., Dierks, T., Rücker, F., Gerken, J.: There's more than meets the eye. In: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. pp. 104–106. ACM, [S.l.] (2020). https://doi.org/10.1145/3371382.3378240
2. Audley, R.J.: A stochastic model for individual choice behavior, vol. 67 (1960). https://doi.org/10.1037/h0046438
3. Baranski, J.V., Petrusic, W.M.: Probing the locus of confidence judgments: Experiments on the time to determine confidence. Journal of experimental psychology. Human perception and performance **24**(3), 929–945 (1998). https://doi.org/10.1037//0096-1523.24.3.929
4. Barden, J., Petty, R.E.: The mere perception of elaboration creates attitude certainty: exploring the thoughtfulness heuristic. Journal of Personality and Social Psychology **95**(3), 489–509 (2008). https://doi.org/10.1037/a0012559
5. Barthelmé, S., Mamassian, P.: Evaluation of objective uncertainty in the visual system. PLoS Computational Biology **5**(9), e1000504 (2009). https://doi.org/10.1371/journal.pcbi.1000504
6. Blanca, M.J., Alarcón, R., Arnau, J., Bono, R., Bendayan, R.: Non-normal data: Is anova still a valid option? Psicothema **29**(4), 552–557 (2017). https://doi.org/10.7334/psicothema2016.383
7. Bodenheimer, B., Meng, J., Wu, H., Narasimham, G., Rump, B., McNamara, T.P., Carr, T.H., Rieser, J.J.: Distance estimation in virtual and real environments using bisection. In: Proceedings, APGV 2007. p. 35. ACM (2007). https://doi.org/10.1145/1272582.1272589
8. Boldt, A., Yeung, N.: Shared neural markers of decision confidence and error detection. Journal of Neuroscience **35**(8), 3478–3484 (2015). https://doi.org/10.1523/JNEUROSCI.0797-14.2015
9. Brenner, E., Smeets, J.B.J.: Depth perception. In: Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, pp. 1–30. John Wiley & Sons, Inc (2018). https://doi.org/10.1002/9781119170174.epcn209
10. Choi, H., Cho, B., Masamune, K., Hashizume, M., Hong, J.: An effective visualization technique for depth perception in augmented reality-based surgical navigation. The International Journal of Medical Robotics and Computer Assisted Surgery **12**(1), 62–72 (2016). https://doi.org/10.1002/rcs.1657
11. Cipiloglu, Z., Bulbul, A., Capin, T.: A framework for enhancing depth perception in computer graphics. In: Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization. p. 141. ACM (2010). https://doi.org/10.1145/1836248.1836276
12. Cutting, J.E., Vishton, P.M.: Perceiving layout and knowing distances. In: Perception of Space and Motion, pp. 69–117. Elsevier (1995). https://doi.org/10.1016/B978-012240530-3/50005-5

13. Diaz, C., Walker, M., Szafir, D.A., Szafir, D.: Designing for depth perceptions in augmented reality. In: 2017 IEEE International Symposium on Mixed and Augmented Reality. pp. 111–122. IEEE (2017). https://doi.org/10.1109/ISMAR.2017.28
14. El Jamiy, F., Marsh, R.: Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality. IET Image Processing **13**(5), 707–712 (2019). https://doi.org/10.1049/iet-ipr.2018.5920
15. Ellis, S.R., Menges, B.M.: Localization of virtual objects in the near visual field. Human Factors: The Journal of the Human Factors and Ergonomics Society **40**(3), 415–431 (1998). https://doi.org/10.1518/001872098779591278
16. de Gardelle, V., Mamassian, P.: Does confidence use a common currency across two visual tasks? Psychological science **25**(6), 1286–1288 (2014). https://doi.org/10.1177/0956797614528956
17. Gherman, S., Philiastides, M.G.: Neural representations of confidence emerge from the process of decision formation during perceptual choices. NeuroImage **106**, 134–143 (2015). https://doi.org/10.1016/j.neuroimage.2014.11.036
18. Gong, L.L., Ong, S.K., Nee, A.Y.C.: Projection-based augmented reality interface for robot grasping tasks. In: Proceedings of the 2019 4th International Conference on Robotics, Control and Automation - ICRCA 2019. pp. 100–104. ACM Press (2019). https://doi.org/10.1145/3351180.3351204
19. Hedayati, H., Walker, M., Szafir, D.: Improving collocated robot teleoperation with augmented reality. In: HRI'18. pp. 78–86. ACM (2018). https://doi.org/10.1145/3171221.3171251
20. Heereman, J., Walter, H., Heekeren, H.R.: A task-independent neural representation of subjective certainty in visual perception. Frontiers in human neuroscience **9**, 551 (2015). https://doi.org/10.3389/fnhum.2015.00551
21. Heinrich, F., Bornemann, K., Lawonn, K., Hansen, C.: Depth perception in projective augmented reality: An evaluation of advanced visualization techniques. In: 25th ACM Symposium on Virtual Reality Software and Technology. pp. 1–11. ACM (11122019). https://doi.org/10.1145/3359996.3364245
22. Hershenson, M.: Size-distance invariance: kinetic invariance is different from static invariance. Perception & psychophysics **51**(6), 541–548 (1992). https://doi.org/10.3758/BF03211651
23. Hibbard, P.B., Haines, A.E., Hornsey, R.L.: Magnitude, precision, and realism of depth perception in stereoscopic vision. Cognitive research: principles and implications **2**(1), 25 (2017). https://doi.org/10.1186/s41235-017-0062-7
24. Howard, I.P.: Depth perception. In: Stevens' handbook of experimental psychology (2002), pp. 77–120
25. Howard, I.P., Rogers, B.J.: Perceiving in depth, vol. 29. Oxford University Press (2012)
26. Jones, A., Swan, J.E., Singh, G., Kolstad, E.: The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In: IEEE virtual reality 2008. pp. 267–268. IEEE (2008). https://doi.org/10.1109/VR.2008.4480794
27. Kiani, R., Corthell, L., Shadlen, M.N.: Choice certainty is informed by both evidence and decision time. Neuron **84**(6), 1329–1342 (2014). https://doi.org/10.1016/j.neuron.2014.12.015
28. Kim, H.Y.: Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restorative dentistry & endodontics **38**(1), 52–54 (2013). https://doi.org/10.5395/rde.2013.38.1.52

29. Kruijff, E., Swan, J.E., Feiner, S.: Perceptual issues in augmented reality revisited. In: 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2010. pp. 3–12 (2010). https://doi.org/10.1109/ISMAR.2010.5643530
30. Krupke, D., Steinicke, F., Lubos, P., Jonetzko, Y., Gorner, M., Zhang, J.: Comparison of multimodal heading and pointing gestures for co-located mixed reality human-robot interaction. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 1–9. IEEE (2018). https://doi.org/10.1109/IROS.2018.8594043
31. Laramee, R., Ware Colin: Rivalry and interference with a head-mounted display. ACM Transactions on Computer-Human Interaction **9**(3), 238–251 (2002). https://doi.org/10.1145/568513.568516
32. Lee, J.H., Herzog, T.A., Meade, C.D., Webb, M.S., Brandon, T.H.: The use of gee for analyzing longitudinal binomial data: A primer using data from a tobacco intervention. Addictive Behaviors **32**(1), 187 – 193 (2007). https://doi.org/https://doi.org/10.1016/j.addbeh.2006.03.030
33. Livingston, M.A., Zanbaka, C., Swan, J.E., Smallman, H.S.: Objective measures for the effectiveness of augmented reality. In: Virtual reality 2005. pp. 287–288. IEEE (2005). https://doi.org/10.1109/VR.2005.1492798
34. Loomis, J., Knapp, J.: Visual perception of egocentric distance in real and virtual environments. In: Virtual and adaptive environments, pp. 21–46. Lawrence Erlbaum (2003). https://doi.org/10.1201/9781410608888.pt1
35. Mamassian, P.: Visual confidence. Annual review of vision science **2**, 459–481 (2016). https://doi.org/10.1146/annurev-vision-111815-114630
36. Microsoft: Microsoft hololens, https://www.microsoft.com/en-us/hololens
37. Microsoft: Microsoft hololens calibration, https://docs.microsoft.com/en-us/hololens/hololens-calibration
38. Microsoft: Mixed reality toolkit (2017), https://github.com/microsoft/MixedRealityToolkit-Unity/releases
39. Mosiello, G., Kiselev, A., Loutfi, A.: Using augmented reality to improve usability of the user interface for driving a telepresence robot. Paladyn, Journal of Behavioral Robotics **4**(3) (2013). https://doi.org/10.2478/pjbr-2013-0018
40. Paris, R., Joshi, M., He, Q., Narasimham, G., McNamara, T.P., Bodenheimer, B.: Acquisition of survey knowledge using walking in place and resetting methods in immersive virtual environments. In: Proceedings of the ACM Symposium on Applied Perception. pp. 1–8. ACM (2017). https://doi.org/10.1145/3119881.3119889
41. Park, J., Ha, S.: Visual information presentation in continuous control systems using visual enhancements. In: Contact-free Stress Monitoring for User's Divided Attention. INTECH Open Access Publisher (2008). https://doi.org/10.5772/6307
42. Petrocelli, J.V., Tormala, Z.L., Rucker, D.D.: Unpacking attitude certainty: attitude clarity and attitude correctness. Journal of Personality and Social Psychology **92**(1), 30–41 (2007). https://doi.org/10.1037/0022-3514.92.1.30
43. Ping, J., Weng, D., Liu, Y., Wang, Y.: Depth perception in shuffleboard: Depth cues effect on depth perception in virtual and augmented reality system. Journal of the Society for Information Display **28**(2), 164–176 (2020). https://doi.org/10.1002/jsid.840
44. Pleskac, T.J., Busemeyer, J.R.: Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. Psychological review **117**(3), 864–901 (2010). https://doi.org/10.1037/a0019737
45. PTC: Vuforia engine in unity, https://library.vuforia.com/articles/Training/getting-started-with-vuforia-in-unity.html

46. Rolland, J.P., Meyer, C., Arthur, K., Rinalducci, E.: Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. Presence: Teleoperators and Virtual Environments **11**(6), 610–625 (2002). https://doi.org/10.1162/105474602321050730
47. Rosales, C.S., Pointon, G., Adams, H., Stefanucci, J., Creem-Regehr, S., Thompson, W.B., Bodenheimer, B.: Distance judgments to on- and off-ground objects in augmented reality. In: Proceedings, 26th IEEE Conference on Virtual Reality and 3D User Interfaces. pp. 237–243. IEEE (2019). https://doi.org/10.1109/VR.2019.8798095
48. S. Stadler, K. Kain, M. Giuliani, N. Mirnig, G. Stollnberger, M. Tscheligi: Augmented reality for industrial robot programmers: Workload analysis for task-based, augmented reality-supported robot control. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). pp. 179–184 (2016). https://doi.org/10.1109/ROMAN.2016.7745108
49. Shimojo, S., Paradiso, M., Fujita, I.: What visual perception tells us about mind and brain. Proceedings of the National Academy of Sciences of the United States of America **98**(22), 12340–12341 (2001). https://doi.org/10.1073/pnas.221383698
50. Tormala, Z.L.: The role of certainty (and uncertainty) in attitudes and persuasion. Current Opinion in Psychology **10**, 6–11 (2016). https://doi.org/10.1016/j.copsyc.2015.10.017
51. Vassallo, R., Rankin, A., Chen, E.C.S., Peters, T.M.: Hologram stability evaluation for microsoft hololens. In: Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment. p. 1013614. SPIE Proceedings, SPIE (2017). https://doi.org/10.1117/12.2255831
52. Walker, M., Hedayati, H., Lee, J., Szafir, D.: Communicating robot motion intent with augmented reality. In: Kanda, T., Ŝabanović, S., Hoffman, G., Tapus, A. (eds.) HRI'18. pp. 316–324. ACM (2018). https://doi.org/10.1145/3171221.3171253
53. Walker, M.E., Hedayati, H., Szafir, D.: Robot teleoperation with augmented reality virtual surrogates. In: HRI'19. pp. 202–210. IEEE (2019). https://doi.org/10.1109/HRI.2019.8673306
54. Williams, T., Hirshfield, L., Tran, N., Grant, T., Woodward, N.: Using augmented reality to better study human-robot interaction. In: Virtual, Augmented and Mixed Reality. Design and Interaction, vol. 12190, pp. 643–654. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-49695-1 43
55. Willis, J., Todorov, A.: First impressions: making up your mind after a 100-ms exposure to a face. Psychological science **17**(7), 592–598 (2006). https://doi.org/10.1111/j.1467-9280.2006.01750.x
56. Wither, J., Hollerer, T.: Pictorial depth cues for outdoor augmented reality. In: ISWC 2005. IEEE (2005). https://doi.org/10.1109/ISWC.2005.41
57. Xie, F., Paik, M.C.: Generalized estimating equation model for binary outcomes with missing covariates. Biometrics **53**(4), 1458 (1997). https://doi.org/10.2307/2533511
58. Yeung, N., Summerfield, C.: Metacognition in human decision-making: confidence and error monitoring. Philosophical transactions of the Royal Society of London. Series B, Biological sciences **367**(1594), 1310–1321 (2012). https://doi.org/10.1098/rstb.2011.0416
59. Yonas, A., Granrud, C.E.: Infants' perception of depth from cast shadows. Perception & psychophysics **68**(1), 154–160 (2006). https://doi.org/10.3758/bf03193665
60. Zollmann, S., Hoppe, C., Langlotz, T., Reitmayr, G.: Flyar: augmented reality supported micro aerial vehicle navigation. IEEE transactions on visualization and computer graphics **20**(4), 560–568 (2014). https://doi.org/10.1109/TVCG.2014.24

7.7 Paper VII. Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues.

Arévalo Arboleda, Stephanie; Rücker, Franziska; Dierks, Tim; and Gerken, Jens, (Eds. 2021). Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '21), ACM, NY. DOI: 10.1145/3411764.3445398.

Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues

Stephanie Arévalo Arboleda
Westphalian University of Applied Sciences
Gelsenkirchen, Germany
stephanie.arevalo@w-hs.de

Franziska Rücker
Westphalian University of Applied Sciences
Gelsenkirchen, Germany
franziska.ruecker@w-hs.de

Tim Dierks
Westphalian University of Applied Sciences
Gelsenkirchen, Germany
tim.dierks@w-hs.de

Jens Gerken
Westphalian University of Applied Sciences
Gelsenkirchen, Germany
jens.gerken@w-hs.de

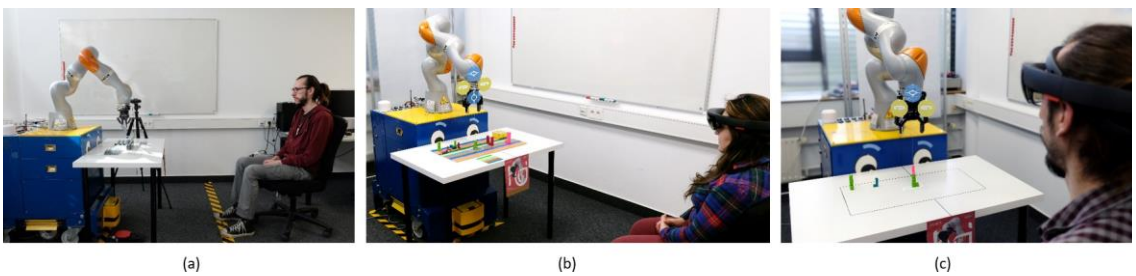


Figure 1: (a) In this work, we explore how to improve performance in manipulation and grasping tasks by using augmented reality to present visual cues in teleoperation of a robotic arm in co-located spaces. To achieve that we designed two types of augmented visual cues (b) Basic Augmented Cues (c) Advanced Augmented Cues

ABSTRACT

Teleoperating industrial manipulators in co-located spaces can be challenging. Facilitating robot teleoperation by providing additional visual information about the environment and the robot affordances using augmented reality (AR), can improve task performance in manipulation and grasping. In this paper, we present two designs of augmented visual cues, that aim to enhance the visual space of the robot operator through hints about the position of the robot gripper in the workspace and in relation to the target. These visual cues aim to improve the distance perception and thus, the task performance. We evaluate both designs against a baseline in an experiment where participants teleoperate a robotic arm to perform pick-and-place tasks. Our results show performance improvements in different levels, reflecting in objective and subjective measures with trade-offs in terms of time, accuracy, and participants' views of teleoperation. These findings show the potential of AR not only in teleoperation, but in understanding the human-robot workspace.

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05. . . $15.00
https://doi.org/10.1145/3411764.3445398

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; *Empirical studies in interaction design*.

KEYWORDS

human-robot interaction, visual cues, augmented reality, robot teleoperation

ACM Reference Format:

Stephanie Arévalo Arboleda, Franziska Rücker, Tim Dierks, and Jens Gerken. 2021. Assisting Manipulation and Grasping in Robot Teleoperation with Augmented Reality Visual Cues. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3411764.3445398

1 INTRODUCTION

Co-located teleoperation of robotic arms can be a challenging task that requires high level of attention, expertise and a good understanding of the environment, the robot, and the objects in it. In scenarios where the operator is not in direct proximity to the robot but at a certain distance (Figure 1a), some problems related to depth and distance perception may occur which in turn can affect performance. A prevailing problem in robot teleoperation is distance estimation in manipulation tasks, which is common in assembly environments. In addition, depth perception problems are part of the lack of situation awareness in teleoperation [15], which can result in lower performance in manipulation tasks. This also relates to the ability to imagine how an object looks from a perspective different to the egocentric view, i.e., spatial orientation, which has been analyzed in remote teleoperation, showing how it affects teleoperation performance [33]. When analyzing co-located teleoperation (operator and robot located in the same physical space), the operator can experience problems related to depth perception and misjudge egocentric distances[43], similar to the problems experienced in remote teleoperation.

There have been a few efforts to improve the depth perception in remote teleoperation using cameras to capture different angles of a scene [23]. However, teleoperation has received fewer attention in co-located spaces. Problems in perception of distances and size can still occur when objects are located within the observer's personal space (2 m), i.e. overestimating distances [9]. The human eye naturally perceives the distance of objects around using visual cues processed by cognitive processes and can be divided into monocular and binocular cues [11]. Monocular cues aid to perceive distances in a near space (< 3m) [6] and comprise static and dynamic cues [22]. Dynamic cues are acquired by the motion of objects and/or observers and are crucial to perceive depth and object structure [46]. However, there are some scenarios where this motion is impossible or difficult (absence of dynamic cues) because the observer or the objects cannot move, e.g., people with mobility impairments, or scenarios where the visual angle cannot be changed due to workplace settings. Here, the observer can only count on natural static cues which might not suffice to get an accurate perception of distance and depth. Inspired by the static cues that humans naturally use to judge depth, we provide enhanced hints with Augmented Reality (AR) that consider elements from monocular cues such as lighting and interposition to provide a better understanding of the environment, the robot, and object affordances. As a use case, we considered a manufacturing environment with an industrial robotic arm where the operator is in a co-located space, teleoperating a robot to perform manipulation and grasping tasks.

The goal of this paper is to contribute to the research efforts of the use of AR for robot teleoperation from a human-centered robotics perspective. We present two designs of visual cues using AR, which we named, Augmented Visual Cues (AVC). These have the potential to improve task performance by providing better awareness of the objects' location and its depth within co-located spaces. In our first design, basic augmented visual cues, we augment the robot and the environment by providing a combination of real and virtual hints. In our second design, advanced augmented visual cues, we utilize sensor-based knowledge of the environment, i.e., using-pose object recognition, to present virtual hints that enhance the robot, environment and objects in it.

In the following sections, we present relevant work related to robot depth perception and the use of AR in robot teleoperation. Then, we introduce the reasoning behind AVC and present each design of basic and advanced augmented visual cues, followed by an experiment where we evaluated the user's performance in teleoperation using our designs of AVC against a baseline condition in a pick-and-place task. Next, we present our results, followed by a discussion section. Finally, we present the limitations of our findings and final conclusions.

2 RELATED WORK

2.1 Depth Perception

The human visual system captures environment information through the eye, transmits this information through the optic nerve to the brain, which in turn interprets the received information [2]. Among this information we find depth perception, which can be defined as the ability to see volume and the relative position of objects in a three-dimensional environment [52]. Further, size perception is closely related to depth perception since humans use depth cues to perceive size. Thus, in environments where the number of depth cues is reduced, humans' abilities to perceive size are also compromised [42].

Depth perception enables our interaction with surrounding objects and helps to determine size and distances, through monocular (provided by one eye) and binocular (provided by both eyes) cues [22]. Monocular cues provide information that allows to determine more or less accurately the distance and size of objects located at an egocentric distance of approximately 3m [6]. Monocular cues can be divided into static, i.e, do not need the observer to change perspective, and dynamic cues which are acquired by motion, e.g., the observer needs to move around the environment to change the egocentric perspective. Static cues comprise of perspective (linear and texture), interposition (occlusion and transparency), lighting (shading and shadow), aerial (optical haze and blur) and dynamic cues involving optical flow, motion parallax and accretion [22]. In general, humans combine different types of cues to understand distance, position, and the relation among objects in space.

From the cognition perspective, it is relevant to discuss cognitive maps. These are individual representations about "the spatial and environmental information of the geographical space" [26]. These maps help humans to locate objects in the environment, and are closely related to individual spatial abilities. In robot teleoperation, spatial orientation (own ability to imagine objects from different visual perspectives) and spatial relation (own ability to mentally manipulate objects) have proven to be relevant in performance [33].

2.2 AR and the Visual Space in Robot Teleoperation

AR explores different ways to superimpose virtual images into the real world seamlessly [18]. The latest widespread of the use of AR head-mounted displays (ARHMD), e.g. Microsoft HoloLens [34] and Magic Leap One [30], has brought along a growing interest in incorporating solutions grounded in the alignment of real and virtual objects in the user's line of sight [36]. Robotics and Human- Robot Interaction (HRI) can benefit greatly from the use of AR. In fact, there has been an increasing body of research on using AR in robotics [31], especially, using AR for robot control [13, 53]. One of the most explored areas of research is using AR to find ways to visualize robot motion, e.g., trajectory planning by previewing planned simulated paths for robot motion [12], focusing on path planning for collision avoidance [7, 12], and co-located teleoperation of aerial robots [21, 47, 48]. Combining robot control with computer vision techniques is a promising area of research, e.g. using deep learning for 3D object pose estimation combined with AR to create a robot system [38].

Augmenting the visual space in a human-robot context has gathered the interest of the HRI community. For instance, AR has been used to provide information that help robot industrial programmers, and it has shown promising results in reducing their cognitive load [41]. Previous work [17], presented a projection-based AR interface to provide instructions that will assist operators in robot programming pick-and-place tasks, and Gacem et al. [14] present a robotic arm that projects a spotlight to localize objects in unfamiliar and dense environments.

Research in AR for assembly environments has also gained popularity. Wang et al. [50] reviewed a number of papers published in 25 years that investigate AR in assembly tasks, which show the influence and potential of AR in assembly systems. Here, enhancing the visual space to facilitate robot control has become an area of interest in manufacturing environments, e.g., Clemente et al. [8] proposed showing visual feedback of the gripper force and closure. Aligned with this line of research, there has been an emphasis on the importance of visually showing the robots' intentions through ARHMD [49, 51], or using projections on the environment [5].

3 AUGMENTED VISUAL CUES (AVC) FOR TELEOPERATION

In this paper, we build on related work by combining AR and robot teleoperation to address distance and depth perception problems, thus, improving task performance. Teleoperating a robot in a co-located assembly environment, requires the operator to process and understand the surroundings and spatial relations between objects and the robot gripper, especially if the visual perspective is fixed, leading to misestimations of the position of the robot gripper relative to target objects. These common misestimations negatively influence accuracy and efficiency during pick-and-place tasks, e.g., it leads to multiple attempts to ft an object in a specific area, repositioning the gripper several times, or failing to grasp a target object. We present two types of AVC, designed for ARHMD, e.g. Microsoft HoloLens, to improve depth perception in pick-and-place tasks. Both approaches capitalize on the conceptual work of Walker et al. [47] who proposed a design framework considering the elements from AR for a user-centered HRI. They present three archetypes: "augmenting the environment", using visual cues embedded in the interaction environment; "augmenting the robot", where visual cues are directly connected to the robot in a way that they might change its real features; or "augmenting the user interface", where ego-centric imagery could provide system information to the user.

In particular, our AVC designs provide a set of visual cues inspired by certain monocular cues, namely lighting and interposition, which humans use to understand their spatial environment. We clarify that lighting relates to the shadows that an object projects over another object, providing information about relative depth [45].

Our AVC designs differ from each other concerning the knowledge of the environment they require. As a baseline, we propose Basic Augmented Visual Cues (Basic Cues) which only require information about the workspace where the robot is operating, such as the height of the robot above the tabletop and its dimensions. This type of cue has certain limitations in how it can assist the user,

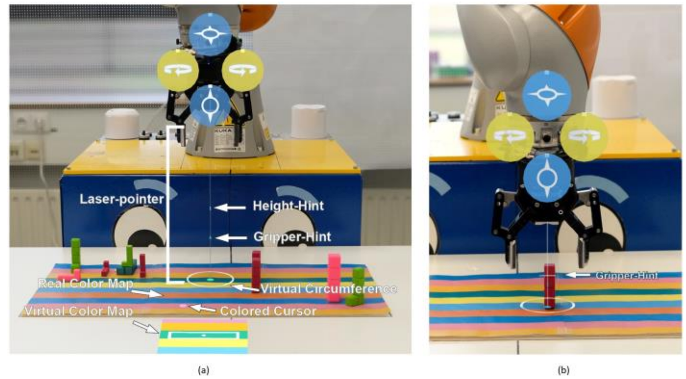


Figure 2: (a) Visual representation of each basic augmented cue. (b) Gripper-hint showing adequate grasping position.

simply because there is no previous knowledge of the environment, for instance, the position of the objects on the workspace, or of the robot. Therefore, our second type of AVC, Advanced Augmented Visual Cues (Advanced Cues), investigates the use of object-pose recognition to build a spatial understanding of the objects in the scene and integrates this knowledge to further assist the operator.

Both AVC designs build on our general interaction principle for robot teleoperation. For this, we apply hands-free multimodal interaction by combining head-movements, using a head-gaze based cursor to point, and speech commands to commit certain actions. We also present a robot control interface that shows basic controls visually and spatially anchored to the gripper of the robot. These controls allow moving the gripper along the x-, y-, z-axis as well as rotating it. We will explain this interaction and interface design in detail in Section 4.7.

3.1 Basic Augmented Visual Cues (Basic Cues)

Our design of Basic Cues can be defined as indirect cues that suggest certain characteristics of the environments and the objects on it. These cues assist the operator to first determine the position of a target (depth perception) to then accurately positioning the gripper over a target. They aim to help teleoperation in unknown environments or when object-pose recognition is not possible.

In order to address distance and depth perception related problems, we designed a combination of physical and virtual hints. As the representative for real hints, we position a poster with a colormap upon the surface of the workspace. The colormap intends to provide a visual landmark that can help determine the exact or relative position of an object, e.g., the object in Figure 2b is located on the light blue stripe. However, it is still difficult to accurately determine the relative x,y position of the gripper above the workspace. Thus, we added a virtual colormap and other virtual hints that would indicate the position of the gripper on the workspace. Figure 2a shows a virtual semi-transparent line, which we named "Laser-pointer" reaching from the gripper towards the tabletop and crossing the dark green stripe on the physical colormap. Correspondingly, we provide a virtual colormap at the front of the table, which, with a rectangle, highlights the color or colors of the relative position of the gripper on the physical colormap.

Additionally, our virtual cues are inspired by the monocular cue of lighting, e.g., we present a virtual circumference under the gripper, emulating a shadow of the gripper on the tabletop. In the following section, we present the design and purpose of each basic cue showed in Figure 2 in detail:

Laser-pointer. This virtual cue consists of a semi-transparent straight vertical line. It starts at the center of the grip region until it reaches the tabletop. It has a small colored point at the end of the straight line on the tabletop. The color of that point matches the color of the physical colormap. The laser-pointer was designed to help to visualize the center position of the gripper and facilitate repositioning it in relation to the target object. Additionally, this semi-transparent line changes length dynamically when the gripper moves up or down along the y-axis (towards the tabletop). It also overlays objects located between the gripper and the tabletop, as it has no means of being aware of their presence or position.

ColorMap. This hint consists of a physical and virtual colormap. For the physical hint, we printed a poster with colored stripes —each 3 cm wide, considering the width of the gripper's fingers, which is of 2.2 cm, and covers the robot's workspace on the tabletop. It presents 7 different colors, which repeat sequentially to match the size of the workspace. The color combination was carefully selected to consider different types of color blindness following Wong's guidelines [55]. The virtual hint can be described as a window with colored stripes —showing the same colors presented in the poster. It is located at the front of the tabletop and is aligned to be in front of the current position of the gripper. Its function is to provide a visualization of the current position of the gripper on the tabletop, in the same way as a camera attached to the gripper facing the tabletop would. In addition, there is a white rectangle drawn over this window, symbolizing the grip region. It also has a white point in the middle of the rectangle that represents the laser-pointer. This cue intends to allow the operator to position the gripper right above a target object, reducing depth perception problems and assisting with partial occlusion.

Colored cursor. The cursor adopts the color of the physical stripe where it is currently pointing. This hint supports the operator by roughly pointing at a specific area on the tabletop, thereby assisting to visualize the position on the table the operator is pointing at.

Colored-hints on the laser-pointer. Along the laser-pointer we added two visual indicators, height-hint and gripper-hint. **The height-hint** highlights the mid-point between the gripper and the tabletop, Figure 2a. **The gripper-hint** is visible after the first movement along the robot's z-axis, and it indicates the future end-position of the gripper once its fingers are closed (grasping position) and it is presented as a white line that intersects the laser-pointer, Figure 2b. It indicates if the gripper is too close to the tabletop and it would collide with it while trying to grasp an object, or if the gripper is too far from the target and would fail to grasp it.

Virtual Circumference. Our design shows a virtual circumference with a diameter that matches the grip region, right below the position of the gripper on the tabletop. The main function of this virtual cue is to provide a better understanding of the diameter of the grip region over the tabletop. It was designed to assist the operator in determining if an object is within the grasping range or to determine if the gripper's fingers would collide with an object in the vicinity of the target object.

3.2 Advanced Augmented Visual Cues (Advanced Cues)

Advance Cues build upon the use of object-pose recognition of the objects in the scene. The resulting cues are therefore able to integrate this object-pose data to become environment-aware and provide explicit hints for interaction. This is achieved using the 3D model of the recognized objects and virtually overlaying them over the workspace. This virtual model is shaded to be invisible to operators and only occlude virtual elements behind it but not the real world. This allowed us to show a virtual representation of grasped objects that behave the same way as real objects. That is why we named the cues derived from object-pose recognition Advanced Cues, as we do not only present cues inspired in the monocular cue of lighting but also add occlusion to this design (Figure 3a, 3c). We propose and describe them as follows:

Laser-pointer and virtual grip region. We maintain this cue from the Basic Cues but added the occlusion capability, e.g., the laser-pointer projects from the center of the gripper down towards the tabletop until it reaches the nearest surface, which can be a target or neighboring objects, see Figure 3. Additionally, we provide a representation of the grip region on the tabletop in the form of a transparent rectangle with bolded borders. The virtual grip region is located just below the current position of the gripper and changes its diameter matching the current aperture of the gripper, e.g. when the fingers are opened or closed with an object in between (Figure 3a, 3b). This cue allows the operator to determine if an object is located within the grip region for picking and it presents the occlusion capability as seen in Figure 3c.

Virtual extensions. The design of this cue augments the robot's gripper with a virtual elongation of the gripper's fingers — similar to the laser-pointer. However, it is only visible before picking an object, e.g. when there is a danger of potential collision of one or both fingers with an object. It is presented by a virtual red semi- transparent plane that starts at the tip of the finger until it intersects with the area of the object that it would collide (Figure 3a). The purpose of this cue is to alert the operator of potential collisions and the need for repositioning the gripper.

Ghost object and ghost gripper. Figure 3b, 3c, and 3d show a virtual copy of the grasped object presented in a gray color. This ghost object appears once an object has been grasped and it follows the operator's gaze. It therefore allows to visualize the future position, pose, and space that an object occupies on a determined area (Figure 3b, 3c) without the need to actually conduct the interaction. The gaze-follow interaction can be activated/deactivated by a virtual button (green bottom button) located around the gripper. When this button is activated it shows the ghost object right under the current position of the gripper. The ghost gripper is related to the ghost object and it is also a virtual representation of the real gripper. It is showed in red, as shown in Figure 3d and appears when the operator points at a position on the tabletop where there would be a risk of some part of the gripper colliding with a neighboring object.

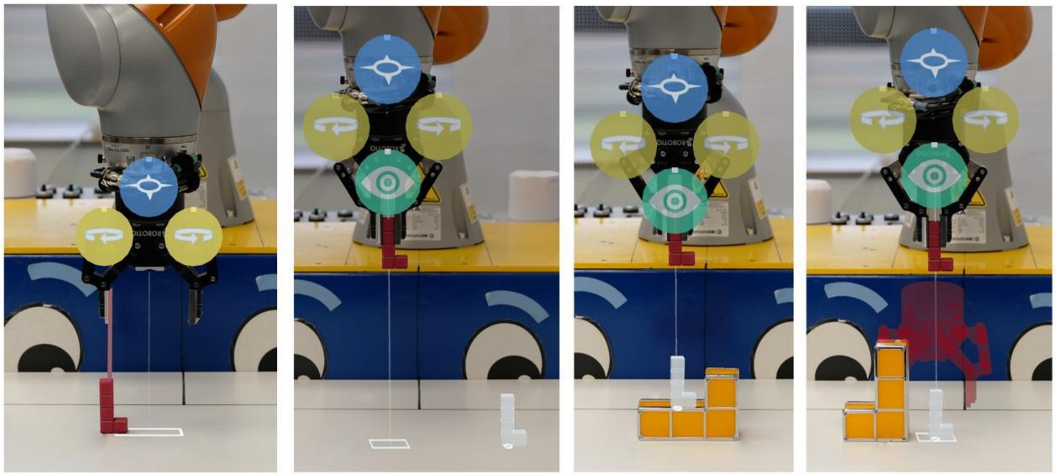


Figure 3: (a) Virtual Extension and grip region showing occlusion. (b) Ghost object following the cursor. (c) Ghost object recognizes other object's pose and position. (d) Ghost object and ghost gripper.

4 EXPERIMENT

We present an experiment where we test different presentations of AVC (Basic, Advanced, and No Cues) in co-located robot teleoperation to analyze their effect on task performance for manipulation and grasping tasks. While designing the different cues, we realized that Basic and Advanced Cues can affect task performance (time, accuracy, errors, subjective measures) on different levels. Additionally, our measure of accuracy could also hint at improvements in depth perception, as shown in previous experiments by Mather & Smith where depth cues improved accuracy and speed of performance [32].

4.1 Hypotheses

H1: We hypothesized that task performance in teleoperation would be the highest when Advanced Cues are used compared to Basic Cues or teleoperation without visual cues, as they explicitly show information of objects in the environment and assist the operator in sensing the environment. Specifically, high accuracy would be achieved with the lowest amount of time and errors across conditions.

H2: We hypothesized that task performance in teleoperation would be higher when Basic Cues are used compared to teleoperation without cues, as having hints that provide some information of the environment are necessary when located at a certain distance of the workspace. Specifically, Basic Cues would have higher accuracy with lesser time and a lower amount of errors than teleoperation without cues.

H3: H3.1. We hypothesized that teleoperation without AVC would cause a high cognitive load, while using Basic or Advanced Cues would have a low cognitive load, as showing visual hints aids to understand better the robot and the environment around it. This avoids the demand that operators must "imagine or guess" the position of the robot gripper and the target objects. H3.2. Additionally, subjective measures of performance would also refect that Basic and Advanced Cues ease teleoperation with higher levels of usefulness, ease of use and learn, satisfaction, enjoyment and lower levels of concentration.

H4: We hypothesized that teleoperating a robotic arm would evoke different views on the presentation of AVC (Basic, Advanced, No Cues). Teleoperating a robot without cues would be perceived negatively, but Basic Cues would prompt neutral views of teleoperation and Advanced Cues would evoke positive views.

4.2 Participants

We recruited a total of 36 males and female participants, all of them students or university staff. They were randomly assigned to one of three experimental groups (No Cues, Basic Cues, Advanced Cues). Average age per condition was No Cues 29.42 (SD = 12.44), basic cues 29.25 (SD = 3.94), Advanced Cues 30.83 (SD = 3.38). The pre-test questionnaire revealed that 18 participants had some experience programming robots, and they were evenly distributed across conditions. Also, 15 participants reported to have some experience with the Microsoft HoloLens. Two participants reported to be colorblind
—one took part in the Basic Cues condition which presented a color map and reported no problems in distinguishing the real and virtual colors. All participants received 7 euros for their participation.

4.3 Experimental Design

We conducted a 3 x 1 between-participants experiment to evaluate how our design of AVC affects task performance in co-located robot teleoperation. Our independent variable is the presentation of AVC (Basic, Advanced, and No Cues). No Cues, i.e., absence of visual cues, served as a baseline condition. In all three conditions, we used the same multimodal interaction (Section 4.7). However, the interface (virtual buttons) was slightly changed to accommodate each design of cues, see Figure 4. In particular, the bottom blue

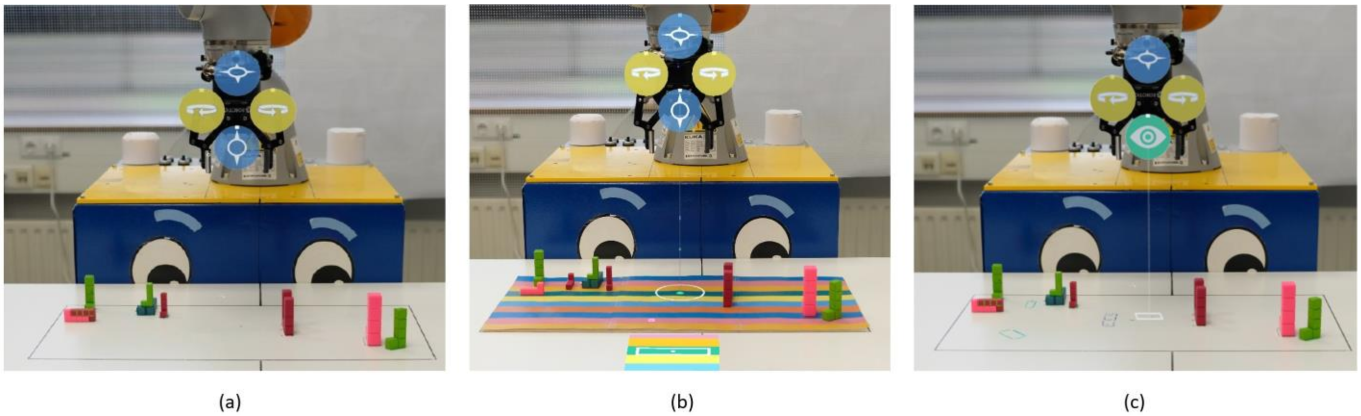


Figure 4: Experimental Conditions. (a) No Cues. (b) Basic Cues. (c) Advanced Cues.

button used in the No Cues and Basic Cues conditions was replaced by a button that shows the ghost object under the current position of the gripper on the tabletop.

We collected the following objective measures to evaluate task performance: time (in seconds) spent positioning the gripper to pick and place an object after pointing to the target position, accuracy in picking and placing (distance in mm to achieve the perfect position, measured by recording the position where the participant grasps or places an object compared to the position where the object is located). Our measure of accuracy is also related to depth perception since it measures the positioning of the gripper above an object, where a higher accuracy would be related to an improvement of distance and depth perception. Also, we collected the number of errors in teleoperation, i.e., participants crashing the gripper against the table or other objects, failing to pick a target object, touching the neighboring objects differently with some part of the gripper, and dropping the target object while picking or placing.

Additionally, we collected subjective measures of task performance by using the NASA Raw-TLX (RTLX) [20] to evaluate workload. This version does not include the weighting of factors and pairwise comparisons. It is a common simplification and has proven to be robust and accurate [19].

In addition, we applied a questionnaire inspired by the USE [29] and Feeling of Flow [16]. We reformulated some items from the questionnaires to ft a human-robot collaborative scenario and evaluated their internal consistency using Cronbach alpha. The dimensions evaluated were enjoyment (4 items, Cronbach $\alpha = .729$), concentration (4 items, Cronbach $\alpha = .762$), usefulness (6 items, Cronbach $\alpha = .873$), ease of use (8 items, Cronbach $\alpha = .895$), ease of learning (4 items, Cronbach $\alpha = .92$), and satisfaction (4 items, Cronbach $\alpha = .887$). In order to understand the participants' views, we asked them to write three words to describe their feelings when controlling the robot and three words to describe the system (see sentiment analysis).

We used a mixed-methods approach to evaluate the data from the experiment. The objective data were analyzed using a one-way ANOVA considering the presentation of AVC as the between-subjects factor. We assessed the normal distribution of our variables using the Shapiro-Wilk test, homogeneity of variances using Levene's Test, and used Bonferroni correction to control Type I errors. Subjective data were analyzed using Kruska-Wallis test with Dunn's post-hoc test using Bonferroni correction. In order to evaluate the participants' views of teleoperation with the different presentations of AVC, we performed a sentiment analysis following a relatively simple approach. We identified the sentiment scores of the three words participants used to describe the system and their feelings. However, not every participant provided three words. The language the participants used was German, and therefore, we used SentiWs, a German-language dictionary for sentiment analysis [40]. If the word was not present in the corpus, the word was not included in the calculations. Once the polarity score had been determined, we calculated the sentiment per participant following a counting method, used in lexicon methods [25]. We reported the total number of sentiments per condition and not the aggregated results to avoid outweighing sentiments, e.g., a positive sentiment with a high score can outweigh two negatives with a low score. In order to look beyond these metrics and better understand the impressions of participants, we combined the sentiment analysis with thematic content analysis following Anderson's approach [3]. As criterion for relevancy, we established that a minimum of 3 different participants should have written the same term or a synonym of it. First, a researcher analyzed and grouped the terms into themes. This analysis was later rated by another researcher.

4.4 Task

The experiment consisted of a pick-and-place task with three subtasks and a similar training task to start the interaction. Each subtask (as well as the training task) involved picking and placing one particular green "L-shaped" object (O1, O2, O3), without crashing or touching the neighboring objects. Once the three objects were placed in their target positions the task was concluded, and the process was repeated 2 more times, resulting in 3 trials in total.

The participants wore the Microsoft HoloLens and were seated in front of the table at 160 cm from the workspace. They were instructed not to move or tilt their heads to the side to change their visual angle. This was done explicitly to evaluate the effect of the AVC in the operator's visual perspective.

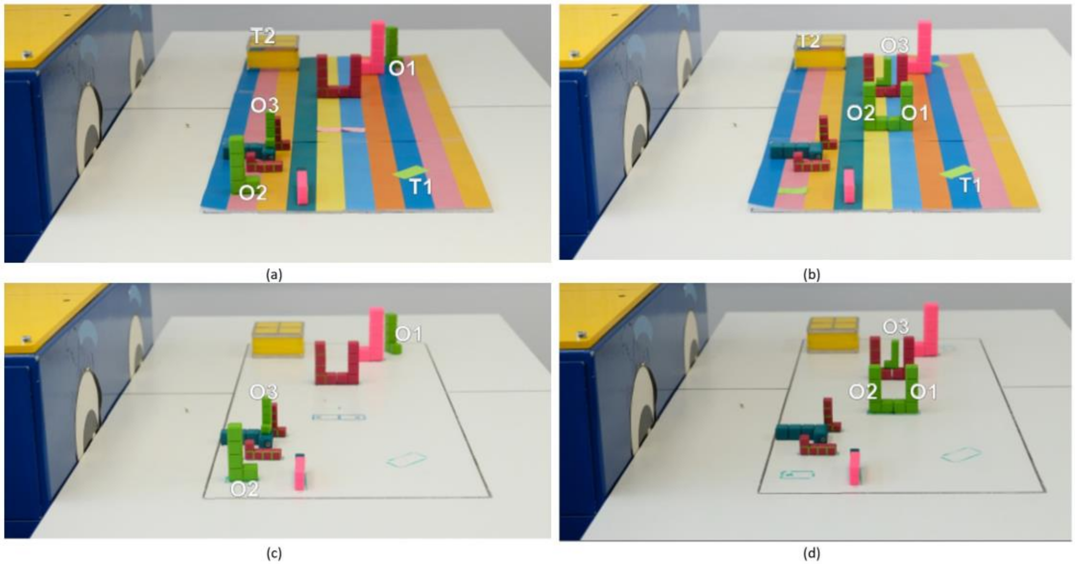


Figure 5: Side view of the experimental task: the numbers represent the order in which each piece would be picked-and-placed. T1 and T2 relate to the positions for the training tasks. (a) Basic Cues workspace before the task. (b) Basic Cues workspace with the completed task. (c) Advanced Cues and No Cues workspace before the task. (d) Advanced Cues and No Cues workspace with the completed task.

Figure 5 shows the workspace with "L-shaped" objects in different sizes and colors. Participants only manipulated the green objects and the numbers represent the order in which these needed to be picked and placed. Each object involved a different type of challenge and the target positions for placing were marked on the workspace, as shown in Figure 5b and 5d. O1 was slightly rotated to the side, requiring the operator to rotate the gripper and align its fingers accordingly. Additionally, participants needed to avoid touching or moving the large pink object next to it. The target position of O1 was in the center of the workspace with (at that time) no other objects in the vicinity. O1 still needed to be rotated further to match the marking. O2 was surrounded by a red and pink object of smaller size, posing some risk of collision on two sides but leaving space for movement around the other surrounding areas. On top of that, the pink object partially occluded O2. The target position of O2, from the operator's perspective, was behind the target position of O1 (Figure 5b, 5d) so O1 and O2 formed a U-shaped figure. O1 thereby occluded the target position and posed risk of crashing onto O1. O3 was smaller in size than O1 and O2, and it was located on top of a blue object (change of height during picking) with two red objects of smaller size in the vicinity (risk of collision). The target position of O3 was in between the U-shaped figure formed by two red objects (Figure 5b, 5d). In order to successfully place O3, the gripper needed to be precisely positioned and gave little space for error, so that it did not touch or crash onto the red objects.

4.5 Procedure

The experiment lasted approximately 60 minutes and consisted of 5 phases (1) introduction, (2) calibration, (3) familiarization and training phase (5 - 15 minutes), (4) task (7 - 30 minutes), (5) subjective questionnaires and interview.

(1) Participants were given an overview of the devices to be used, the goal of the experiment was explained and they were handed a consent form. (2) Then, they were shown an introductory video on how to wear the HoloLens and proceeded to calibrate it. (3) After that, they were shown a set of short clips regarding the interaction modalities, the interface, and the visual cues to be used. Participants on the No Cues condition were only shown the first two sets of videos. Then, they proceeded to perform the training task. During this phase, participants were instructed to use all the commands and virtual buttons to familiarize themselves with the interface and the interaction itself. Once they felt comfortable with the interface and interaction, we instructed them to take off the HoloLens and take a short break while filling out a background questionnaire. This was performed to avoid fatigue effects derived from extended periods of wearing the HoloLens. (4) Next, participants completed the task with three repetitions. (5) Finally, participants were given post-experiment questionnaires that evaluate their experience, and we conducted a short interview.

4.6 Implementation and Devices

We used a Kuka iiwa 7 R800 lightweight robotic arm, designed especially for HRI, with an attached Robotiq adaptive 2-Finger Gripper, both controlled via the Kuka iiwa's control unit. It has a reach of 80 cm, a payload of 7 kg and torque sensors in each joint for a safe collaboration with humans. For the user interface, we used the Microsoft HoloLens 1, which is equipped with inside-out tracking, an HD video camera and microphones that allow the use of head movement, and speech and gestures as input options. The field of view is $30^{\circ} \times 17.5^{\circ}$ with a resolution of 1268×720 px per eye. The application running on the HoloLens was programmed

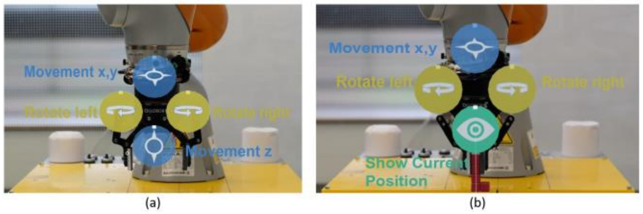


Figure 6: (a) Robot controls for the basic cues: yellow buttons rotate the gripper, top blue button moves the gripper on x,y axis and bottom blue button moves the gripper on the z axis. (b) Robot controls for the Advanced Cues: yellow buttons rotate the gripper, top blue button moves the gripper on the x,y axis, and the green button shows the current position of the gripper on the tabletop.

with Unity 2018.1.2 and the HoloToolkit Plugin [35], as well as, the Vuforia Plugin [1], using C#.

The Kuka iiwa and the HoloLens communicate via User Datagram Protocol (UDP) in a local network. The control unit of the Kuka iiwa runs a specially designed backend application that processes the received messages, moves the robot according to the receiving data, and returns to its current status. In order to display the augmentation at the right pose in the real world, we use the Vuforia plugin to detect a marker attached to the robot and align it by placing a virtual copy in the HoloLens' environment at the detected position.

To communicate pose data from the HoloLens to the control unit of the robot, we first converted the pose from Unity's left- handed coordinate system to the robot's right-handed coordinate system and use common length units (mm). Then, this cartesian position, along with rotation and velocity is sent via UDP. The backend recognizes the command and allows the robot to plan the movement via internal inverse kinematics to then moving the robot arm physically.

For a full object-pose recognition, first, a 2D object detection and the 6D pose recognition is necessary. In order to enable this we followed Sundermeyer et al.'s [44] approach. However, we did not actively scan and perform object-pose recognition for the experiment. After several trials, we had a failure of pose recognition at a mean of 7.5% overlap between objects, which was not suitable for our workspace with several objects that occlude each other. Thereby to keep the accuracy of object-pose recognition stable along the experiment, we decided to hardcode the object-pose from the workspace in the system. This avoided having different experiences in the use of Advanced Cues, which depend on the accuracy of object-pose recognition. Further, the analysis of the accuracy of object-pose recognition was not a subject of study in this paper but was the mean to develop our Advanced Cues.

4.7 Robot Control and Interaction

This research is part of a larger project in which we also investigate hands-free interaction. Therefore, in order to control the robotic arm for picking and placing tasks, we opted for multimodal hands-free interaction with a combination of head-movements for pointing and speech for committing an action. These two modalities have been proved to fit well together [28], are common interaction modalities in AR/Mixed-Reality (MR) which are natively supported by the HoloLens, and are a natural way to interact when giving commands. As our contribution of this paper lies within the Visual Cues, we kept the interaction modality stable to avoid confounding effects.

Our interface design shows virtual robot controls anchored to the gripper of the robot (Figure 6). This allows operators to have these controls not only within the user's line of sight but also precisely in the current operator's focus. Thereby, an interface anchored to the operator's current focus avoids unrequired attention shifts, which are common when using a control pad or an external screen for teleoperation.

In our scenario, this interaction involves pointing at a position on the tabletop where a target object is located, commanding the robotic arm to move the robot to that position by saying the command "Move", finding an adequate grasping position and adjusting the position of the gripper (using the top blue button, Figure 6a). Once the Operator has determined that the gripper is located at an adequate gripping position, the command "pick" commits the action of grasping, i.e. the gripper moves in the robot's z-axis towards the tabletop, closes its fingers, and moves to the initial position on the z-axis (height) again. Once the object has been successfully grasped, the operator points at another target position, commands the robotic arm to move to that position, adjusts the position, and then places the grasped object in the target position by saying the command "place".

If there is no previous knowledge of the environment, the operator needs to manually control and adjust the gripper's height (z-axis, up/down) to find the adequate grasping position. This is achieved using the blue bottom button as seen in Figure 6a. Also, the commands to pick and place objects function slightly differently when using object-pose recognition. When the workspace and the objects around are known by the system, the gripper can move automatically towards the target, after pointing to the object and repositioning on the x,y plane. The operator then says the command "pick" or "place" and the robot will move towards the target object, close the fingers and move to the initial height (z-axis).

On top of the commands, we added two sets of virtual buttons anchored to the real gripper, which are activated by dwelling (Figure 6). The yellow buttons rotate the robot's gripper at a pace of 10 degrees per second and allow the operator to ensure that the fingers of the gripper match the affordances of the object to grasp it. The top blue button controls the gripper's movements on the robot's x,y axis (right, left, back, front). It follows the operator's head movements and is intended to be used for fine adjustment of the position of the gripper. It can move from 0 to 800 mm/s following the operator's head movements, and the operator controls the speed by directing the head gaze further from the gripper for faster movements, and closer to the gripper for slow movements, e.g., if the operator gazes at the corners of the tabletop, the gripper will move faster, and it will move slower when the operator gazes to the sides of the gripper. The bottom blue button is only present when there is no knowledge of the environment (Basic Cues), and it controls the gripper movements on the z-axis (up/down) to manually adjust the height when approaching an object on the tabletop. It behaves in a similar manner as the x,y movement button and follows the operator's head movements. The operator can also control the speed by directing the gaze further or closer to the gripper, e.g., the gripper moves faster if the operator gazes below the table or 30 cm above the gripper, and it moves slower if the operator gazes somewhere close to the center of the gripper. When a button is activated the rest of buttons are hidden to avoid the Midas-touch problems [24], especially when the robotic arm is following head movements. This bottom blue button is replaced by a green button in the Advanced Cues. Figure 6b shows the ghost object at the current position of the gripper.

4.8 Results

4.8.1 Objective Measures. We analyzed our data with a One-Way Anova with type of cues (No cues, Basic Cues, Advanced Cues) as the between-subjects factor and time, accuracy, and errors as dependent variables.

For time, the data were normally distributed, as assessed by the Shapiro-Wilk test ($\alpha = .05$). Homogeneity of variances was asserted using Levene's Test which showed that equal variances could be assumed (p = .124). The mean task-time was statistically significant for the different levels of presentation of cues, Figure 7, $F(2, 33) = 13.62$, $p < .001$, $\eta p^2 = .452$. Post-hoc pairwise comparisons (Bonferroni adjusted) show that the time spent for picking and placing decreased significantly with Advanced Cues (M=29.35 s, SD=8.67s) when compared to Basic Cues (M=62.02 s, SD=22.70), ($p < .001$) and to the No Cues condition (M=57.86 s, SD=15.66), (p = .001) but no significant difference was found between the No Cues and Basic Cues condition.

For accuracy, the data were normally distributed, as assessed by the Shapiro-Wilk test ($\alpha = .05$). Homogeneity of variances was asserted using Levene's Test, which showed that equal variances could be assumed (p = .414). The mean accuracy differed statistically for the different levels of type of cues, Figure 7, $F(2,33) = 6.61$, p =.004, $\eta p^2 = .29$. Post-hoc pairwise comparisons, applying Bonferroni correction, revealed a statistically significant improvement in accuracy when using Advanced Cues (M=10.97 mm, SD=1.84) compared to No Cues (M=13.54 mm, SD=2.70), (p = .021), and a significant improvement when using Basic Cues (M=10.54 mm, SD=1.89), (p = .006) compared to No Cues. No significant difference was found between Basic Cues and Advanced Cues.

Figure 7 shows a plot of the amount of time spent vs accuracy. Here, we can see the improvement in time and accuracy, especially for the Advanced Cues.

For the number of errors, the data was normally distributed for Basic Cues and Advanced Cues, but not for the No Cues condition, as assessed by the Shapiro-Wilk test ($\alpha = .05$). As a result, we applied a non-parametric test statistic on the data, the Kruskal-Wallis, followed by pairwise comparisons with the Dunn's test. The Kruskal-Wallis test, which was corrected for tied ranks, was significant $\chi^2(2, N = 36) = 11.10$, p = .004. For pairwise comparisons, the post- hoc Dunn's test showed significant differences for Advanced Cues (N=12, Md=3) compared to Basic Cues (N=12, Md=6) with p=0.008, and compared to No Cues (N=12, Md=6), with p=0.017 (Bonferroni corrected). However, no significant difference was found between No Cues and Basic Cues.

Table 1: Results of Subjective Measures per AVC

	No Cues		Basic Cues		Adv. Cues			
	M	SD	M	SD	M	SD	$\chi^2(2)$	p
Usefulness	3.18	1.50	5.08	1.01	5.39	0.67	10.88	.004
Ease of Use	4.38	1.51	4.85	1.09	5.47	.92	3.63	.16
Ease of Learning	4.94	1.78	6	.94	6.33	.54	4.47	.11
Satisfaction	4.79	1.51	5.56	.64	5.9	.61	4.56	.1
Enjoyment	5.96	.82	6.06	.69	6.33	.59	1.68	.43
Concentration	5.75	.8	6.19	.74	6.21	.6	2.61	.27

Table 2: Sentiment Analysis Results

	System		Feelings	
	No. Positive	No. Negative	No. Positive	No. Negative
No Cues	14	6	9	13
Basic Cues	16	5	10	10
Advanced Cues	21	2	18	6

4.8.2 Subjective Measures. We did not find significant differences of our designs of AVC after performing the Kruskal-Wallis test in ease of use, ease of learning, satisfaction, enjoyment, and concentration, see Table 1. However, we found significant difference in usefulness ($\chi^2(2) = 10.88$, p = .004). Post-hoc Dunn's test for pairwise comparisons, applying Bonferroni correction, showed a higher perception of usefulness when teleoperating the robot using Advanced Cues (p = .005) compared to teleoperation with No Cues, and teleoperation using Basic Cues (p = .048) compared to teleoperation with No Cues. No significant differences were found between Advanced and Basic Cues.

After evaluating the NASA RTLX using the Kruskal-Wallis test, we found significant differences between the different designs of AVC only in the performance factor ($\chi^2(2) = 7.93$, p = .019); see Figure 8. Post-hoc Dunn's Test for pairwise comparisons, applying Bonferroni correction, revealed a significant difference only when comparing Basic Cues (p = .026) to the No Cues condition.

4.8.3 Sentiment Analysis. We used sentiment analysis to evaluate the participants' views. In order to enable that, we considered the words participants used to describe the system and their feelings with different presentations of AVC. We found that in general, the system evoked a positive sentiment across conditions. However, we identified nuances in their feelings while performing the experiment dependent on the conditions each group of participants experienced (Table 2). Here, the No Cues condition prompted a negative sentiment which is supported by some comments by the participants during the interview, e.g. P15, "Something with the control system was counter-intuitive;" P17, "There is a visual feedback missing;" P20, "You have certain spatial imagination but that was not enough to find the correct position." Participants who performed the experiment using Basic Cues had different sentiments regarding their feelings while teleoperating the robot; half of them expressed positive sentiments and the other half expressed rather negative

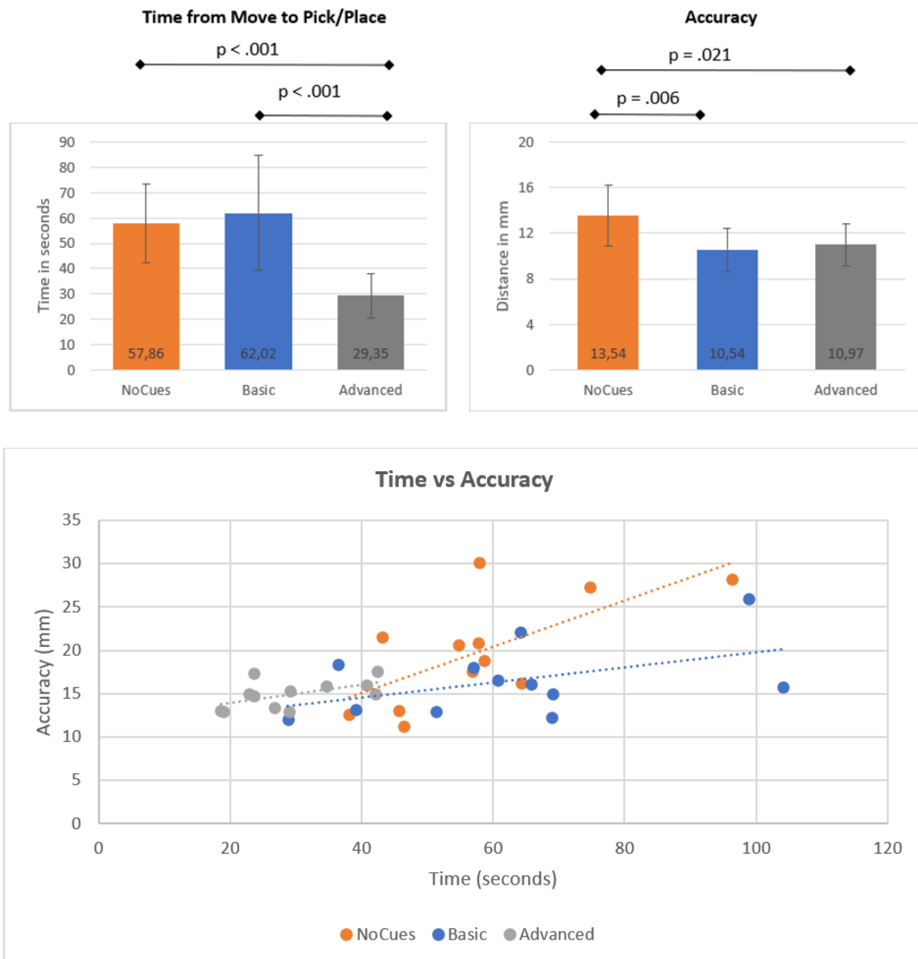


Figure 7: Top: Task performance measures of time and accuracy. Bottom: Plot chart of time vs accuracy

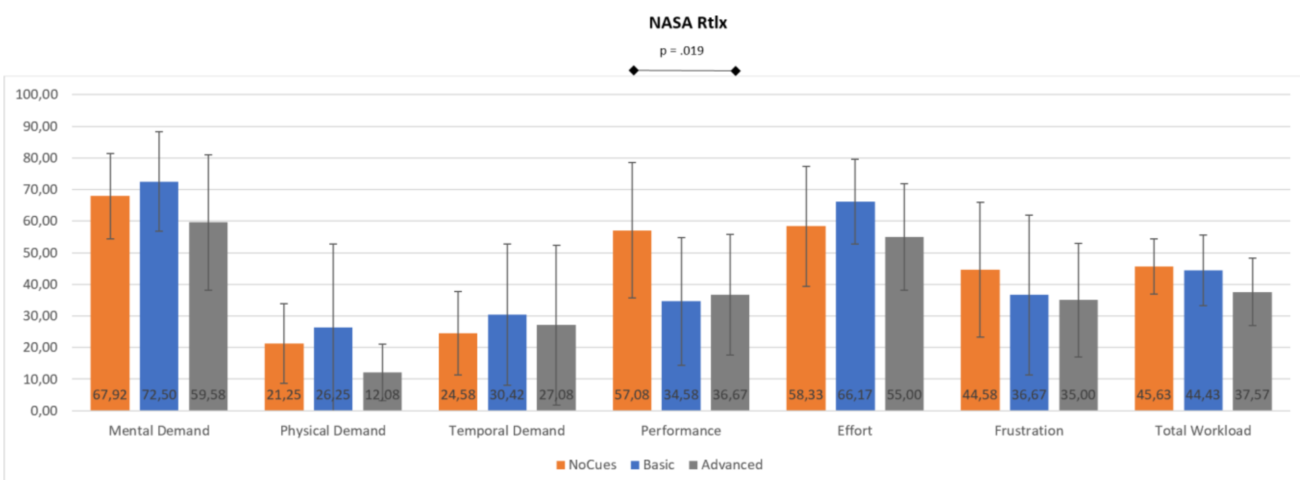


Figure 8: NASA Rtlx of Perceived Workload under the three conditions.

sentiments. Therefore, a further analysis is needed to understand the reasons behind this better (Section 4.8.4). The Advanced Cues aroused positive sentiments without any negative sentiment about teleoperation. Consequently, most participants expressed a positive sentiment about how they felt while teleoperating the robot using Advanced Cues.

4.8.4 Thematic Content Analysis. We considered it relevant to perform thematic content analysis since it could reveal underlying impressions of Basic and Advanced Cues. Hence, after the sentiment analysis, we looked for possible themes, and identified three that captured participants' impressions of both designs of AVC, and each design.

Both designs of AVC were considered innovative, interesting, and intuitive. The participant mentioned terms like "innovative" (3 times), "interesting" (3 times), and "intuitive" (4 times) on our Basic and Advanced Cues conditions. Teleoperating a robot using AR to augment the robot and its workspace was something novel to most of our participants, consequently evoking a sense of innovation. Despite having some participants with previous experiences with robots and AR/VR but not particularly robotic arms or teleoperation, it seemed that this combination (teleoperation with AR) was perceived as innovative. Related to innovation, participants referred to our system as "interesting" which might be also a consequence of this innovation and novelty factor, as presented in previous findings [10, 37]. Moreover, participants described our system as "intuitive". We relate this to our design of visual cues. P26 mentioned, "The ghost object was very easy to understand, as well as the red ghost gripper that appeared and showed collisions". However, P22, expressed, "It took me a while to trust the robot and the visuals. Once I did, it worked better for me." Also, we design our cues to be minimalistic and avoid overloading the operator's visual space.

Basic Cues were referred to as futuristic and complex. Under this condition participants used terms like "futuristic" (3 times) and complex (3 times). Most participants do not use robots in their day-to-day life and hence, teleoperating robotic arms with AR can still be perceived as something that is not quite in the present but rather forward-looking. Additionally, a complexity factor was mentioned; we attribute this to the design of Basic Cues since it had a rather large number of visual hints (6) visible at all times. This was also reflected when comparing Basic to Advanced Cues on our objective measures of task performance, e.g., a longer task time and higher number of errors. This hinted that for some participants, it was difficult to actively use all our Basic Cues, which could have led to ignoring some hints, e.g., P6 expressed, "You need to concentrate fully on moving the gripper, so you are looking at the gripper, and then also concentrate on the colors and that is hard." Participants' comments on Basic Cues were mostly directed towards the combi- nation of virtual and real color maps. In fact, these were considered the most helpful hints by 9 out of 12 participants from the Basic Cues condition, e.g., as P3 said, "Without the color map it would be impossible to perceive depth." P22 stated, "The real color map and the virtual rectangle showing the color where the gripper was located were the most helpful."

Advanced Cues were regarded as visionary. Participants mentioned mostly a visionary aspect (5 times) while using Advanced Cues. We attribute this to the "ghost object and gripper" hint and the occlusion capability added to our cues. Being able to see different future states of an object and determine what would be the best position based on those states adds a visionary aspect to the interaction. Here, the ghost object was mentioned by half of the participants as the most helpful Advanced Cue, e.g., P10 said, "The ghost object was the most helpful." This was followed by the laser-pointer (5 participants), about which P29 expressed, "The visualization of the object (ghost object) and the laser-pointer were the most helpful." Finally, the grip region was also mentioned by 4 participants, e.g., P8 said, "The visualization of the gripping area was really helpful for picking and for placing, as well as the ghost object." Participants rarely referred to only one hint from the Advanced Cues as helpful; it was mostly to a combination of them. A possible explanation is that they showed only two different cues at a time.

5 DISCUSSION

In this section, we compare our hypotheses to the results obtained from the experiment. First, we will discuss in detail our hypotheses (H1, H2) related to the objective measures of task performance. Further, we will discuss our subjective measures (H3) together with the qualitative evaluation from our sentiment and thematic content analysis (H4).

In **H1**, we hypothesized that teleoperation with Advanced Cues will have the highest task performance. This should be reflected by the highest accuracy, the shortest amount of time, and the lowest number of errors across conditions. Our results partially support H1. These indeed show that teleoperation with Advanced Cues improved task performance compared to teleoperation without any cues. However, when comparing teleoperation with Advanced Cues to teleoperation with Basic Cues, we only found significant improvements in terms of time and number of errors (not in accuracy). These results suggest that both designs of AVC improve and support accuracy in teleoperation similarly. In fact, improvements in accuracy in task performance suggest having a close relation to depth perception [32]. Related to these results, we further question if this trade-off of time to favor accuracy affects the perception of the interaction. When looking closely at the performance factor of cognitive load, we found similar performance perceptions between Basic and Advanced Cues. This could indicate that having a high level of accuracy influences self-evaluation of performance. Still, when aiming to optimize pick-and-place tasks in terms of time and errors, Advanced Cues have shown to have a lower number of errors and a shorter task execution.

Regarding **H2**, we hypothesized that teleoperation with Basic Cues would improve task performance in terms of time, accuracy, and the number of errors, compared to teleoperation without cues. We can partially support this hypothesis. Our results show similar performance in amount of time and errors, but a significant improvement in terms of accuracy. This leads us to the conclusion that in the No Cues condition, people were able to trade-off time and errors for reduced accuracy. From the perspective of the Basic Cues design, we also believe that a certain amount of visual clutter due to the number of hints presented could have affected time and errors. We showed 6 different cues that were always visible. Nonetheless, during the interview, participants only referred to 3 of them as most helpful. It may be possible that participants might have tried to use all of them at the beginning but eventually ignored some of the hints as the experiment progressed. Our thematic content analysis also supported this line of thinking, by our Basic Cues being referred to as "complex" (Section 4.8.4). We highlight that in a real-world setting, a physical colormap requires changes in the workspace. We evaluated one type of landmark (combination of the virtual and physical colormap) that proved to increase the accuracy and usefulness in task performance. This can make this change on the physical workspace worthwhile. Yet, other types of landmarks that blend-in with the physical space might provide similar benefits.

Our results from H1 and H2 build upon similar results from Brizzi et al. [4] who showed how presenting visual cues through AR improved the accuracy and efficiency in task performance in pick-and-place tasks. Also, our findings align partially with those of [31] in 2D images, where a higher number of depth cues improved accuracy in a shorter time. We indeed found improvements in accuracy and time for the Advanced Cues but a time trade-off for the basic cues. This difference might be a result of the 3D environment and the design of our cues. Still, this time trade-off is in line with previous research about visual search in complex environments, where the number of items in the visual space causes extra effort (longer time) in discriminating targets [39].

H3 regards our subjective measures. In H3.1., we hypothesized about a lower cognitive load of both designs of AVC. Our results partially support this hypothesis. We only found a significant difference in the perceived performance factor of the Nasa Rtlx when comparing Basic Cues to the No Cues condition. We believe that improvements in accuracy is also important for people's subjective assessment of their performance. H3.2 was also partially supported. Our results show only the significant differences in the perception of the usefulness of Basic and Advanced Cues compared to the No Cues condition. Our results did not show other significant differences in the rest of our subjective measures. A possible reasoning behind this can be the level of precision required for the task, added to the fact that most participants had never operated a robotic arm before. Expertise effects have been found to influence performance in teleoperation for pick-and-place tasks [4]. However, these can be leveled out when providing feedback through AR. Additionally, the metrics used could also have influenced these results. To better understand the subjective factor of our presentation of AVC, we explored the views of participants through sentiment and thematic content analysis.

In **H4**, we hypothesized about different views evoked by the presentation of our cues. A sentiment analysis supported this hypothesis in terms of the feelings expressed by the participants. Although participants were not able to compare the different presentations of AVC (because of a between-subjects experimental design), these were perceived differently and evoked contrasting views. The No Cues condition evoked a negative sentiment among participants, our Basic Cues aroused the same number of positive and negative sentiments, while the Advanced Cues gathered mostly positive sentiments. Moreover, our thematic content analysis tried to reveal hidden views of our designs of AVC. The positive sentiment is manifested with both designs of AVC as they were referred to as interesting, innovative, and intuitive. However, we are cautious with this result since the novelty factor might have played a definitive role. We consider the fact that participants referred to the Basic Cues as complex interesting. As discussed in H3, we attribute this to the number of cues presented to participants which could have hindered the positive sentiment that they could have evoked. Finally, our design of Advanced Cues was regarded as "visionary", which is also reflected by the high number of positive sentiments. This leads us to think that using previous knowledge of the environment to present explicit visual cues has a positive impact not only on objective performance but also on how the task and teleoperation is perceived.

6 LIMITATIONS

We acknowledge that our results might have limited generalizability due to our experimental design and our metrics. We decided to modify the existing subjective questionnaires (USE, Feeling of Flow) since they were constructed based on human-computer interaction systems and they did not completely transfer to HRI, e.g., teleoperating robotic arms and the use of AR. We recognize that different subjective questionnaires could have been used. However, we performed a sentiment and thematic content analysis that further explored the subjective aspect of the presentation of AVC.

Another limitation comes in terms of the type of interaction used. We recognize that multimodal hands-free interaction is a particular choice for robot teleoperation. However, we kept the interaction stable among conditions to minimize interaction effects and did not identify any major influence of the type of interaction in the evaluation of AVC. None of our participants reported problems with the general control of the robot that could be ascribed to the interaction modality. Head-gaze and speech are common interaction modalities in AR and MR and were natively supported by the technology we used, i.e., the Microsoft HoloLens 1. Additionally, this multimodal interaction modality is much simpler than learning to operate a robot control pad, which requires extensive training. On a different note, this work is particularly relevant for the growing body of research for hands-free interaction in HRI [28], [36].

The study is limited to the current AR technology, i.e., HoloLens 1. The tracking capabilities of the device are not precise enough to correctly allow for real-world objects to occlude the virtual ones. Besides, issues such as disparity planes and focal rivalry might arise prominently, as noted by Kruijf et al. [27], resulting in uncomfortable focus switches, although the location of virtual and real objects might be the same. Future or different AR technology might improve some of these issues. For example, see-through based AR (Varjo XR1) might reduce the visual clutter when virtual cues overlay or occlude real world objects, but it would need to be tested since it might arise other type of problems.

7 CONCLUSION

In this paper, we explored the design space of AR to enhance the operator's perception of the robot and its environment. We designed an interface with basic robot controls and used a multimodal interaction approach for teleoperation. This was used in combination with two designs of AVC (Basic and Advanced Cues) with the goal of providing a better understanding of depth and distances and thus improving task performance. In our first design, Basic Cues, we augmented the robot and the environment by providing a combination of real and virtual indirect hints. In our second design, Advanced Cues, we utilized previous knowledge of the environment —using object-pose recognition, to present explicit virtual hints that enhance the robot, the environment, and the objects in it. We evaluated our designs against a baseline (No Cues) in a user study where participants teleoperated a robotic arm to perform pick-and-place tasks that required some level of precision.

We hypothesized about the improvement of task performance at different levels in objective and subjective measures using both designs of AVC. Our findings show that both of our designs of AVC improve task performance in terms of accuracy, which is closely related to distance and depth perception compared to the baseline. However, our design of Advanced Cues also presents improvements in terms of time and number of errors. Our subjective measures hinted at nuances in participants' perception of self-performance and the usefulness of both designs of AVC. That is why we further explored the participants' views through sentiment analysis. Both designs of AVC gathered positive sentiments about the system. Nonetheless, we found differences in participants' feelings towards teleoperating a robot with Basic and Advanced Cues. In particular, our Advanced Cues evoked a positive sentiment while the Basic Cues aroused rather mixed sentiments. A possible reason behind this can be found in the results of our thematic content analysis. Here, we identified that our design of Basic Cues was regarded as complex, which we attribute to the number of hints presented to the operator.

All in all, our results show that both Basic and Advanced Cues offer a promising approach to enhance the visual space. They help to understand the relation between the environment and the robot in the workspace, thus, assisting the teleoperation of a robotic arm and increasing accuracy.

In future work, we will continue to explore the use of visual cues to enhance the operator's awareness of the workspace. We intend to go beyond improving depth perception and task performance and provide implicit visual cues that will help the operator to determine the changes and the potential problems derived from the interaction, e.g., objects falling, out-of-view objects, and partially occluded objects. A novel line of research has investigated robot deictic gestures used by social robots to provide instructions [54]. We consider it closely related to our augmented visual cues and we intend to extend this line of research in manufacturing environments with industrial robots.

ACKNOWLEDGMENTS

We would like to thank in particular our co-authors Franziska Rücker and Tim Dierks. Their tremendous work during their master studies made this paper possible.

This research is supported by the German Federal Ministry of Education and Research (BMBF, FKZ: 13FH011IX6).

REFERENCES

[1] 07/09/2020. SDK Download | Vuforia DeveloperPortal. https://developer.vuforia.com/downloads/sdk

[2] 2011. Appendix A: Human Visual Perception. In *Practical Image and Video Processing Using MATLAB®*, Oge Marques (Ed.). John Wiley & Sons, Inc, Hoboken, NJ, USA, 591–610. https://doi.org/10.1002/9781118093467.app1

[3] Rosemarie Anderson. 2007. Thematic content analysis (TCA). , 4 pages.

[4] Filippo Brizzi, Lorenzo Peppoloni, Alessandro Graziano, Erika Di Stefano, Carlo Alberto Avizzano, and Emanuele Rufaldi. 2018. Efects of Augmented Reality on the Performance of Teleoperated Industrial Assembly Tasks in a Robotic Embodiment. *IEEE Transactions on Human-Machine Systems* 48, 2 (2018), 197–206. https://doi.org/10.1109/THMS.2017.2782490

[5] Ravi Teja Chadalavada, Henrik Andreasson, Robert Krug, and Achim J. Lilienthal. 2015. That's on my mind! robot to human intention communication through on-board projection on shared foor space. In *2015 European Conference on Mobile Robots*, European Conference on Mobile Robots and Tom Duckett (Eds.). IEEE, Piscataway, NJ, 1–6. https://doi.org/10.1109/ECMR.2015.7403771

[6] E. Laurence Chalmers. 1952. Monocular and Binocular Cues in the Perception of Size and Distance. *The American Journal of Psychology* 65, 3 (1952), 415. https://doi.org/10.2307/1418762

[7] J.W.S. Chong, S. K. Ong, A.Y.C. Nee, and K. Youcef-Youmi. 2009. Robot programming using augmented reality: An interactive method for planning collision-free paths. *Robotics and Computer-Integrated Manufacturing* 25, 3 (2009), 689–701. https://doi.org/10.1016/j.rcim.2008.05.002

[8] Francesco Clemente, Strahinja Dosen, Luca Lonini, Marko Markovic, Dario Farina, and Christian Cipriani. 2017. Humans Can Integrate Augmented Reality Feedback in Their Sensorimotor Control of a Robotic Hand. *IEEE Transactions on Human-Machine Systems* 47, 4 (2017), 583–589. https://doi.org/10.1109/THMS.2016.2611998

[9] James E. Cutting and Peter M. Vishton. 1995. Perceiving Layout and Knowing Distances. In *Perception of Space and Motion*. Elsevier, 69–117. https://doi.org/10.1016/B978-012240530-3/50005-5

[10] Andreas Dünser, Raphaël Grasset, Hartmut Seichter, and Mark Billinghurst. 2007. Applying HCI principles to AR systems design. (2007).

[11] Fatima El Jamiy and Ronald Marsh. 2019. Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality. *IET Image Processing* 13, 5 (2019), 707–712. https://doi.org/10.1049/iet-ipr.2018.5920

[12] H. C. Fang, S. K. Ong, and A.Y.C. Nee. 2012. Interactive robot trajectory planning and simulation using Augmented Reality. *Robotics and Computer-Integrated Manufacturing* 28, 2 (2012), 227–237. https://doi.org/10.1016/j.rcim.2011.09.003

[13] Markus Funk, Andreas Bächler, Liane Bächler, Thomas Kosch, Thomas Heidenreich, and Albrecht Schmidt. 2017. Working with Augmented Reality?. In *PETRA 2017 (ICPS: ACM international conference proceeding series)*, Unknown (Ed.). ACM, New York, NY, USA, 222–229. https://doi.org/10.1145/3056540.3056548

[14] Hind Gacem, Gilles Bailly, James Eagan, and Eric Lecolinet. 2015. Finding Objects Faster in Dense Environments Using a Projection Augmented Robotic Arm. In *Human-computer interaction - INTERACT 2015*, Julio Abascal (Ed.). LNCS sublibrary. SL 3, Information systems and applications, incl. Internet/Web, and HCI, Vol. 9298. Springer, Cham, 221–238. https://doi.org/10.1007/978-3-319-22698-9{_}15

[15] Yiannis Gatsoulis, Gurvinder S. Virk, and Abbas A. Dehghani-Sanij. 2010. On the Measurement of Situation Awareness for Efective Human-Robot Interaction in Teleoperated Systems. *Journal of Cognitive Engineering and Decision Making* 4, 1 (2010), 69–98. https://doi.org/10.1518/155534310X495591

[16] Jawaid A. Ghani and Satish P. Deshpande. 1994. Task Characteristics and the Experience of Optimal Flow in Human—Computer Interaction. *The Journal of Psychology* 128, 4 (1994), 381–391. https://doi.org/10.1080/00223980.1994.9712742

[17] L. L. Gong, S. K. Ong, and A. Y. C. Nee. 2019. Projection-based Augmented Reality Interface for Robot Grasping Tasks. In *Proceedings of the 2019 4th International Conference on Robotics, Control and Automation - ICRCA 2019*, Unknown (Ed.). ACM Press, New York, New York, USA, 100–104. https://doi.org/10.1145/3351180.3351204

[18] Scott A. Green, Mark Billinghurst, XiaoQi Chen, and J. Geofrey Chase. 2008. Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design. *International Journal of Advanced Robotic Systems* 5, 1 (2008), 1. https://doi.org/10.5772/5664

[19] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909

[20] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human mental workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, Amsterdam and Oxford, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[21] Hooman Hedayati, Michael Walker, and Daniel Szafr. 2018. Improving Collocated Robot Teleoperation with Augmented Reality. In *HRI'18*, Takayuki Kanda, Selma Ŝabanović, Guy Hofman, and Adriana Tapus (Eds.). Association for Computing Machinery, New York, New York, 78–86. https://doi.org/10.1145/3171221.3171251

[22] I. P. Howard. 2012. *Perceiving in Depth, Volume 1: Basic Mechanisms*. Oxford University Press. https://books.google.de/books?id=A26JAgAAQBAJ

[23] J. A. Gomer, C. H. Dash, K. S. Moore, and C. C. Pagano. 2009. Using Radial Outfow to Provide Depth Information During Teleoperation. *Presence: Teleoperators and Virtual Environments* 18, 4 (2009), 304–320. https://doi.org/10.1162/pres.18.4.304

[24] Robert J. K. Jacob. 1995. Eye Tracking in Advanced Interface Design. In *Virtual Environments and Advanced Interface Design*. Oxford University Press, Inc, USA, 258–288.

[25] Alistair Kennedy and Diana Inkpen. 2006. SENTIMENT CLASSIFICATION of MOVIE REVIEWS USING CONTEXTUAL VALENCE SHIFTERS. *Computational Intelligence* 22, 2 (2006), 110–125. https://doi.org/10.1111/j.1467-8640.2006.00277.x

[26] R. Kitchin. 2001. Cognitive Maps. In *International encyclopedia of the social & behavioral sciences*, Neil J. edt Smelser and Paul B. edt Baltes (Eds.). Elsevier, Amsterdam and New York, 2120–2124. https://doi.org/10.1016/B0-08-043076-7/02531-6

[27] Ernst Kruijf, J. Edward Swan, and Steven Feiner. 2010. Perceptual issues in augmented reality revisited. In *9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2010*, Tobias Höllerer (Ed.). IEEE, Piscataway, NJ, 3–12. https://doi.org/10.1109/ISMAR.2010.5643530

[28] Dennis Krupke, Frank Steinicke, Paul Lubos, Yannick Jonetzko, Michael Gorner, and Jianwei Zhang. 10/1/2018 - 10/5/2018. Comparison of Multimodal Heading and Pointing Gestures for Co-Located Mixed Reality Human-Robot Interaction. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 1–9. https://doi.org/10.1109/IROS.2018.8594043

[29] Arnold M. Lund. 2001. Measuring usability with the use questionnaire12. *Usability Interface* 8, 2 (2001), 3–6.

[30] Magic Leap. [n.d.]. Magic Leap One. https://www.magicleap.com/

[31] Zhanat Makhataeva and Huseyin Varol. 2020. Augmented Reality for Robotics: A Review. *Robotics* 9, 2 (2020), 21. https://doi.org/10.3390/robotics9020021

[32] George Mather and David R. R. Smith. 2004. Combining depth cues: efects upon accuracy and speed of performance in a depth-ordering task. *Vision Research* 44, 6 (2004), 557–562. https://doi.org/10.1016/j.visres.2003.09.036

[33] M. Alejandra Menchaca-Brandan, Andrew M. Liu, Charles M. Oman, and Alan Natapof. 2007. Infuence of perspective-taking and mental rotation abilities in space teleoperation. In *Proceedings of the 2007 ACM*, Cynthia Breazeal, Alan C. Schultz, Terry Fong, and Sara Kiesler (Eds.). ACM, New York, 271. https://doi.org/10.1145/1228716.1228753

[34] Microsoft. [n.d.]. Microsoft Hololens. https://docs.microsoft.com/en-us/hololens/hololens1-hardware

[35] Microsoft'. 2017. Mixed Reality Toolkit. https://github.com/microsoft/MixedRealityToolkit-Unity/releases

[36] Susanna Nilsson, Torbjörn Gustafsson, and Per Carleberg. 2007. Hands Free Interaction with Virtual Information in a Real Environment: Eye Gaze as an Interaction Tool in an Augmented Reality System. *Proceedings of COGAIN* (2007), 53–57.

[37] Thomas Olsson and Markus Salo. 2012. Narratives of satisfying and unsatisfying experiences of current mobile augmented reality applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (ACM Digital Library)*, Joseph A. Konstan (Ed.). ACM, New York, NY, 2779. https://doi.org/10.1145/2207676.2208677

[38] Yoon Jung Park, Hyocheol Ro, and Tack-Don Han. 2019. Deep-ChildAR bot. In *ACM SIGGRAPH 2019 Posters*. Association for Computing Machinery, [S.l.], 1–2. https://doi.org/10.1145/3306214.3338589

[39] Rajesh P.N. Rao, Gregory J. Zelinsky, Mary M. Hayhoe, and Dana H. Ballard. 2002. Eye movements in iconic visual search. *Vision Research* 42, 11 (2002), 1447–1463. https://doi.org/10.1016/S0042-6989(02)00040-8

[40] Robert Remus, Uwe Quasthof, and Gerhard Heyer. 2010. SentiWS - A Publicly Available German-language Resource for Sentiment Analysis.

[41] S. Stadler, K. Kain, M. Giuliani, N. Mirnig, G. Stollnberger, and M. Tscheligi. 2016. Augmented reality for industrial robot programmers: Workload analysis for task-based, augmented reality-supported robot control. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 179–184. https://doi.org/10.1109/ROMAN.2016.7745108

[42] Bennett L. Schwartz and John H. Krantz. 2019. *Sensation & perception* (second edition ed.). SAGE, Los Angeles.

[43] Jose Aparecido Da Silva. 1985. Scales for Perceived Egocentric Distance in a Large Open Field: Comparison of Three Psychophysical Methods. *The American Journal of Psychology* 98, 1 (1985), 119. https://doi.org/10.2307/1422771

[44] Martin Sundermeyer, Zoltan-Csaba Marton, Maximilian Durner, Manuel Brucker, and Rudolph Triebel. [n.d.]. Implicit 3D Orientation Learning for 6D Object Detection from RGB Images. https://arxiv.org/pdf/1902.01275

[45] C. T. Swain. 1996. Integration of monocular cues to create depth efect. In *1997 IEEE international conference on acoustics, speech, and signal processing*. IEEE Comput. Soc. Press, 2745–2748. https://doi.org/10.1109/ICASSP.1997.595357

[46] Laurence P. Tidbury, Kevin R. Brooks, Anna R. O'Connor, and Sophie M. Wuerger. 2016. A Systematic Comparison of Static and Dynamic Cues for Depth Perception. *Investigative ophthalmology & visual science* 57, 8 (2016), 3545–3553. https://doi.org/10.1167/iovs.15-18104

[47] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafr. 2018. Communicating Robot Motion Intent with Augmented Reality. In *HRI'18*, Takayuki Kanda, Selma Ŝabanović, Guy Hofman, and Adriana Tapus (Eds.). Association for Computing Machinery, New York, New York, 316–324. https://doi.org/10.1145/3171221.3171253

[48] Michael E. Walker, Hooman Hedayati, and Daniel Szafr. 3/11/2019 - 3/14/2019. Robot Teleoperation with Augmented Reality Virtual Surrogates. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 202–210. https://doi.org/10.1109/HRI.2019.8673306

[49] Stuart Walker and Ed Dorsa. 2001. Making Design Work: Sustainability, Product Design and Social Equity. *The Journal of Sustainable Product Design* 1, 1 (2001), 41–48. https://doi.org/10.1023/A:1014412307092

[50] X. Wang, S. K. Ong, and A. Y. C. Nee. 2016. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing* 4, 1 (2016), 1–22. https://doi.org/10.1007/s40436-015-0131-4

[51] Atsushi Watanabe, Tetsushi Ikeda, Yoichi Morales, Kazuhiko Shinozawa, Takahiro Miyashita, and Norihiro Hagita. 2015. Communicating robotic navigational intentions. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Wolfram Burgard (Ed.). IEEE, Piscataway, NJ, 5763–5769. https://doi.org/10.1109/IROS.2015.7354195

[52] Marcus R. Watson and James T. Enns. 2016. Depth Perception. In *Reference module in neuroscience and biobehavioral psychology*, John Stein (Ed.). Elsevier, [Place of publication not identifed]. https://doi.org/10.1016/B978-0-12-809324-5.06398-7

[53] Rong Wen, Wei-Liang Tay, Binh P. Nguyen, Chin-Boon Chng, and Chee-Kong Chui. 2014. Hand gesture guided robot-assisted surgery based on a direct augmented reality interface. *Computer methods and programs in biomedicine* 116, 2 (2014), 68–80. https://doi.org/10.1016/j.cmpb.2013.12.018

[54] Tom Williams, Matthew Bussing, Sebastian Cabrol, Elizabeth Boyle, and Nhan Tran. 3/11/2019 - 3/14/2019. Mixed Reality Deictic Gesture for Multi-Modal Robot Communication. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 191–201. https://doi.org/10.1109/HRI.2019.8673275

[55] Bang Wong. 2011. Color blindness. *Nature methods* 8, 6 (2011), 441. https://doi.org/10.1038/nmeth.1618

7.8 Paper VIII. Does One Size Fit All? A Case Study to Discuss Findings of an Augmented Hands-Free Robot Teleoperation Concept for People with and without Motor Disabilities.

Arévalo Arboleda, Stephanie; Becker, Marvin; and Gerken, Jens. 2022. Does One Size Fit All? A Case Study to Discuss Findings of an Augmented Hands-Free Robot Teleoperation Concept for People with and without Motor Disabilities. Technologies, 10(1), 4. DOI:10.3390/technologies10010004.

technologies

Article

Does One Size Fit All? A Case Study to Discuss Findings of an Augmented Hands-Free Robot Teleoperation Concept for People with and without Motor Disabilities

Stephanie Arévalo Arboleda *, Marvin Becker and Jens Gerken *

The Human-Computer Interaction Group, Department of Media Informatics and Communication, Westphalian University of Applied Sciences, 45897 Gelsenkirchen, Germany; marvin.becker@w-hs.de
* Correspondence: stephanie.arevalo@w-hs.de (S.A.A.); jens.gerken@w-hs.de (J.G.)

Abstract: Hands-free robot teleoperation and augmented reality have the potential to create an inclusive environment for people with motor disabilities. It may allow them to teleoperate robotic arms to manipulate objects. However, the experiences evoked by the same teleoperation concept and augmented reality can vary significantly for people with motor disabilities compared to those without disabilities. In this paper, we report the experiences of Miss L., a person with multiple sclerosis, when teleoperating a robotic arm in a hands-free multimodal manner using a virtual menu and visual hints presented through the Microsoft HoloLens 2. We discuss our findings and compare her experiences to those of people without disabilities using the same teleoperation concept. Additionally, we present three learning points from comparing these experiences: a re-evaluation of the metrics used to measure performance, being aware of the bias, and considering variability in abilities, which evokes different experiences. We consider these learning points can be extrapolated to carrying human–robot interaction evaluations with mixed groups of participants with and without disabilities.

Keywords: robot teleoperation; augmented reality; learning points; case study; hands-free interaction; people with motor disabilities

check for updates

Citation: Arevalo Arboleda, S.; Becker, M.; Gerken, J. Does One Size Fit All? A Case Study to Discuss Findings of an Augmented Hands-Free Robot Teleoperation Concept for People with and without Motor Disabilities. *Technologies* **2022**, *10*, 4. https://doi.org/10.3390/technologies10010004

Academic Editors: Jeffrey W. Jutai and Luc de Witte

Received: 7 December 2021
Accepted: 31 December 2021
Published: 6 January 2022

1. Introduction

Robots and assistive devices have taken a caregiving role for the aging population and people with disabilities [1–3]. For people with physical impairments, such as motor disabilities, assistive robotics provides a set of opportunities that can enhance their abilities. People with Motor Disabilities (PMD) have restricted mobility of the upper and/or lower limbs and could perform reaching and manipulation tasks using robotic arms. An example of this is the Lio robot, which is presented as a "multifunctional arm designed for human-robot interaction and personal care assistant tasks" and has been used in different assistive environments and health care facilities [4].

Many PMD maintain their ability to use eye gaze, head movements, and speech and use those abilities to interact with other people and the environment. These abilities can be used as hands-free input modalities which can be suitable for robot control [5]. Further, a multimodal approach to robot teleoperation may complement the use of some modalities. For instance, head movements complement eye gaze for pointing and can be coordinated with speech [6]. Park et al. [7] proposed using eye gaze and head movements as a hands-free interaction to perform manipulation tasks with a robotic arm. Added to multimodality in teleoperation, augmented reality (AR) has gained popularity in robotics as a medium to exchange information, primarily using Head-Mounted Displays (HMDs) [8]. Moreover, using AR in robot teleoperation allows to enhance the operator's visual space and provide visual feedback within the operator's line of sight.

The experiences of people with and without disabilities may vary greatly, even when handling an experiment under the same environmental and instrumental conditions. These changes derive from the differences in abilities from each group. Culen and Velden [9] discuss methodological challenges when using participatory design with vulnerable groups derived from differences in cognitive and physical abilities. Moreover, Shinohara et al. [10] specified crucial differences about the use and perception of technology from people with and without disabilities. PMD may have different needs and expectations even when sharing a diagnosis [11] or having a common goal, e.g., at sheltered workshops [12]. We build upon those findings to present a perspective from an Individual with Motor Disabilities (IMD) about a teleoperation concept that was already tested with participants without disabilities [13].

The contribution of this paper is to provide insights from an IMD about a teleoperation concept of a robotic arm using AR (Microsoft HoloLens 2 [14]) in a pick-and-place task. We briefly present our interaction concept to allow for a detailed description of the experience of our participant. Further, we put that experience in context with those of people without disabilities. This allows us to present some reflections about similarities and differences in their experiences and learning points from carrying out such analysis. Our goal is to spread the perspectives of PMD and add to the body of knowledge about Disability Interaction (DIX), specifically to the third point in Holloway's DIX action plan by reporting in-the-wild studies testing solutions with people with disabilities [15].

2. An Augmented Hands-Free Teleoperation Concept

We use a multimodal hands-free approach through head orientation (yaw and pitch) and voice commands to teleoperate a robotic arm in a co-located space to perform pick-and-place tasks. This interaction concept, akin to "Put-that-there" [16], combines speech with deictic gestures. However, to adapt to the abilities of PMD, we considered head pose to point instead of hand gestures and combined it with voice commands. The interaction flow starts with head pointing to mark a position on a workspace and voice commands to commit an action, e.g., moving the robotic arm to an intended position and grasping/releasing objects, see Figure 1. We describe our interaction concept as follows:

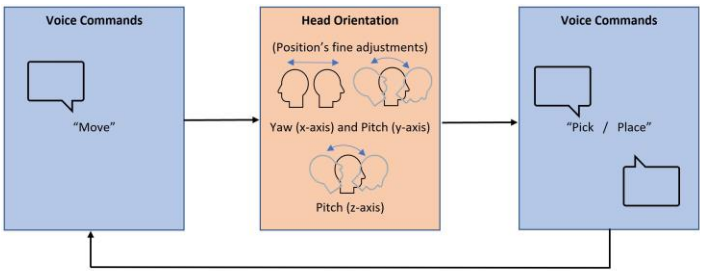


Figure 1. Diagram of the multimodal hands-free teleoperation.

1. The operator sees a cursor that follows the movements of the head (head pointing).
2. The operator uses head pointing to determine a position on the workspace, then the voice command "move" directs the robotic arm above the pointed position. The operator hears a click and the gaze pointer changes sizes to indicate that the command has been recognized. We set this to provide 1 second to keep or modify the position.
3. Once the operator has determined that the gripper is above a target, the command "pick" commits the action of grasping, i.e., the gripper moves in the robot's z-axis toward the tabletop, closes its fingers, and moves back to the initial position on the z-axis (height) again.
4. To release a grasped object, the command "place" moves the gripper in the z-axis toward the tabletop, opens its fingers to release the object, and moves to the initial position on the z-axis (height) again.

Additionally, our control interface presents two sets of virtual buttons anchored to the real gripper, which are activated by dwelling (Figure 2). The first set of buttons (yellow) rotate the robot's gripper (10 degrees per second) to the left or right. The second set of virtual buttons (blue) activate a precision mode where the robotic arm follows the operator's head yaw and pitch. The top blue button uses head's yaw to move the robotic arm left or right and pitch to move it front or back, i.e., x,y pane. The lower blue button uses head's pitch to move the gripper toward or away from the workspace, i.e., *z*-axis. This mode is intended to allow for fine adjustment of the position of the robot above a target object. The speed we use ranges from 0 to 800 mm/s and the operator can control the speed by directing the cursor further from the gripper (fast movement) or closer (slow movement) to it. For instance, if the operator moves the cursor to the workspace's borders, the gripper will move faster. When any of the buttons is activated, the rest of them are hidden to avoid Midas touch problems [17].

We added shortcut commands that perform some actions: the command "turn" rotates the gripper on/to 90° and the command "down", which executes the same action as the z-movement button, but it reduces the time needed to execute an action. It moves the gripper to half of the calculated distance from its current position to the tabletop.

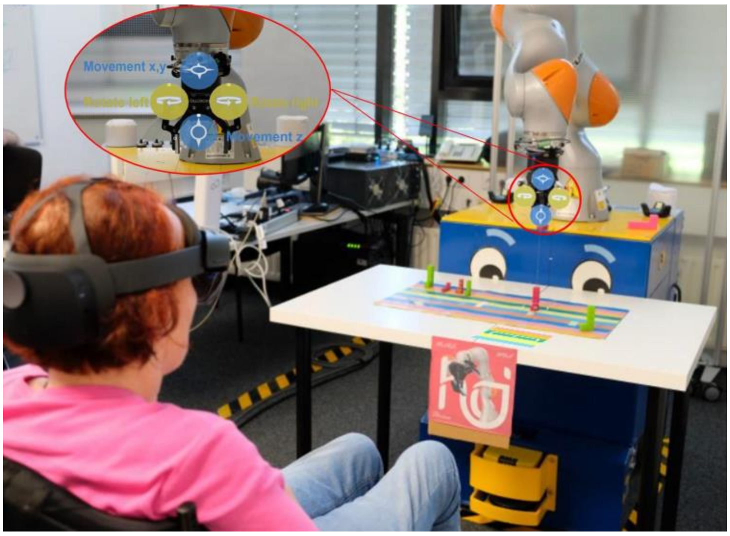


Figure 2. Study scenario with a view of the interface from the HoloLens.

We have previously reported the conceptualization and experiences of people without motor disabilities using two types of augmented visual cues (basic and advanced) in [13]. This case study builds upon that work and further investigates the use of basic augmented visual cues for PMD. Participants without disabilities did not express having problems with the input modalities (voice commands, head movements) or identifying the virtual elements in the environment.

3. A Case Study—Miss L.

We report on the experience of an IMD, Miss L., when teleoperating a robotic arm using our multimodal hands-free teleoperation concept. Miss L. has been involved in testing different assistive technologies for over 10 years and therefore possesses a certain level of experience. She has teleoperated different robotic arms, e.g., Kinova [18], Kuka iiwa [19], UR5 [20], tested different devices and sensors, as well as tried out various teleoperation interaction concepts. This allows her to provide broader insights and establish comparisons between different technologies and interactions grounded on her experience. Additionally, her assessments may be more critical and realistic than the ones from a novice user. In fact, we believe that her experiences minimize the influence of novelty and therefore allows a more ecological valid evaluation of the teleoperation interaction concept.

3.1. A Personal Background

Miss L. is an optimistic and kind person who enjoys reading and being outdoors. She studied journalism and worked as an editor. In 1988, she was diagnosed with multiple sclerosis. Today, she cannot move her legs and arms but maintains her energy and joyful personality. She describes her diagnosis as *"the disease of 1000 faces"* and says "I am lucky that I can still see, but I am shortsighted". Around 10 years ago, the German Department of Labour contacted her since they were looking for a person with motor disabilities to join a research team at the University of Bremen. That led her to move from Berlin to Bremen, where she joined the Institute for Automation Technology. Since then, she has been testing different assistive technologies developed at the University of Bremen and the Westphalian University of Applied Sciences.

She recalls her first experiences with the FRIEND (Functional Robot arm with user frIENdly interface for Disabled people) robot, which was a robotic arm mounted on a wheelchair [21]. It was designed to support people with motor disabilities to perform manipulation tasks in a workplace setting. She describes her experience with it, "I was a bit hesitant because it was very big and I was always sitting in it." She adds that, around that time, she was asked if she wanted to have a robotic arm at home which she declined as she preferred to have a "human assistant." However, she expressed, "Honestly, nowadays after all my experiences with robots, I would like to have a robotic arm at home. Caregivers have sometimes crazy mood swings, and a robotic arm does not have moods. It will not have good mood but there is no bad mood either. I would like very much to work with a neutral robotic arm".

3.2. Study

We present a study where Miss L. teleoperated a robotic arm to perform a basic pick-and-place task using our augmented hands-free teleoperation concept, as seen in Figure 3. The goal was to collect impressions to identify the strengths and downsides about the overall interaction concept and the technology used for a person with motor disabilities.

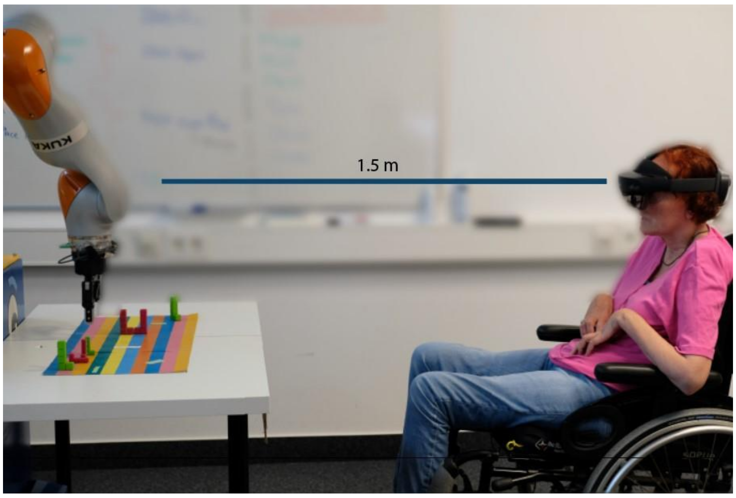


Figure 3. Side view of the scenario with Miss L. wearing the Microsoft HoloLens2 while teleoperating the robotic arm.

The task comprised teleoperating the robotic arm to pick an L-shaped object, rotating it, and placing it at a specific position on the workspace. The task was performed wearing the Microsoft HoloLens 2 at 1.5 m away from the workspace, see Figure 3. We collected subjective impressions using the NASA-TLX [22] to evaluate workload, the System Usability Scale (SUS) [23] for usability, the Godspeed Questionnaires Series [24]. Our goal was to determine the robot's perception, considering four dimensions (animacy, likeability, perceived intelligence, and safety), and carried out an interview. All the information was collected in German, analyzed in German, and then translated to English to report it in this paper.

The session lasted approximately three hours with several breaks in between. First, we introduced the session, the devices to be used, the goal of the study, asked for verbal consent, and carried out a short interview. Then, we showed an introductory video and proceeded to calibrate the HoloLens. Next, we performed a training task to familiarize her with the devices, modalities, and overall interaction. After that, we proceeded to perform the task, followed by the subjective questionnaires and a final interview.

3.3. Findings

We aimed to collect objective and subjective measures of performance of our teleoperation concept. However, we experienced a new series of unexpected challenges that influenced Miss L.'s experiences. She reported having difficulties visually identifying the cursor that follows head movements. The cursor was needed to activate the robot controls on the virtual menu (see Figure 2). Additionally, she mentioned that on that particular day, her voice volume was rather soft and she was not able to speak out very loud. This resulted in failures to recognize the voice commands.

3.3.1. Questionnaire Results

We present the questionnaire results together with some comments from Miss L. For some items of the different questionnaires, she elaborated on her line of reasoning behind them.

The results from the NASA-TLX workload showed mental demand = 50, weight (5); physical demand = 20, weight (1); temporal demand = 5, weight (0); performance = 75, weight (4); effort = 15, weight (2); frustration = 50, weight (3); with a total score of 50. These results show a relatively high frustration and mental demand compared to the other dimensions, see Figure 4.

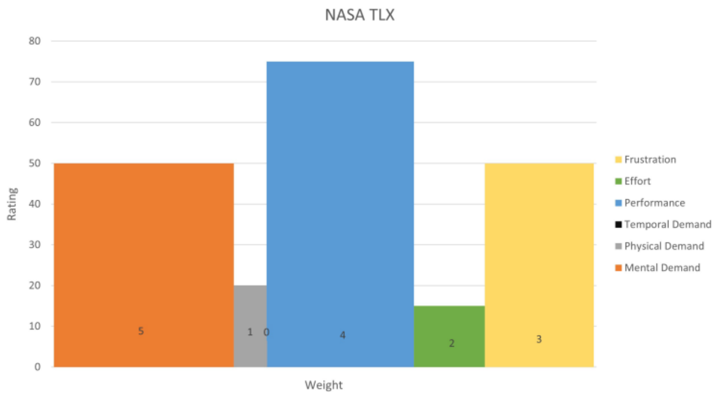


Figure 4. Participant's workload (NASA TLX plot).

The Godspeed Questionnaire prompted the following scores (5-point Likert scale). On the animacy scale (4.83), Miss L. emphasized the positive characteristics, e.g., "very alive", "very lively", "very lifelike", "very interactive". When Miss L. was asked about the robotic arm being mechanic or organic, she added, "Well, I always give a robot a soul, so I think the robot was indeed organic". Regarding the likeability dimension (5), she also expressed a rather positive attitude, e.g., "I like the robot very much", "It is a nice robot, it gets a 5", "It was a nice and spontaneous experience". In the perceived intelligence dimension (5), she added, "I do not think that in any situation the robot was unreasonable". Concerning safety (4), she mentioned, "Well, today I have not been on my greatest mood so I have not been relaxed".

The SUS Questionnaire (5-point Likert scale) prompted an overall score of 82.5 out of 100. During the questionnaire, she mentioned, "I do not think the system is unnecessary complex. I sometimes just need a while to understand something but after doing the task for a few times I understand it".

3.3.2. Interview

We present the answers, in verbatim, that Miss L. gave us during the interview. This interview took place after she teleoperated the robotic arm.

Question 1. What do you think about working with robotic arms? Have you had any problems in the past?

> "In my opinion there were never problems because there is always an emergency button, so I am not afraid of robotic arms. I do not get easily scared but it is important to have an emergency button. The hardest part is always to understand how everything works and what I should do. I do not always know what to do and I take some time until I fully understand everything."

Question 2. What do you think about controlling the robot by moving your head and using voice commands?

> "Usually, I like voice commands, but today I do not have a lot of energy to speak. So that is why the voice commands were not working today for me. The voice commands are fantastic, but today I am missing energy.
>
> I had another experiment where I had to move my head a lot, and it was very very very exhausting. But here I did not have to do that. I liked that everything felt compact."

Question 3. How did you feel controlling the robot? Can you mention three feelings that you experience while controlling the robot?

> "I think it's really cool! That is of course rather imprecise. But I felt like a part that contributed to the development. I thought that it was great.
>
> I was also feeling good, I found it interesting, but I felt frustrated. The buttons were constantly activating, and I was not able to deactivate them, thus, I could not use them. When the robot was following the movements of my head, I was not sure where to look. And I was not able to see where the cursor was. I am almost blind from the left eye, so maybe that is the reason behind it."

Question 4. Did you feel any pain or were you uncomfortable when using the HoloLens or teleoperating the robot?

> "No, I did not feel any pain or discomfort."

Question 5. We know that in previous experiments, you have used an external display to control the robot. How does it feel to have the robot controls within your line of sight?

> "When I need to look at the visual cues and then use the buttons to control the robot it feels like working with a different display. However, I found those elements easy to understand. What I really liked was having the buttons just in front of me, it was like a screen within the screen. I like that they are attached to the robot."

Question 6. You had virtual elements that indicated to you the position on the workspace such as different colors. Which element helped you the most?

> "At the beginning, I was not using any of those things. But when you started asking me about them, I realized that I could actually use them. I found the colors useful, it was easier to find where objects were with the colored stripes. I liked the virtual rectangle with the colors at the front. It showed me where the gripper is. Once I realized that, it got clear to me. But at the beginning I was just ignoring all of them.
>
> In other experiments when I have teleoperated a robot. I never knew how far the robot's gripper was relative to an object. However, when I saw that small virtual

"rectangle at the front showing me the color, it helped a lot to realize where I was on the workspace."

Question 7. Can you mention three words to describe the system and three words to describe your feelings while teleoperating the robotic arm?

"I think it is very good, interesting, and easy to understand. I have positive feelings towards it but I also felt frustrated because I could not see the cursor, so I could not activate the menu and then I could not perform the task at all."

3.3.3. Discussion

The NASA-TLX questionnaires showed a relatively high frustration and mental demand compared to the other dimensions. These results align with the comments expressed during the interview, where "frustrated" was one of the words used to describe the experience. We believe that this portrays the challenges that Miss L. experienced when teleoperating the robotic arm and thus had a decisive role in the workload.

The score of the SUS shows high usability of our overall interaction concept which, according to Bangor et al. [25], can be interpreted as high acceptability. These results align with those from the Godspeed questionnaire and hint at positive impressions of the overall interaction concept. However, we remark the high scores on animacy and perceived intelligence, as we expected that a collaborative industrial robotic arm and the teleoperation would evoke low scores on both dimensions.

Industrial robotic arms have anthropomorphic characteristics that relate to their functionality rather than their sociality, such as end-effectors similar to hands and fingers [26]. We consider that these anthropomorphic human-like features, e.g., an arm with soft edges and a gripper as end-effector, can indeed evoke high scores of animacy. Bartneck et al. [24] argue that animacy can be attributed to robots as they can react to stimuli and industrial robotic arms show physical behavior to manipulate objects in the environment in a similar way to that of a human hand. Additionally, we considered that teleoperation itself would decrease the perceived intelligence of the robot due to its lack of autonomy. However, perceived intelligence is more related to competence [24]. Although Miss L. experienced challenges performing the task, she did not attribute them to the robot but to the characteristics of her situation on that particular day. Further, seeing the evolution of the robots and technology used may have influenced the perceived intelligence of the robot. In Miss L.'s previous experiences, she teleoperated a larger robot with an extra external display and used a more demanding teleoperation concept as she expressed in the interview (Question 2). The use of AR to present basic robot controls within the operator's line of sight may have also simplified teleoperation and influenced the perceived intelligence of the robot.

4. Comparing Findings from People with and without Motor Disabilities

We designed an interaction concept that both people with and without disabilities could use. We aimed for simplicity in design and presented basic robot controls that novice users could quickly learn without further training in robotics. However, we encountered new challenges that an IMD faced, which we did not consider in previous studies with people without disabilities [13]. We present these findings and the challenges that Miss L. experienced in two topics, hands-free multimodal teleoperation and augmented visual cues.

In [13], participants highlighted and often used the terms *"useful and precise"* to describe the system. We found those words also mentioned by Miss L. However, we believe that this needs to be put in context. Miss L. found certain virtual elements useful, especially to improve depth perception as it can be seen in her answers to Question 6. She did not refer to the whole teleoperation concept as useful, but she established comparisons to other teleoperation concepts that she found exhausting and she described our teleoperation as compact (Question 2). Regarding finding the concept "precise", Miss L. described it as rather lacking precision. We believe that this comment represents the challenges she experienced when trying to move the robot using the virtual basic controls. However, she also emphasized the improvements on depth perception by using the augmented visual cues.

4.1. Hands-Free Multimodal Interaction

4.1.1. Similarities

People without motor disabilities and Miss L. did not experience problems using their head orientation (yaw and pitch) to make the robot move in the workspace. Additionally, the choice of words used as voice commands was easily remembered by both groups.

4.1.2. Differences

The voice commands presented challenges for Miss L. This was expressed in a comment mentioning that she feels that she has to speak louder but she was lacking energy. This resulted in constraints to move the robotic arm around the workspace. The volume of someone's voice had not been previously identified as an issue. This aspect invites reflection on the wide range of variability in the physical abilities of PMD [11]. Besides, an individual might be comfortable to move their head further left or right on one day but less on the next one.

4.2. Augmented Visual Cues and the Use of AR

4.2.1. Similarities

Both groups commented about the helpfulness of the visual cues to realize the position of the robot relative to a target. Moreover, both groups expressed positive and negative subjective impressions about the augmented visual cues. Overall, both groups expressed that using augmented visual cues improved their perception of depth.

4.2.2. Differences

For our participant with motor disabilities, the lack of visual acuity in one eye did not allow us to identify the virtual cursor clearly. This element was crucial to use the dwell buttons and execute fine adjustments to position the gripper. We believe that increasing the size of the cursor might mitigate this issue, although further testing would be needed for this affirmation to be accurate. Derived from this problem, the virtual menu—containing the basic robot controls—was unintentionally activated several times, possibly pointing to the downsides of having the menu within the operator's line of sight. We considered that having the basic controls in the line of sight at all times would reduce the workload evoked by gaze shifts. However, we realized that it also led to unintentional activations derived from the unawareness of the cursor's location. We noticed that this led to higher levels of frustration expressed by the IMD, which could evoke a feeling of lack of agency.

5. Participants with and without Motor Disabilities: Learning Points

We present learning points when comparing the experiences of people with and without disabilities with our robot teleoperation concept. We grouped them in three different topics that may provide a bigger picture of pragmatic factors to consider when handling experiments with these two different groups of participants.

5.1. Re-evaluate the Experimental Metrics

Handling experiments with people with and without disabilities brings a pertinent question about the type of metrics to be collected when evaluating any given concept. However, using the exact same objective metrics might not provide a real picture of the experience of people with disabilities. For instance, in our study, the lack of acuity in one eye did not always allow our participant to identify the virtual cursor and activate the basic robot controls. This led us toward not considering metrics such as time or distance in the evaluation. Instead, we emphasize the subjective impressions when using our teleoperation concept. Hence, we highlight the relevance of collecting subjective measures and qualitative analysis when carrying out studies with people with disabilities. We consider that qualitative analysis and subjective measures point at specific design decisions that could challenge the usability of a solution for people with disabilities.

5.2. Beware of the Bias

PMD may often feel excluded from design decisions in technology and most of them are eager to participate in design sessions as well as user studies, but it could also lead to a rather positive bias. For instance, Miss L. provided several positive comments about our augmented hands-free teleoperation concept reflected in the questionnaires used and the interview. However, during our observations, she struggled to teleoperate the robot because of the factors we previously mentioned (voice's volume and not being able to see the cursor at all times). This leads us to consider that her vast and positive experiences with robotic arms together with being involved in design sessions might have played a determining role. Although we call for caution about this positive bias, it also shows the influence of including participants in design sessions as a way to reduce fear or negative bias toward robots and even contributes to the notion of "design for empowerment" of people with disabilities [27].

Miss L.'s experience leads us to consider that individual and situational characteristics have a major influence on the impressions of interaction concepts. This aligns with previous findings of such factors in modality choices and may be even applied to general interaction concepts [28]. That is why we suggest performing periodical tests to familiarize participants not only with the technology but interaction concepts. We consider that this would allow reducing the influence of external and internal factors, e.g., individual characteristics of participants.

5.3. Consider Variability in Abilities Evoking Different Experiences

We believe that our findings emphasize the importance of usability testing and performing studies with people with disabilities, as it identifies problems that only PMD could possibly experience. Our results regarding hands-free multimodal teleoperation lead us to design to allow for multiple options in teleoperation. This can reduce the impact of having differences in participants' abilities and could even mitigate fatigue factors. For instance, in the morning, voice commands and head movements could be used as input modalities for teleoperation, and by the end of the day, when the voice may be feeble, the GUI could allow using solely head movements with a context menu displayed at the cursor. Therefore, we consider that presenting a base interaction that allows for multiple adaptations could improve the subjective experience of teleoperation and possibly performance.

6. Limitations

Our main limitation is presenting a single-subject study. We initially planned to carry out the study with more PMD. However, the impact of SARS-CoV-2 hindered us from inviting more participants as most of them belong to the high-risk group from this disease. Nevertheless, we align with Busso et al. [29] who remark the relevance of single-case studies to indicate intersubject variability.

In retrospect, we could have rescheduled the appointment for the experiment to minimize the effects of the situational characteristics of Miss L. However, it would have added more burden to our participant, as she needs special assistance and transportation to attend to experiments outside of her home. Therefore, we decided to carry out the experiment and report the subjective measures collected together with some reflections on her experience.

7. Conclusions

All in all, our findings align with Holloway's philosophy of "one size fits one" [15], as they point toward the absence of a design or interaction concept that allows people with and without disabilities to achieve the same level of performance or experience. In fact, there might not even be a singular design for a group of people sharing a diagnosis.

In this paper, we present the impressions of an augmented multimodal teleoperation concept from the perspective of a person with motor disabilities, Miss L. These suggest the potential of including AR in HRI for PMD and the challenges that may arise from certain design decisions, e.g., presenting robot controls within the line of sight and the use of voice commands. Further, we put in context the experiences of people without motor disabilities with those of Miss L. This led us to provide three learning points when evaluating an interaction concept with people with and without disabilities: (1) re-evaluating the metrics used to measure performance, (2) being aware of the bias, and (3) considering abilities' variability which evokes different experiences. Finally, our findings do not intend to present a definite conclusion about interaction design decisions but aim to enable discussions about the perspectives of people with disabilities in HRI design.

Author Contributions: Conceptualization, S.A.A.; methodology, S.A.A.; formal analysis, S.A.A., M.B.; investigation, S.A.A.; experiment, S.A.A. and M.B.; writing—original draft preparation, S.A.A.; writing—review and editing, S.A.A. and J.G.; supervision, J.G.; project administration, S.A.A. and J.G.; funding acquisition, J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the German Federal Ministry of Education and Research (BMBF, FKZ: 13FH011IX6).

Institutional Review Board Statement: Ethical reviews by an "Ethics Committee/Board" are waived in Germany as they are not compulsory for fields outside medicine.

Informed Consent Statement: Our participant gave informed consent before participating in the study and approved the text concerning the background in this manuscript.

Acknowledgments: We would like to thank our participant for her openness and willingness to participate in our study. We further would like to thank Henrik Dreisewerd for his support with the translations.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AR	Augmented Reality
DIX	Disability Interaction
GUI	Graphical User Interface
HMD	Head-Mounted Display
HRI	Human–Robot Interaction
IMD	Individual with Motor Disabilities
PMD	People with Motor Disabilities
SUS	System Usability Scale

References

1. Fraunhofer Institute for Manufacturing Engineering and Automation. Care-O-bot 4. Available online: https://www.care-o-bot. de/en/care-o-bot-4.html (accessed on 4 June 2021).
2. Guerreiro, J.; Sato, D.; Asakawa, S.; Dong, H.; Kitani, K.M.; Asakawa, C. CaBot: Designing and Evaluating an Autonomous Navigation Robot for Blind People. In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility—ASSETS '19, Pittsburgh, PA, USA, 28–30 October 2019; Bigham, J.P., Azenkot, S., Kane, S.K., Eds.; ACM Press: New York, NY, USA, 2019; pp. 68–82. doi:10.1145/3308561.3353771.
3. Kyrarini, M.; Lygerakis, F.; Rajavenkatanarayanan, A.; Sevastopoulos, C.; Nambiappan, H.R.; Chaitanya, K.K.; Babu, A.R.; Mathew, J.; Makedon, F. A Survey of Robots in Healthcare. *Technologies* **2021**, *9*, 8. doi:10.3390/technologies9010008.
4. Miseikis, J.; Caroni, P.; Duchamp, P.; Gasser, A.; Marko, R.; Miseikiene, N.; Zwilling, F.; de Castelbajac, C.; Eicher, L.; Fruh, M.; et al. Lio-A Personal Robot Assistant for Human-Robot Interaction and Care Applications. *IEEE Robot. Autom. Lett.* **2020**, *5*, 5339–5346. doi:10.1109/LRA.2020.3007462.

5. Chadalavada, R.T.; Andreasson, H.; Krug, R.; Lilienthal, A.J. That's on my mind! robot to human intention communi- cation through on-board projection on shared floor space. In Proceedings of the 2015 European Conference on Mobile Robots, Lincoln, UK, 2–4 September 2015; Robots, E.C.O.M., Duckett, T., Eds.; IEEE: Piscataway, NJ, USA, 2015; pp. 1–6. doi:10.1109/ECMR.2015.7403771.
6. Qvarfordt, P. Gaze-informed multimodal interaction. In *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations – Volume 1*; Oviatt, S., Schuller, B., Cohen, P.R., Sonntag, D., Potamianos, G., Krüger, A., Eds.; ACM:New York, NY, USA, 2017; pp. 365–402. doi:10.1145/3015783.3015794.
7. Park, K.B.; Choi, S.H.; Lee, J.Y.; Ghasemi, Y.; Mohammed, M.; Jeong, H. Hands-Free Human–Robot Interaction Using Multimodal Gestures and Deep Learning in Wearable Mixed Reality. *IEEE Access* **2021**, *9*, 55448–55464. doi:10.1109/ ACCESS.2021.3071364.
8. Makhataeva, Z.; Varol, H. Augmented Reality for Robotics: A Review. *Robotics* **2020**, *9*, 21. doi:10.3390/robotics9020021.
9. Culén, A.L.; van der Velden, M. The Digital Life of Vulnerable Users: Designing with Children, Patients, and Elderly. In *Nordic Contributions in IS Research*; van der Aalst, W., Mylopoulos, J., Rosemann, M., Shaw, M.J., Szyperski, C., Aanestad, M., Bratteteig, T., Eds.; Lecture Notes in Business Information Processing; Springer: Berlin/Heidelberg, Germany, 2013; Volume 156, pp. 53–71. doi:10.1007/978-3-642-39832-2_4.
10. Shinohara, K.; Bennett, C.L.; Wobbrock, J.O. How Designing for People With and Without Disabilities Shapes Student Design Thinking. In Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility —ASSETS '16, Reno, NN, USA, 24–26 October 2016; Feng, J.H., Huenerfauth, M., Eds.; ACM Press: New York, NY, USA, 2016; pp. 229–237. doi:10.1145/2982142.2982158.
11. Arevalo Arboleda, S.; Pascher, M.; Baumeister, A.; Klein, B.; Gerken, J. Reflecting upon Participatory Design in Human-Robot Collaboration for People with Motor Disabilities: Challenges and Lessons Learned from Three Multiyear Projects. In Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference, Corfu, Greece, 29 June–2 July 2021; ACM: New York, NY, USA, 2021; pp. 147–155. doi:10.1145/3453892.3458044.
12. Arboleda, S.A.; Pascher, M.; Lakhnati, Y.; Gerken, J. Understanding Human-Robot Collaboration for People with Mobility Impairments at the Workplace, a Thematic Analysis. In Proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy, 31 August–4 September 2020; pp. 561–566. doi:10.1109/RO- MAN47096.2020.9223489.
13. Arevalo Arboleda, S.; Rücker, F.; Dierks, T.; Gerken, J. 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, Yokohama, Japan, 8–13 May 2021; Kitamura, Y., Quigley, A., Isbister, K., Igarashi, T., Bjørn, P., Drucker, S., Eds.; ACM: New York, NY, USA, 2021; pp. 1–14. doi:10.1145/3411764.3445398.
14. Microsoft. Microsoft Hololens. Available online: https://www.microsoft.com/en-us/hololens/hardware (accessed on 31 July 2018).
15. Holloway, C. Disability interaction (DIX). *Interactions* **2019**, *26*, 44–49. doi:10.1145/3310322.
16. Bolt, R.A. "Put-that-there": Voice and gesture at the graphics interface. *ACM SIGGRAPH Comput. Graph.* **1980**, *14*, 262–270. doi:10.1145/965105.807503.
17. Jacob, R.J.K. *What You Look at Is What You Get: Eye Movement-Based Interaction Techniques*; ACM Press: New York, NY, USA, 1990; pp. 11–18. doi:10.1145/97243.97246.
18. Kinova. Professional Robots for Industry. Available online: https://www.kinovarobotics.com/en/products/ gen2-robot (accessed on 23 April 2020).
19. KUKA. LBR iiwa | KUKA AG. New York, NY, USA. Available online: https://www.kuka.com/en-de/products/robot-systems/industrial-robots/lbr-iiwa (accessed on 21 April 2020).
20. Universal Robots. UR5—The Flexible and Collaborative Robotic Arm, New York, NY, USA. Available online: https://www. universal-robots.com/products/ur5-robot/ (accessed on 30 June 2020).
21. Graser, A.; Heyer, T.; Fotoohi, L.; Lange, U.; Kampe, H.; Enjarini, B.; Heyer, S.; Fragkopoulos, C.; Ristic-Durrant, D. A Supportive FRIEND at Work: Robotic Workplace Assistance for the Disabled. *IEEE Robot. Autom. Mag.* **2013**, *20*, 148–159. doi:10.1109/MRA.2013.2275695.
22. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*; Hancock, P.A., Meshkati, N., Eds.; Advances in Psychology; North-Holland: Amsterdam, The Netherlands; Oxford, UK, 1988; Volume 52, pp. 139–183. doi:10.1016/S0166-4115(08)62386-9.
23. Brooke, J. Sus: A quick and dirty usability scale. In *Usability Evaluation in Industry*; CRC Press: Boca Raton, FL, USA, 1996; Volume 189.
24. Bartneck, C.; Kulić, D.; Croft, E.; Zoghbi, S. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *Int. J. Soc. Robot.* **2009**, *1*, 71–81. doi:10.1007/s12369-008-0001-3.
25. Bangor, A.; Kortum, P.; Miller, J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Stud.* **2009**, *4*, 114–123.
26. Millo, F.; Gesualdo, M.; Fraboni, F.; Giusino, D. Human Likeness in Robots: Differences between Industrial and Non-Industrial Robots. In Proceedings of the ECCE 2021: European Conference on Cognitive Ergonomics 2021, Siena, Italy, 26–29 April 2021 doi:10.1145/3452853.3452886.
27. Ladner, R.E. Design for user empowerment. *Interactions* **2015**, *22*, 24–29. doi:10.1145/2723869.

28. Arevalo, S.; Miller, S.; Janka, M.; Gerken, J. What's behind a choice? Understanding Modality Choices under Changing Environ- mental Conditions. In *Proceedings of the ICMI'19*; Gao, W., Ling Meng, H.M., Turk, M., Fussell, S.R., Schuller, B., Song, Y., Yu, K., Meng, H., Eds.; The Association for Computing Machinery: New York, NY, USA, 2019; pp. 291–301. doi:10.1145/3340555.3353717.
29. Busso, C.; Narayanan, S.S. Interrelation Between Speech and Facial Gestures in Emotional Utterances: A Single Subject Study. *IEEE Trans. Audio Speech Lang. Process.* **2007**, *15*, 2331–2347. doi:10.1109/TASL.2007.905145.

DOI: 10.17185/duepublico/76485
URN: urn:nbn:de:hbz:465-20220908-085412-0