

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Advancing Support for XR-Mediated Remote Collaboration on Physical Tasks

Permalink

<https://escholarship.org/uc/item/3q67b3s1>

Author

Johnson, Janet George

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Advancing Support for XR-Mediated Remote Collaboration on Physical Tasks

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Computer Science and Engineering

by

Janet George Johnson

Committee in charge:

Professor Nadir Weibel, Chair
Professor Mark Billinghurst
Professor William Griswold
Professor David Kirsh
Professor Scott Klemmer
Professor Kristen Vaccaro

2023

Copyright

Janet George Johnson, 2023

All rights reserved.

The Dissertation of Janet George Johnson is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

To my parents, George and Valsa. And to my brother, Jiju.

I owe all that I am to you.

EPIGRAPH

“The goal then becomes identifying needs which are not ideally met in the medium of physical proximity, and evolving mechanisms which leverage the strengths of the new medium to meet those needs.”

– Hollan & Stornetta, 1992

TABLE OF CONTENTS

Dissertation Approval Page	iii
Dedication	iv
Epigraph	v
Table of Contents	vi
List of Figures	x
List of Tables	xiii
Acknowledgements	xiv
Vita	xix
Abstract of the Dissertation	xxi
Chapter 1 Supporting Remote Collaborative Physical Tasks	1
1.1 The Promise of Extended Reality	2
1.2 The Challenge with Physical Tasks	5
1.3 Towards Enhancing Support for Collaborative XR	7
1.4 Thesis Statement and Contributions	10
Chapter 2 A Conceptual Framework for Designing Collaborative Systems	15
2.1 Collaboration through Mediated Communication	15
2.2 The Mechanics of Communication for Physical Tasks	17
2.3 Technology-Mediated Collaborative Physical Tasks	22
2.3.1 The Impact of Distance on Communication Bandwidth	23
2.3.2 Collaborators' Communicative Needs	25
2.3.3 The Impact of the Medium on Support for Communication	26
2.3.4 The Impact on Collaborators' Situated Experience	29
2.4 The Capabilities of XR as a Collaborative Medium	30
2.5 A Native Model for XR-Mediated Remote Collaboration	32
2.5.1 A Hybrid Virtual-Physical Space for Local Users	33
2.5.2 A Virtual Task Environment for Remote Users	34
2.5.3 Asymmetric Experiences of the Mutual Collaborative Space	36
2.6 Design Goals for Remote Guidance	37
2.7 Conclusion	38
Chapter 3 Evaluating Mixed Reality Guidance for Novices in Time-Critical Settings .	39
3.1 Introduction and Motivation	39
3.2 Background and Related Work	42
3.2.1 Evaluation of MR Guidance in Realistic Settings	42

3.2.2	Supporting Bystander-Initiated CPR	43
3.3	Designing Mixed Reality Guidance Cues	45
3.3.1	The Iterative Design Process	46
3.3.2	Design Challenges and Proposed Solutions.....	48
3.3.3	The Resulting Prototype: HoloCPR	52
3.4	Evaluating MR Guidance against Traditional Media	56
3.4.1	Evaluation Procedure	57
3.4.2	Results	60
3.4.3	Discussion	64
3.5	Conclusion	65
Chapter 4	Evaluating An Immersive Interface for Experts in Time-Critical Settings ..	66
4.1	Introduction and Motivation	66
4.1.1	Supporting Telementoring for Trauma Surgery	68
4.2	Designing an Immersive Expert Interface	69
4.2.1	Participatory Design and Prototyping Sessions	69
4.2.2	Expert Needs, Envisioned Features, and Workarounds	70
4.3	Evaluating Collaborative XR for Casualty Surgery	77
4.3.1	Feasibility Study with Perfused Cadaver Models	77
4.3.2	Qualitative Observations of Interface Use	78
4.4	Implications for Designing Remote Expert Interfaces	85
4.5	Conclusion	89
Chapter 5	A Review of XR-Mediated Remote Collaboration for Physical Tasks	91
5.1	Introduction and Motivation	91
5.1.1	Review Methodology	92
5.2	The Design Space of Collaborative XR Systems	95
5.2.1	Facilitating Environment Awareness	98
5.2.2	Facilitating Collaborator Awareness.....	101
5.2.3	Facilitating Communicative Actions	103
5.3	The Situated Experience with Current Virtual Task Environments	106
5.4	Support for Object-Focused Communication Today	111
5.5	Conclusion	117
Chapter 6	Enhancing Support for Referencing During Physical Tasks	119
6.1	Introduction and Motivation	119
6.2	Research Approach	122
6.3	Prototype Design and Implementation	124
6.3.1	Worker's First-Person View.....	124
6.3.2	A Map of Task Objects	124
6.3.3	A List of Task Objects	126
6.3.4	MR System Guidance through the HoloLens	126
6.4	Experiment	127
6.4.1	Participants and Procedure	128

6.4.2	Experiment Task and Environment	129
6.4.3	Experiment Measures and Analysis	131
6.4.4	Limitations of our study	131
6.5	Results	132
6.5.1	Task Performance	133
6.5.2	Efficiency of Communication	133
6.5.3	Helper's Referencing Behaviors	134
6.5.4	Collaborator Task Load	137
6.5.5	Participant Perceptions	138
6.6	Understanding Affordances	140
6.6.1	Effects of 2D Spatial Information (R1)	140
6.6.2	Effects of MR System Guidance (R2)	141
6.6.3	Combining Spatial Information and MR Guidance (R3)	143
6.7	Implications for Designers and Developers	144
6.7.1	Offloading Referencing to MR to Increase Efficacy	144
6.7.2	Making Support for Referencing More Accessible	146
6.8	Conclusion	146
Chapter 7	Leveraging Experts' Situated Experiences for Remote Collaboration	148
7.1	Introduction and Motivation	148
7.2	Our Design Strategy	151
7.3	Our Research Approach	153
7.4	Prototype Design and Implementation	155
7.4.1	System Overview and User Features	156
7.4.2	Study-Specific Features	158
7.4.3	Implementation Details	159
7.5	Experiment	160
7.5.1	Experiment Task	160
7.5.2	Recruitment	162
7.5.3	Experiment Setup	163
7.5.4	Study Procedure	164
7.5.5	Experiment Measures and Data Analysis	165
7.6	Results	166
7.6.1	Task Completion and Search Time	167
7.6.2	Collaborator Task Load	168
7.6.3	VR Interface Interactions	168
7.6.4	Communication Efficiency and Patterns	169
7.6.5	Self-Reported Questionnaire Scores	172
7.7	Observational Results and Participant Feedback	173
7.7.1	Behavioral and Experience Differences Between Mapped and UnMapped	174
7.7.2	Learning from Repeated Use	178
7.8	Discussion	178
7.8.1	The Impact of a Familiar Environment on Remote Guidance	179
7.8.2	The Implications of a Spatially Stable Environment on Repeated Use	180

7.8.3	Designing for Practical Use and Limitations	181
7.9	Conclusion	183
Chapter 8	Conclusion	186
8.1	A Summary and Discussion	186
8.2	Looking forward	189
8.3	Closing Remarks	194
Appendix A	Papers Reviewed	196
Appendix B	Helper Task Instructions	197
Appendix C	Cake Decorating Task	204
Bibliography		212

LIST OF FIGURES

Figure 1.1.	Milgram and Kishino’s reality-virtuality continuum.....	2
Figure 1.2.	A vision of how XR-mediated collaboration can differ from screenbound experiences.	4
Figure 2.1.	A conceptual model depicting collocated collaboration for physical tasks..	17
Figure 2.2.	A conceptual model of the mechanics of technology-mediated collaboration during physical tasks.	22
Figure 2.3.	Each collaborative system feature impacts ease of communication and the situated experience.	23
Figure 2.4.	The cognitive and physical cycles for individual users during technology-mediated collaboration.	25
Figure 2.5.	Dimensions of support for communication with technology-mediated collaboration.....	27
Figure 2.6.	A conceptual model for XR-mediated collaboration.....	33
Figure 2.7.	Dimensions of the situated experience provided by the virtual task environment or the interface for remote experts.....	35
Figure 3.2.	The envisioned MR prototype for expert-guided CPR.	46
Figure 3.3.	Prototyping with HoloLens and our HoloCPR mannequin.	48
Figure 3.4.	Instructions for cardiopulmonary resuscitations, as seen by the wearer. ...	49
Figure 3.5.	The “Red Arrow” guides users by indicating where the important things are located and where they should act.	50
Figure 3.6.	The “Circle-of-attention” points the user to important information even with a narrow field-of-view.	51
Figure 3.7.	Dividing instructions into “What”, “How” and “Where”	54
Figure 3.8.	An artistic depiction of the HoloCPR scenario.	55
Figure 3.9.	CPR instructions and the experiment set up.....	59
Figure 3.10.	Response time, transition time, time spent searching for the AED, and time taken to attach AED pads across both conditions.	61

Figure 4.1.	In this chapter we focus on the choice of immersive technology to provide a virtual task environment for remote experts.	67
Figure 4.2.	Model operating room (left) and expert space (right) set up in the lab for iterative design sessions.	70
Figure 4.3.	Design sessions with domain experts, technology researchers, and design researchers.	70
Figure 4.4.	Features that provide remote experts with a view of of the task environment and collaborators.	71
Figure 4.5.	Features that help remote experts communicate technique.	74
Figure 4.6.	Sketch of the features that could help novices identify tools.	75
Figure 4.7.	An artistic rendition of the features integrated into final design for both the remote expert and the local novice.	77
Figure 5.1.	The PRISMA flow diagram for article selection and screening.	93
Figure 5.2.	Distribution of papers presenting systems designed for MR-mediated guidance for physical tasks.	96
Figure 5.3.	Distribution of features based on the primary collaborator need they satisfy.	96
Figure 5.4.	Features that support environment awareness for remote physical tasks. ..	98
Figure 5.5.	Features that support collaborator awareness for remote physical tasks. ...	101
Figure 5.6.	Features that support communicative actions for remote physical tasks. ...	104
Figure 5.7.	Technology used to provide remote users with a virtual task environment..	107
Figure 5.8.	Support for referencing with current collaborative XR systems.	110
Figure 5.9.	Support for referencing with current collaborative XR systems.	117
Figure 6.1.	2x2 mixed factorial experiment design that studies the effect of MR guidance for workers and two different helper aids.	123
Figure 6.2.	Helper's space and worker's environment in our experiment.	125
Figure 6.3.	First-person view of the worker as they are guided to an object by the MR guidance	126

Figure 6.4.	Helper's aids. (a) <i>Map</i> showing the task objects and the worker's position. (b) <i>List</i> showing all the task objects.	127
Figure 6.5.	Snapshots of the worker's environment and task objects.	128
Figure 6.6.	Average completion time across participants and conditions. (a) Total completion time across both subtasks. (b) Completion time for Subtask1. (c) Completion time for Subtask2.	132
Figure 6.7.	Top: Mean number of words per task by helper; Bottom: Mean number of words per task by worker.	135
Figure 6.8.	Results of the NASA TLX questionnaire. Top: Task load on helpers; Bottom: Task load on workers.....	137
Figure 6.9.	Perceived importance of the map and list aids. Top: Perceived importance on general collaboration; Bottom: Perceived importance on remote guidance.	139
Figure 7.1.	An artistic rendering of the UnMapped prototype that leverages remote experts' situated experiences by providing an immersive virtual environment that reflects their own workspace.	150
Figure 7.2.	A view of the Mapped and UnMapped VR interface for remote experts. ..	156
Figure 7.3.	MR features for the novice in the local task environment.	158
Figure 7.4.	The novice space in our experiment including the active task space and the novice using MR-mediated guidance to decorate a cake.	161
Figure 7.5.	Left: Experiment space for the remote expert in our study that also included space for the Wizard; Right: One of the two novice environments in our study.	163
Figure 7.6.	Average task completion (left) and search time (right) across study conditions with significant differences between Interface Types depicted for each Session.	167
Figure 7.7.	Average task load for experts (left) and novices (right) across study conditions with significant differences between Interface Types depicted for each Session.	168
Figure 7.8.	Top: Expert's overall utterance patterns with the Mapped and UnMapped interface; Bottom: Expert's overall reference patterns with the Mapped and UnMapped interface	170

LIST OF TABLES

Table 2.1.	Classification of the constituents of awareness necessary during collaborative physical tasks.	20
Table 2.2.	Key communication processes collaborative physical tasks rely on.	21
Table 2.3.	XR capabilities grouped according to what it offers in terms of the user experience.	31
Table 5.1.	A visualization of the how the design space for collaborative XR is currently populated.	97
Table 5.2.	The situated experience current collaborative XR systems provide remote users.	108
Table 6.1.	Observed referencing behavior across the presence or absence of MR guidance and the different aids.	136
Table 7.1.	Expert interactions within the VR interface across study conditions. Significant differences are highlighted in green.	169
Table 7.2.	Collaborators' social presence and perceived guidance quality scores across study conditions.	172
Table 7.3.	Experts' spatial presence scores and cognitive representation of the novice's environment across study conditions.	172
Table 8.1.	A visualization of the how different XR capabilities have been leveraged to support the various communication needs.	191

ACKNOWLEDGEMENTS

When I first moved to California to pursue a Master's degree, I didn't know where life would take me, but I was certain a PhD wasn't on the cards. And yet, here we are. I could not have, in my wildest dreams, imagined the journey this would be. My accomplishments within this dissertation and beyond are a result of the collective influence and unwavering support of so many wonderful people I've had the blessing of crossing paths with.

To Nadir, thank you for seeing potential in me before I ever could. Thank you for teaching me the ropes of research and academia and for always encouraging me to listen to my gut. Thank you for being more than an academic advisor and holding space for all parts of my life – in you I have found a mentor and a friend, and for that, I cannot thank you enough.

To the rest of my committee – Scott, Bill, Mark, David, and Kristen – it has been an absolute pleasure to learn from you. Thank you Scott, for teaching me how to frame research and for instilling in me a love for teaching. Thank you Bill, for gently asking me questions that nudge me towards the right direction and for all the words of encouragement. Thank you Mark, for being a true inspiration and for patiently helping me think more deeply about the field. Thank you David, for the riveting research discussions and for the best eclairs I've had till date. Thank you Kristen, for the candid discussions about research and the future.

To everyone in the HXI lab, you've made my time here so memorable and have (mostly) succeeded in keeping me sane. Thank you in particular – To Steven, for being like a second advisor to me and for being someone I can always turn to. To Vish, for the questionable long drives and the even more questionable conversations. To Danilo, for encouraging me to pursue a PhD and for all the ragestorming. To Tommy, for being a creative presence and for being the most reliable friend and colleague one could ask for. To Matin, for the endless fits of laughter and for reminding me to question the things I take for granted. To Manas, for making me feel like a wise owl and for giving me company when work felt impossible. And to Chen, for being ever ready to lend a helping hand.

To the students I have had the privilege of mentoring, I have learnt more from you

than I could ever hope to teach you. And needless to say, none of the work presented in this dissertation would have been possible without you. Thank you in particular – To Evan, for being an absolutely brilliant analysis partner, and for being so patient with me as I fumbled through the research and learnt what it meant to mentor someone. To Cynthia, for bringing so much skill and the most calming presence to chaotic times. To Danica, for being so enthusiastic about the work and for doing whatever needed to be done for us to move forward. To Ruijie, for being the quickest learner and for asking the hard questions. To Ren, for listening to me ramble and helping me ensure our research was sound.

To all of my collaborators and colleagues both within academia and outside, thank you for your indispensable support. It has been a privilege to work alongside you. Thank you in particular – To Venkatesh, for leading the way in a challenging environment and for making it look easy. To Preetham, for picking up the phone when I most needed it and for your endless passion towards research. To Jürgen, for having faith in my XR design abilities even when I was new to it. To Wanze, for making those designs a reality. To Andrea, Paul, and everyone on the PwC team, for the excitement that led to bold designs. To Ming, Emilia, Allison, and everyone on the Open&Ask and VOLI projects, thank you for trusting me to take your research forward. To Konrad, Roland, and everyone at the Naval Medical Center, thank you for the opportunity to work on a project with such tangible impact and for all our conversations that undoubtedly shaped this dissertation. To Yuanyuan, Robin, Ru, Enrique, and everyone else on the ARTEMIS team, thank you for the experiences that showcased the power of teamwork. To everyone at Reality Labs and XRSI, thank you for being so welcoming and for the unique insights that shape how I think about the field. To Paolo, you have been nothing short of an inspiration. Thank you for believing in me right from the start and for always being open to a conversation. To Stu and Jacquie, thank you for showing me how to be a compassionate researcher and reminding me to stay focused on bringing value to my work. To Jo, Fiona, and Sophiya, thank you for your friendship and for being incredible role models. To Kavya, thank you for fearlessly tackling the challenging questions and for the empathetic but resolute leadership you exemplify.

To everyone in the Design Lab and the CSE department, thank you for creating a vibrant community that welcomes discourse. Thank you in particular – To Nadyne, for being someone I felt comfortable approaching when the going got tough. To Wendy, for invaluable wisdom and for all the quick hallway greetings that turned into long conversations. To Julie, thank you for being a constant source of support. To Teenah, Nivardo, and everyone else on the dlab operations team, thank you for your invaluable role in bringing our research dreams to life. To Lilly, Jim, Don, Steven, and all of the dlab faculty, thank you for your warm smiles and open doors. To the students I’ve been fortunate to forge friendships with, you make the long grind that is a PhD so much richer – Vineet, Tone, Shrishti, Dylan, Colin, Varun, and Jessica, thank you for the camaraderie and the joyous support.

To the inspirational teachers from my earlier years, thank you for being a guiding light. Thank you in particular – To Srikanth Prabhu and Harish S V for pushing me to do more than I thought I was capable of. To Ninan sir, Lali ma’am, Bindu ma’am (BJ), Everest ma’am (EV), Mini ma’am (MM), and all of Little Rock’s teachers, for creating an environment where excellence flourished and for teaching us of our humanity. To Sona ma’am (SO), for providing me with the strong foundations I have built my career on and for your constant friendship.

To my dearest friends, thank you for the many adventures and for instilling confidence in me even when you weren’t sure of what I was doing. Thank you in particular – To Diksha, Kiran, Preethish, Bhuvana, Arpan, Raveena, and Mischa, for being my earliest cheerleaders and for unwavering support I can only hope to reciprocate. To Ariana, Marisol, Rob, Tricia, Ailie, and Chris, for genuine companionship and for embracing the weirdness. To Shreyas, Shrey, Nanzo, Kriti, Ayush, Apurva, Pallavi, Kanishka, Abeer, Dhruv, Anupriya, and Shivika, for celebrating every win and for listening to me complain about every pitfall. To Shashank, Arpan, Markus, Tobias, Rushil, Tushar, and Ravi, for somehow making me feel so deeply understood while poking fun at me and for the often needed brutal honesty. To Dhiman, Gökçe, Shravan, and Nishant, for patiently serving as my sounding board (even when it drove you crazy) and for holding my hand when life got rough. To the many others I haven’t named here, thank you.

To my family, thank you for your endless faith in me. Dad, thank you for teaching me to dream big, for instilling in me a drive for excellence, and for reminding me that it didn't matter what I did as long as I did it well. I hope I've made you proud. Mom, your resilience and nurturing guidance have paved the path I'm on. Thank you for showing me what empathy meant and for teaching me to look for the unique ways I can add value to those around me. Bhaiyah, I've tried to emulate you for as long as I can remember – so it's a good thing you're as inspirational as you turned out to be. Thank you for being someone I can be authentically myself with and for being a reliable mentor, protective sibling, and cool friend all rolled into one. Olivia, thank you for always looking out for me and for teaching me to stand up for myself. Lucas and Atlas, watching you grow has brought me joy that words cannot capture. Thank you for reminding me of the world's innocence and potential. Subra, thank you for the countless sacrifices you have made to make my education and life as I know it possible. To Sequiera aunty, Maria aunty, Sicy aunty, Babu uncle, Elsy aunty, Matilda aunty, Alex uncle, Kuttu, Kevin, Midhun, Shibby, Joel, Georgie, Anu, Abish, Gavin, and the rest of my extended family who have always had a place for me in your homes, thank you.

Chapter 3, in part, includes portions of material as it appears in *HoloCPR: Designing and Evaluating a Mixed Reality Interface for Time-Critical Emergencies* by Janet G. Johnson, Danilo Gasques Rodrigues, Madhuri Gubbala, and Nadir Weibel in the Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare, ACM, 2018. The dissertation author was the primary investigator and author of this material.

Chapter 4, in part, includes material that appears in *ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring* by Danilo Gasques, Janet G Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Yifei Zhang, Wanze Xie, Xinming Zhang, Konrad Davis, Michael Yip, and Nadir Weibel in the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. The dissertation author was one of the primary investigators of this material.

Chapter 4, in part, also includes material that appears in *I Want to Be a Surgeon! Role Playing for Remote Surgery in Mixed Reality* by Thomas Sharkey, Janet G. Johnson, Danilo Gasques, and Nadir Weibel in Workshop on Interactive Systems, 2019 CHI Conference on Human Factors in Computing Systems. The dissertation author was one of the primary investigators of this material.

Chapter 6, in part, includes portions of material as it appears in *Do You Really Need to Know Where “That” Is? Enhancing Support for Referencing in Collaborative Mixed Reality Environments* by Janet G. Johnson, Danilo Gasques, Tommy Sharkey, Evan Schmitz, and Nadir Weibel in the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. The dissertation author was the primary investigator and author of this material.

Chapter 7, in part, includes portions of material as it appears in *UnMapped: Leveraging Experts’ Situated Experiences to Ease Remote Guidance in Collaborative Mixed Reality* by Janet G. Johnson, Tommy Sharkey, Iramuali Cynthia Butarbutar, Danica Xiong, Ruijie Huang, Lauren Sy, and Nadir Weibel in the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. The dissertation author was the primary investigator and author of this material.

Chapter 2 and 5, in part, are currently being prepared for submission for publication of the material by Janet G. Johnson, Nadir Weibel, and Mark Billingham. The dissertation author is the primary investigator and author of this material.

VITA

- 2014 Bachelor of Engineering in Computer Science and Engineering, Manipal Institute of Technology, Manipal University, India
- 2020 M.S. in Computer Science, University of California San Diego
- 2023 Ph.D. in Computer Science, University of California San Diego

PUBLICATIONS

Janet G. Johnson, Tommy Sharkey, Iramuali Cynthia Butarbutar, Danica Xiong, Ruijie Huang, Lauren Sy, and Nadir Weibel. “*UnMapped: Leveraging Experts’ Situated Experiences to Ease Remote Guidance in Collaborative Mixed Reality*”, In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.

M.D. Tadlock, E.J. Olson, D. Gasques, R. Champagne, M.J. Krzyzaniak, S.A. Belverud, V. Ravindra, J. Kerns, P. M Choi, J. Deveraux, J. Johnson, T. Sharkey, M. Yip, N. Weibel, K. Davis. “*Mixed reality surgical mentoring of combat casualty care related procedures in a perfused cadaver model: Initial results of a randomized feasibility study*”, In Surgery 2022.

Janet G. Johnson, Danilo Gasques, Tommy Sharkey, Evan Schmitz, and Nadir Weibel. “*Do You Really Need to Know Where “That” Is? Enhancing Support for Referencing in Collaborative Mixed Reality Environments*”, In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.

Danilo Gasques, Janet G Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Yifei Zhang, Wanze Xie, Xinming Zhang, Konrad Davis, Michael Yip, and Nadir Weibel. “*ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring*”, In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.

Chen Chen, Janet G Johnson, Kemeberly Charles, Alice Lee, Ella T. Lifset, Michael Hogarth, Alison A. Moore, Emilia Farcas, and Nadir Weibel. “*Understanding Barriers and Design Opportunities to Improve Healthcare and QOL for Older Adults through Voice Assistants*”, In Proceedings of ASSETS 2021, International ACM SIGACCESS Conference on Computers and Accessibility.

W. Xie, Y. Liang, J. Johnson, A. Mower, S. Burns, C. Chelini, P. D’Alessandro, N. Weibel, and J.P. Schulze. “*Interactive Multi-User 3D Visual Analytics in Augmented Reality*”, In Proceedings of IS&T International Symposium on Electronic Imaging 2020.

Janet G. Johnson, Evan Schmitz, Venkatesh Ramnath, and Nadir Weibel. “*Nurture-Empower-Support: A Human-Centered Approach to Understand and Support ICU Families*”, In Proceedings of The 13th EAI International Conference on Pervasive Computing Technologies for Healthcare, pp. 119-128. ACM, 2019.

Janet G. Johnson, Danilo Gasques Rodrigues, Madhuri Gubbala, and Nadir Weibel. “*HoloCPR: Designing and Evaluating a Mixed Reality Interface for Time-Critical Emergencies*”, In Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare, pp. 67-76. ACM, 2018.

FIELDS OF STUDY

Major Field: Human Computer Interaction (Computer Science)

ABSTRACT OF THE DISSERTATION

Advancing Support for XR-Mediated Remote Collaboration on Physical Tasks

by

Janet George Johnson

Doctor of Philosophy in Computer Science and Engineering

University of California San Diego, 2023

Professor Nadir Weibel, Chair

Technology-mediated collaboration for physical tasks – where a novice manipulates real objects in their space with the guidance of a remote expert – helps extend expertise to remote and often resource-strapped environments. The spatial and immersive nature of emerging technologies like eXtended Reality (XR) promise to make this collaboration more natural and intuitive in ways that traditional screenbound collaborative systems can't. However, a plethora of challenges remain before it can fulfill this promise, and several key aspects of the design space are yet to be explored.

In this dissertation, I first introduce a framework for XR-mediated collaborative physical tasks that aims to serve as a cohesive resource for both the design of collaborator interfaces

and the exploration of the unique capabilities spatial and immersive technology bring to remote collaboration. I then design Mixed and Virtual Reality interfaces for both experts and novices and evaluate them in realistic settings. These evaluations expand our understanding of XR as a collaborative medium and highlight the challenges associated with supporting the situational awareness experts need to partake in object-focused communication. An in-depth review of existing collaborative XR systems adds to the discourse and uncovers key design opportunities to enhance support for communication during physical tasks. This dissertation also highlights a core assumption within the field and challenges the need to fully support experts' environment awareness through a view of the task environment. I do this by leveraging the context-aware capabilities of the technology to provide alternate information representations and communication mechanisms that minimize the information foraging collaborators need to perform within the task environment to sustain object-focused communication. Controlled experiments evaluate the impact of these techniques on the collaborative experience and showcase how they reduce task load while increasing task performance and communication efficiency.

Put together, this work lays the foundations for rich collaborative experiences that go beyond the constraints of face-to-face interactions while making it easier and more sustainable to scale real-world expertise.

Chapter 1

Supporting Remote Collaborative Physical Tasks

From baking and construction, to flying an aircraft and performing surgical procedures, purposeful *physical tasks* where we alter real world objects around us have driven human development and fundamentally shaped our societies. While some of these tasks have found digital equivalents as technology developed, those that are intricately tied to the environment they occur in remain crucial across a variety of domains like manufacturing, design, and healthcare. These physical tasks are typically complex in nature and require a high degree of skill and tacit knowledge that is gained through exclusive training and honed over time with repeated practice. Today, this specialized knowledge is becoming increasingly distributed with skilled personnel often being concentrated in urban centers [208] – which means the expertise required to perform tasks like surgery [253] or aircraft repair [191] is not always locally available.

Advances in telecommunications and collaborative technology over the past few decades have made working with each other over distances an easier, richer, and more ubiquitous experience. This technology-mediated remote collaboration has allowed for more distributed teams and shifted how we structure and leverage the expertise required for tasks. With technological systems that support remote collaboration, *remote experts* who are geographically removed from the task environment can provide procedural guidance to *local novices* who manipulate the objects in the task environment [93]. The expertise, tacit knowledge, and decision making skills

offered by the expert can make task completion easier, more efficient, and less error prone for the novices, especially in time-critical settings. For example, technology has allowed remote experts to assist with crime scene investigations [242] and enabled trauma surgeons to remotely guide emergency patient care at a moments notice [183].

Most collaborative systems, however, are built to support activities that can be performed without reference to a spatial or physical environment [93] (like video conferencing or editing a document together) and provide insufficient support for physical tasks that rely on it [94]. Collaborative systems explicitly designed for physical tasks is therefore key to expanding access to the expertise required for it in both geographically distributed teams as well as in resource-strapped scenarios.

1.1 The Promise of Extended Reality

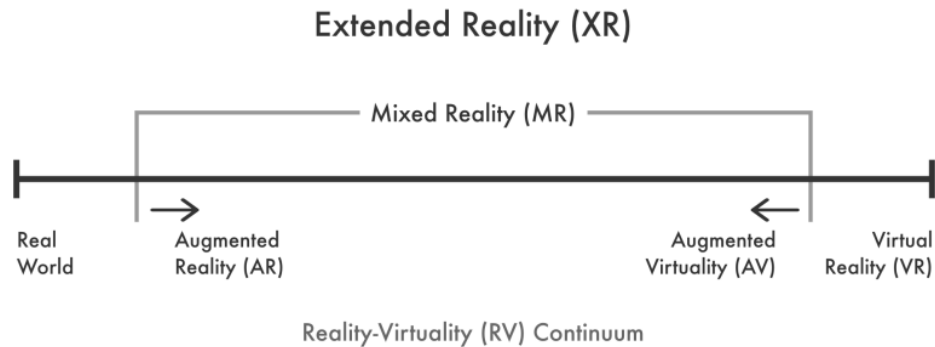


Figure 1.1. Milgram and Kishino’s reality-virtuality continuum.

The virtuality continuum (Fig. 1.1), initially introduced by Milgram and Kishino [211] describes one end of the continuum as “reality”, which is the perception of the physical environment around us without the use of any technology. The other end of the spectrum is Virtual Reality (VR) where users are immersed in a completely virtual, computer-generated environment. Anything in between is referred to as Mixed Reality (MR) which includes both augmenting the physical environment with digital content (Augmented Reality), and integrating real objects into a virtual environment (Augmented Virtuality). Extended Reality (XR) is an umbrella term that

covers any technology-mediated experience within this continuum.

Today, the concept of MR and XR has expanded beyond visual displays and can mean a variety of different things [282]. Fundamentally, in this dissertation, I think of XR as a collection of display, input, and tracking technologies that enable spatial and immersive user experiences that can enhance human perception [282, 274]. I recognize that no single device today captures the full potential of what XR can offer, and that XR experiences often include additional capabilities like haptics, human-environment interaction, hand tracking, sensors, and motion tracking [290, 263].

A Natural Collaborative Medium for Physical Tasks

Given the inherent spatial nature of physical tasks, the immersive and spatial nature of XR makes it well-suited to be a collaborative medium for it because it can provide an environment where collaborators can feel like they are in the same 3-dimensional space [289] (Fig. 1.2 shows an artistic depiction of this vision). This can improve their shared context and restore their natural ability to understand and interact with digital information, collaborators, and both local and remote environments with relative ease. In general, these technologies allow for communication behaviors similar to face-to-face collaborations, especially when compared to traditional collaborative interfaces on screens [26].

Mixed Reality (MR), in particular, seamlessly blends our physical and digital worlds [212, 211] and is capable of understanding the real world and presenting virtual content within it. When compared to video conferencing, previous research has shown that MR facilitates a reduced separation between task and communication spaces, a higher degree of social presence, and more natural communication behaviors [26]. Recent years have thus seen an increased interest in commercial applications as well as an uptick of research efforts towards MR-mediated collaboration [323, 77] for physical tasks.

While experiencing the digital content alongside the real world is beneficial for collaborators in the physical task environment, the environment of the geographically separated user is

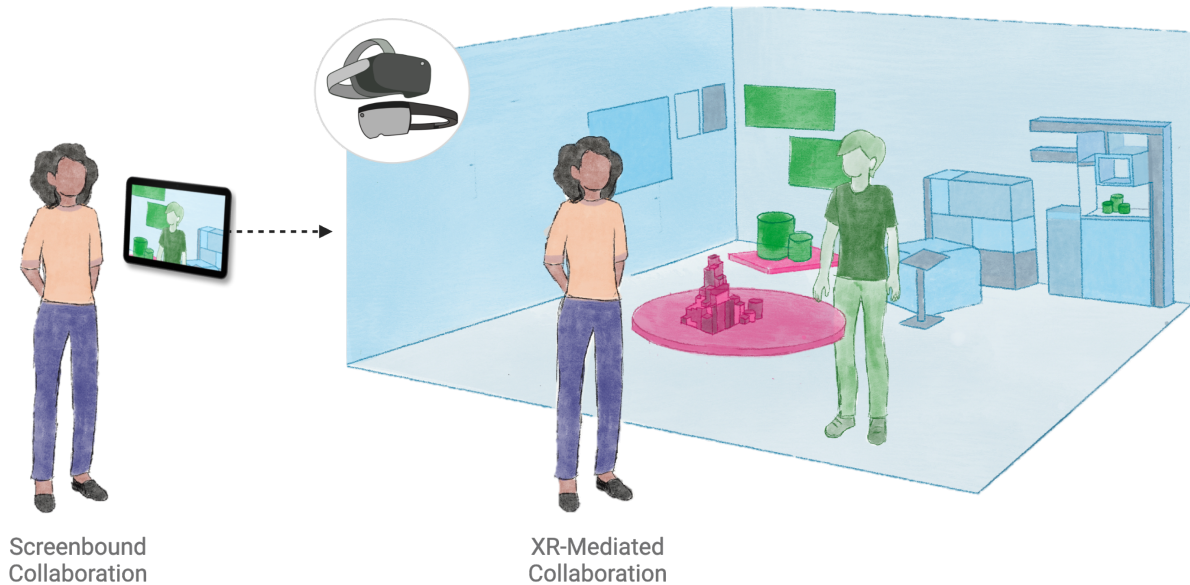


Figure 1.2. A vision of how XR-mediated collaboration can differ from screenbound experiences.

typically not relevant. The flexibility of technology choice this provides have led researchers to use a range of devices for remote collaborators including Virtual Reality (VR) which can present 3D visual content in a spatial manner which enables full depth perception and allows users to perceive environments in a similar manner as they would a physical space in the real world. The recent advances in the technology including better display technologies, environment understanding, more processing power, and higher network bandwidth have made it possible for features like real-time 3D reconstructions or digital twins to recreate the experience of being in a real physical space that is geographically separated from the user [227, 46, 190, 301]. With this, remote experts can engage with the task environment more naturally [278] and provide guidance to local novices in manners similar to collocated settings – something that can be especially beneficial for complex physical tasks [136].

Beyond Imitating Collocated Settings

While XR has the ability to recreate the experience of collocated settings, the constraints of the technology like restricted field-of-view, limited resolution and color depth, and varied ergonomic issues [26] means that viewing the world through these devices is simply not the same

as viewing it through the naked eye. It is important to note that the impact of these constraints will be minimized as technology advances. However, I argue that the seams and discontinuities in the collaborative experience [140] will remain even the technology-mediated collaborative space is indistinguishable from one in the real world. This is primarily because even if collaborators experience virtual content through stimulation of their extroceptive senses, their introceptive senses will likely remain unaltered [274]. For example, even if remote collaborators were fully immersed in a flawless reconstruction of the task environment that's impossible to tell apart from the real thing, they would still be aware of the fact that they aren't physically present in that space.

Moreover, while face-to-face interactions are often treated as the gold standard systems that support collaboration should strive towards, CSCW researchers have challenged this approach in the past. Most notably, Hollan and Stornetta called for telecommunication systems that go beyond "being there" [132], focusing on leveraging the unique capabilities of technology will help us create collaborative systems that are seen as more than just replacements for physical proximity and could even help us overcome the deficiencies of collocated communication.

The specific technology used as the collaborative medium heavily impacts the design space available to explore communication mechanisms that diverge from those used when collaborators are present in the same space. In fundamentally modifying the perceptive and sensory experience of a space, XR in particular offers us (probably to a larger extent than traditional media) the flexibility of malleable space-time perceptions [77] and new types of social interactions [113] that can go beyond what physical proximity offers.

1.2 The Challenge with Physical Tasks

While most synchronous collaboration technologies (like video conferencing) are designed to support tasks that can be performed without reference to the real world [94], physical tasks require it as they are fundamentally tied to the environment they occur in and the objects

within it. Providing communicative support for a collaborator removed from this environment this becomes a non-trivial task. When it comes to collaborative systems that allow remote experts to provide MR-mediated guidance for physical tasks, researchers have explored a number of rich interaction techniques that leverage the unique capabilities of the technology. For example, remote collaborators can use hand gestures [10, 237], virtual replicas [224] of task objects, or environment annotations [210, 45, 105] to engage in epistemic actions [169] and communicate technique in ways that are not entirely possible when working remotely without the visual and spatial capabilities of XR.

However, these communication mechanisms require substantial environment awareness to partake in. For example, to gesture towards a task object or sketch a circle around it to direct a novices attention towards it, a remote expert would need to know the position of that object in the task environment. In fact, the role shared visual information of the task environment plays in physical tasks is very well documented in CSCW literature, and is almost always seen as indispensable for remote collaborators to gain situational awareness and ground their object-focused communications more effectively [174]. Providing remote collaborators with a view of the task environment is therefore typically seen as a non-negotiable feature.

A consequence of this has been that despite the increased interest in commercial applications from the public and an uptick of research efforts towards XR-mediated collaboration over the recent years [323, 77], a key goal for remote collaborator interfaces have been to increase its users sense of presence in the task environment by immersing them in computationally recreated versions of it. In general, the physicality of the task, the importance of workplace awareness, and the ease with which we gain this awareness and communicate in real physical spaces have made it extremely challenging to design interfaces that truly go beyond recreating physical proximity. Today, we don't know what an interface that breaks past this boundary would entail and it is unclear what the benefits of it would be.

1.3 Towards Enhancing Support for Collaborative XR

The concept of adjacent possible, originally proposed by Kauffman [154], is an abstract notion that has been used as a framework to structure innovation in a variety of research fields including design [28]. When applied to the design of technological systems, the *actual* describes the current state of the field which includes the systems in use today as well as the knowledge of the underlying conceptual components and the relationships between them. The *adjacent possible* refers to the opportunities available for innovation that are outside but near the current state of knowledge, where known technologies and approaches are combined in generative ways to yield novel results. The ability to engage in design explorations that grow the boundaries of the adjacent possible requires a systematic understanding of the actual. A crucial step towards understanding the state of the field is to distinctively outline what it means to support physical tasks at a distance, and understand the impact of the technology capabilities available for it. Today, there are two key aspects missing that make this endeavour challenging:

The Lack of Theoretical or Conceptual Frameworks Native to Collaborative Extended Reality

In the past, models and theories that provide an understanding of human communication have played a crucial role in facilitating the progress of traditional collaborative systems in less hit-or-miss ways [1]. They offer a humanistic perspective of how people interact with technology [123, 20], assist in the framing of new research paths, act as a guide for innovation and design practice, and offer new ways to evaluate these systems [257, 128]. However, in going beyond screen-bound experiences, the 3-dimensional and immersive interaction capabilities of XR comes with its own unique characteristics and challenges. Additionally, in presenting virtual content that can be perceived and interacted with in manners similar to real objects, XR can fundamentally change the collaborator's situated experience in their environment. Therefore, while extremely valuable, insights from research on collocated and technology-mediated com-

munication might break down when applied to MR, and existing communication models alone provide an incomplete picture of the different aspects of MR-mediated collaboration.

MR-specific collaboration theories or frameworks for physical tasks are even more scarce with Marques et al.'s ontology for evaluation [204] and Chastine et al.'s model for referencing in collaborative AR [44] being the two major efforts in the space. Additionally, most of the work that offers a review of the space use dimensions that primarily focus on the underlying technology but don't provide a way to describe the qualities of the user experience [77]. As a result, researchers and practitioners are left with very few design guidelines and theoretical scaffolding to guide them [148].

While it is natural for emerging technologies to lack theoretical foundations, advances in MR are changing the nature of human interactions in ways we don't quite understand [30]. Creating impactful, usable systems that move the field towards maturity will require a more structured approach where we build on existing theories in generative ways [20] and develop more "native" models that specifically describe the XR-mediated collaborative experience. These native models, in turn, allow us to articulate guiding questions that drive the successful exploration of the adjacent possible.

The Lack of User-Centered Design and Evaluation of XR Interfaces for Remote Collaboration in Realistic Settings

General research on MR-mediated remote collaboration dates back to the 1990's with work like the SharedView [179]. Since then researchers have built collaborative MR systems with various features like presenting collaborators as avatars to each other, sharing virtual content to supplement communication, and allowing a remote user to place spatial cues in the local collaborator's workspace. When looked at through the lens of technology development models like the BRETAM trajectory [96], MR-mediated collaboration is very much at the replication stage where there is little grounding foundation for research explorations and the collective insights gained primarily stem from empirical reports of successes and failures. There is,

therefore a sense that we have just begun exploring the vast design space of collaborative MR systems, with several key research directions yet to be pursued. Understanding how to cull this design space and build effective collaborative systems is key to furthering this research field in the most impactful manner.

The majority of current research efforts focus on creating individual systems and exploring technological innovations and interaction techniques [204, 77]. A large reason for this is because the technology has only recently matured to a point where researchers can focus on the nuances of supporting technology, and not on creating the enabling technology [77]. In a survey of MR user studies, Dey et al. [66] found that only 15 of the 369 studies conducted between 2005 and 2014 involved collaborative systems, and only a few of those collected communication or behavior focused measures with most of the studies and design frameworks primarily focused on usability [231]. Additionally, there has been very little research focusing on user-based interface design approaches. Until 2004, out of all the publications in primary venues for AR, only 14% of articles were HCI-related and only 8% of those contained any user-based experiments [288]. A more recent review of publications reporting on AR applications for assembly tasks found that while 39% of the articles deal with assembly guidance, only 8% focused on effectiveness evaluation and only 4% conducted usability evaluations [330].

More recent works for MR-mediated collaboration provide a range of contributions, from enabling technology [?] to user evaluations of new interfaces [224] to new interaction modalities [239]. However, even these recent efforts take place in lab settings that don't fully resemble a realistic task and provide a very high level evaluation of the system. Therefore, while existing work has showcased the potential for XR technologies to provide richer collaborative experiences for remote guidance, we don't fully understand how these features impact the quality of remote collaboration since studies of collaborator communication and behaviors have been relatively scarce in the literature. Therefore, there is a need to both design these interfaces with users for real-world tasks and more in depth evaluations to understand the impact of XR features in realistic scenarios [77].

1.4 Thesis Statement and Contributions

In my dissertation, I seek to understand how the additional visual and spatial affordances of XR can best be leveraged to advance the support collaborative XR systems offers its users for physical tasks. A primary objective of this research has been to explore design choices that could break past the constraints the physicality of the task brings. More specifically, I aim to enable remote experts to provide effective guidance that novices can easily comprehend and act upon without the need to accumulate extensive knowledge of the task environment for it. I firmly believe that progress towards this goal is key to making the the use of collaborative systems for physical tasks more sustainable and ensuring access to the expertise required for it over larger areas. To that end, the thesis I have formed posits that:

Designing information representations and communicative mechanisms that minimize the environment awareness required for object-focused communication during remote physical tasks provides the flexibility to create collaborative environments that do not fully replicate physical proximity while enhancing task performance and reducing the joint effort required for collaborative success.

I support my thesis statement through a series of research initiatives that have progressively brought me closer to designing remote collaborative interfaces that go beyond recreating physical proximity while enhancing support for both novices and experts. My overarching research approach has been to first systematically uncover the *actual* – or the current state of the field – to understand the key challenges that keep the collaboration tied to the task environment. I then expand the *adjacent possible* by designing and evaluating mechanisms that leverage the unique capabilities of XR to overcome these challenges and provide novel ways of supporting object-focused communication during physical tasks.

Here I describe how a report of these initiatives are organized across the chapters in this

dissertation and list the specific contributions to the research field they offer.

A Conceptual Framework and Native Model for Collaborative XR

Chapter 2 distills the key theories and models from a variety of domains to present a conceptual framework for designing collaborative systems. This framework describes the relationships between the many factors that influence collaborative success, and outlines the key design goals and collaborator needs a system should support during physical tasks. It also proposes the dimensions of support for communication technology offers, and discusses what I consider to be key mechanisms conducive to building systems that go "beyond being there" [132]. This chapter also goes further to define the capabilities XR offers as a collaborative medium and extends the framework to introduce a native model for XR-mediated remote collaboration that showcases its impact on users' situated experiences.

In addition to being a cohesive resource researchers and practitioners can use to design collaborative systems that support physical tasks, this framework serves as the foundational and theoretical scaffolding that guides the subsequent research efforts in this dissertation.

Empirical Evidence of the Support XR Offers in Time-Critical Settings

A key aspect of understanding where we currently stand in terms of supporting physical tasks with XR is evaluating how the technology fares in realistic scenarios for real-world tasks. Supporting remote guidance for physical tasks, in particular, is most impactful for true novices in complex and time-critical settings where enabling accurate and fast completion of the task is critical. Moreover, the limitations of human cognition and goals of the task in time-critical settings ensure generalizability since features that effectively serve more constrained settings typically remain impactful even in more flexible scenarios. Designing for time-sensitive physical tasks, therefore, provides useful constraints that can help guide the creation of effective interfaces by restricting the seemingly limitless design space for collaborative XR. Additionally, there have been very few research efforts that evaluate the impact of using XR for physical task guidance in

this setting. For these reasons, I opted to evaluate the utility of XR for physical task guidance in time-critical scenarios as a way of understanding both the challenges of designing collaborative interfaces and facilitating communication with XR technology today.

Chapter 4 first set out to understand the utility and feasibility of Mixed Reality guidance cues for novices performing cardiopulmonary resuscitation during an emergency. A user-centered iterative design process highlights a number of key considerations for designing alongside the limitations of technology today, and an evaluation in a realistic setting outlines the benefits the added visual and spatial capabilities of XR have over traditional media for providing procedural guidance to novices in time-critical settings.

Chapter 5, then, provides an understanding of how the visual and spatial capabilities of XR impacts the ease with which remote experts can provide MR-mediated procedural guidance for novices performing trauma surgery. A participatory design study with experts and a subsequent feasibility study with the resulting XR system highlights expert's domain-specific information and communication needs, and discusses how technological constraints impacts the support collaborative systems provide for it.

Both these chapters offers empirical evidence that expands our understanding of the benefits and limitations the visual and spatial capabilities of XR bring to remote collaborative physical tasks. Chapter 5, in particular highlights the key challenges associated with supporting the situational awareness required for experts to partake in object-focused communication.

A Communication-Focused Overview of the Design Landscape

A systematic understanding of the design space existing XR-mediated collaborative systems occupy is a crucial step towards furthering the field and enhancing the support available for object-focused communication in a methodical manner. Chapter 6 presents a review of the field and analyzes the communication-focused features of existing systems that enable working together on a physical task from a distance today. An in-depth analysis of the support offered for referencing – the communication process most tied to the physical task environment, as well as

the situated experience associated with experts' task environments uncovers key research gaps that could drive future design explorations.

The Design and Evaluation of Interaction Techniques that Reduce the Joint Effort Required for Object-Focused Communication

Armed with an in-depth understanding of the phenomenon of collaborative XR, the nuances of supporting physical tasks, the current design landscape for it, and the key challenges ahead of us, the final research thread in my dissertation sought to expand the repertoire of design choices available to practitioners with interaction techniques that reduced the effort required from collaborators for object-focused communication.

Chapter 6 presents two design strategies that ease the referencing for collaborators during physical tasks – explicitly representing the spatial information required for it, and (partially) offloading aspects of the communication process to the collaborative system. It also presents a controlled experiment that evaluates two interaction techniques that embody these strategies, and discusses the impact explicit representations and offloaded communication has on experts' need for environment awareness and spatial information during collaborative physical tasks. A key takeaway of this exploration is that spatial information can be superfluous when communicative responsibilities are shifted from the expert to a spatially-aware collaborative system.

Chapter 7, then, exploits this key takeaway to design an immersive interface for remote experts that is hybrid between their own familiar workspace in the real-world and the physical task environment the novice in need of guidance is operating from. A comparative study evaluated the use of this interface against one that fully recreates the novice's environment on both single and repeated use to uncover the impact a familiar and spatially-stable interface has on the experts' ability to provide guidance. Results showcase that such an interface leverages expert's existing spatial memories to enhance their ability to provide effective procedural guidance while maximizing the benefits of the immersive nature of XR.

These two chapters expand the design space for collaborative XR and demonstrate how

remote physical tasks can be supported without fully replicating physical proximity. It also showcases the benefits of designing interaction techniques minimize the experiential overhead remote collaborators need to undertake to use their expertise during collaborative physical tasks. In particular, chapter 6 and 7 offers alternate communication mechanisms and information representations that individually and collectively reduce the task load and visual search both novices and experts need to perform within the task environment to sustain object-focused communication while increasing task performance and communication efficiency.

Overall, the work in this dissertation provides a conceptual lens and specific design ideas that empower practitioners to create rich collaborative experiences for physical tasks that go beyond the constraints of face-to-face interactions while making it easier and more sustainable to scale real-world expertise.

Authorship Statement

Although I am the principal author of the research discussed in this dissertation, it is also the culmination of years of collaboration with my advisor, Nadir Weibel, and my colleagues at the University of California San Diego. To acknowledge the contributions of my collaborators, I use first-person plural pronouns in the following chapters when appropriate.

Chapter 2

A Conceptual Framework for Designing Collaborative Systems

Thesis Question: What are the design considerations associated with supporting collaborative physical tasks at a distance? How does Extended Reality impact the experience of remote collaboration?

In this chapter, I present a conceptual framework that can serve as a comprehensive guide for researchers and practitioners looking to design effective remote collaborative systems. The framework includes summative models that provide an overview of the various communications mechanisms at play during collaborative physical tasks, and isolates the user needs and design goals to be supported for collaborative success. It also introduces and discusses the dimensions of communication support technology can offer to outline the design space for remote collaborative systems. Finally, it defines the capabilities of XR as a collaborative medium and introduces a native model for XR-mediated collaborative physical tasks that highlights its impact on collaborators' situated experiences.

2.1 Collaboration through Mediated Communication

At its core, collaboration is about people communicating with each other to coordinate the action required to achieve a common goal. Therefore, collaborators' attention and efforts during any task are split between two functional tasks and spaces [50]. The first is the *domain task* that occurs within the task space where effort is spent towards completing the task at hand. The second is the *communication task* that occurs within the communication space and involves exchanging and understanding information between each other. The two heavily influence each

other – complex domain tasks increase how much effort collaborators’ need to put into the communication task, and the difficulty of engaging in communication makes completing the domain task harder. However, the complexity of the domain task isn’t fundamentally dependent on the technology, and the effort it takes to partake in the communication fundamentally defines the success and adoption of a collaborative system [111].

Given that, **the primary design goal for any collaborative system should be to reduce the effort required for the communication task.** Collaborative physical tasks, in particular, require extensive and complex coordination that is intricately tied to the task environment [174, 50]. Designing interfaces for geographically separated collaborators, therefore, demands an understanding of both the communication processes required for this coordination and how technology can support it.

There are a plethora of research articles and design resources that help us better understand how we make sense of the world, the various communication mechanisms at play, and how distance and technology impacts them [1, 337]. However, isolating the relevant design factors and how they relate to each other is non-trivial with relevant research being extremely fragmented and vastly multidisciplinary [115, 243] spanning multiple fields from psychology, philosophy, and linguistics to cognitive science and computer-supported cooperative work (CSCW). In the rest of chapter, I discuss key takeaways from communication and cognitive theories to isolate the various factors that need to be considered when designing for remote collaboration on physical tasks. I distill these factors into summative models for both collocated and technology-mediated collaboration and outline the collaborator needs that need to be supported for a successful collaboration. I then introduce a model for XR-mediated collaboration that highlights how the visual and spatial capabilities of the technology impact collaborators’ embedded and embodied experiences during collaborative physical tasks.

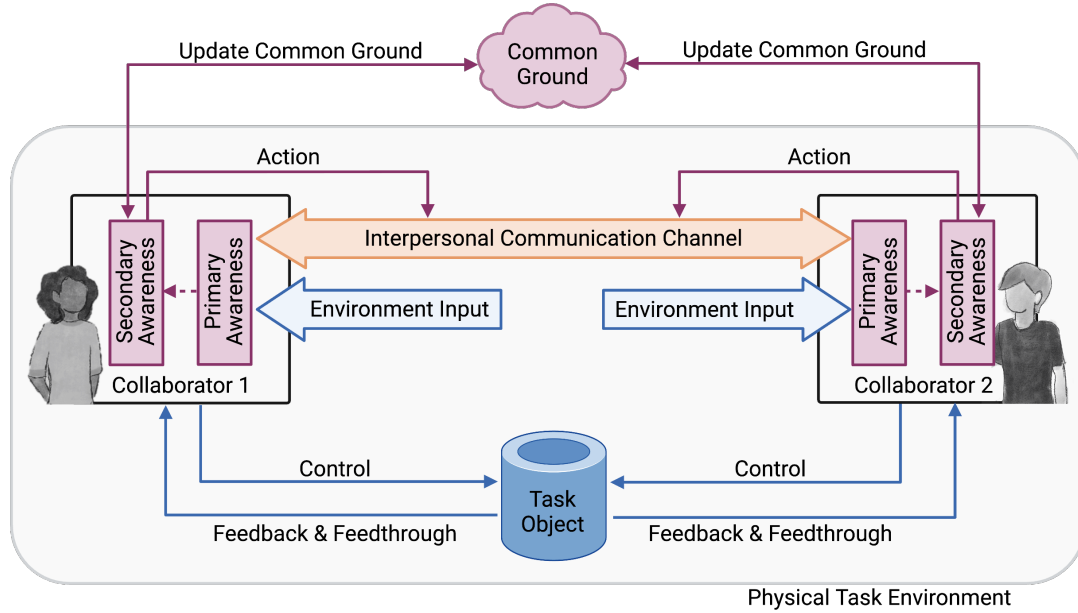


Figure 2.1. A conceptual model for collocated collaboration where two collaborators share the physical task environment. Depicted is the interpersonal communication channel in orange, the environment input and interactions with task objects in blue, and cognitive-physical cycle of generating awareness, building common ground, and taking action in pink.

2.2 The Mechanics of Communication for Physical Tasks

Physical tasks where people perform actions on concrete objects in the real world [94]

The mechanics of communication provides an overview of the various concepts that come into play during a collaborative physical task that drives the communication task forward.

INTERPERSONAL COMMUNICATION CHANNEL

A number of existing models conceptualize communication from a primary lens of how information moves between collaborators. Classic models of communication under information theory like the Shannon-Weaver model of communication [270] and Berlo’s SMCR model [23] describe communication between two people as information that flows between a sender and a receiver through a channel or medium. While these models tend to treat communication as a one-way process, others like Schramm’s Model of Communication [268] treats communication as a circular process where all participants are both receivers and senders. In general, most models

that focus on information flow and those that derive from it retain the concept of an *interpersonal communication channel*. The model in Fig. 2.1, depicts the interpersonal communication channel in orange as bi-directional arrows between collaborators through which both verbal and non-verbal communication is exchanged.

ENVIRONMENT INPUT AND THE ROLE OF TASK OBJECTS

Physical tasks in particular, are inherently spatial since the work is primarily carried out on or through real objects in the task environment. Collaborators also use the physical features and spatial relationships of these *task objects* in the direct conversation they engage in with each other. Both the domain task and the communication task for physical tasks is therefore intricately tied to the position and relations between collaborators, objects, and their environment [174, 226]. In general, collaborators observe, make decisions, and act on the *environment input* they receive from their physical world to partake in the collaboration.

The model in Fig. 2.1 includes elements from the people-artifact framework [68, 84] and depicts the functional relationship between the task objects in the environment and collaborators in blue. These objects are acted upon by one or more collaborators who are also able to perceive the state of the task through their appearance, physical features, and spatial relationships. For example, when a collaborator acts on or *controls* a task object, they experience *feedback* from the object itself. The other collaborator then observes the effects of this action through a phenomenon called *feedthrough*.

AWARENESS, COMMON GROUND, AND COMMUNICATIVE ACTION

Individual contributors of a physical task engage in a sophisticated process of generating awareness, building common ground, and taking purposeful and meaningful action to move the task forward. Models like the perception-action cycle [95, 2] and the human-action cycle [222] emphasise how we make sense of the world through interaction with the people and artifacts in it. The key aspects of these theories are captured in Fig. 2.1 in pink and can be divided into

two processes – the *cognitive cycle* which describes the mental aspect of communication that isn’t directly observable, and the *physical cycle* which talks of the observable physical-bodily aspect of communication where collaborators make decisions and take actions based on their understanding of what needs to be done.

The Cognitive Cycle and Information Needs

A major focus of existing literature has been how collaborators generate situational awareness [73] and the mutual knowledge required for successful communication [52]. One of the most notable concepts raised by Clark and Brennan is that of grounding in communication [51]. Grounding is the interactive process of establishing *common ground* – knowledge that collaborators have in common (and are aware that they do) – without which there can be no collaboration. Collaborators continuously accrue common ground based on the information they have available to them as the communication and the task progresses [226].

Grounding fundamentally requires collaborators to have an awareness of their environment, of the state of the task, and of their collaborators’ actions and intentions. Simply put, awareness is “knowing what is going on” [78] and answers “who, what, where, when, and how” questions [121] using information gleaned from the communication channel or directly from the environment itself. The human perspective on how awareness is achieved has been covered in a number of theories and models [219, 121], with Endsley’s model of situational awareness [78] being highly cited in CSCW. These models often describe awareness in a hierarchical fashion with the perception and comprehension of lower level information leading up to compounded, higher levels of awareness [285].

We capture the essence of these hierarchies to outline collaborators’ information needs during physical tasks by classifying the different elements that constitutes awareness into two broad categories (Table 2.1) – *primary awareness* which refers to what collaborators can directly observe from the task environment and their collaborators, and *secondary awareness* which constitutes higher levels of awareness.

Table 2.1. Classification of the constituents of awareness necessary during collaborative physical tasks.

Key Constituents of Awareness During Physical Tasks		
Primary Awareness	Primary Environment Awareness	Refers to awareness of the task environment and the task objects (like tools) within the environment. This includes being able to observe the active task space, being aware of the presence of task objects, and knowledge of the spatial layout between collaborators and objects in the environment.
	Primary Collaborator Awareness	Refers to the ability to perceive other collaborators working on the physical task. This includes an awareness of who is there (presence), what they are saying (speech), where they are (location), where they are looking (gaze), what they can see (view), what can they reach (reach), and what they are currently doing (action).
Secondary Awareness		Refers to knowledge that is inferred from the information that makes up primary awareness and to make up higher levels of awareness and understand their significance with respect to the context at hand. For example, task status awareness is a secondary form of awareness that can be gleaned from knowledge of the task objects and collaborator actions. Another form of secondary awareness could be inferring collaborator intent based on their gaze.

The Physical Cycle and Communicative Needs

Collaborators' awareness and perception of common ground influence the communicative *actions* they take in the physical cycle that is then relayed to each other through the communication channel. These actions can be both verbal and non-verbal. The verbal spoken dialogue exchanged between collaborators is typically a crucial aspects of interpersonal communication. People also use non-verbal forms of communication that include (but are not limited to) facial expressions, posture, eye gaze, head direction, movement, and hand gestures. These non-verbal gestures can be illustrative or indicative (when people gesture as a part of conversation, either to emphasize or illustrate aspects of dialogue - like pointing) or emblematic (where the gesture is made in place of a speech act - like a thumbs up for "okay").

Typically, the more common ground collaborators have, the easier it is for collaborators to communicate and work together, and the more it allows them to use mechanisms that make

communication more seamless. If collaborators realize that their action isn't understood by others, they will engage in further action to repair any misinterpretations and revise their assumptions about the common ground [51]. However, this can be quite taxing, and theories like Collaborative Discourse Theory [252, 53] and the Principle of Least Collaborative Effort [54] emphasize that people will preemptively choose actions that minimize the collaborative effort required to present and understand it during a collaborative task. For example, experts will use domain-specific jargon if they assume others have a similar expertise level, but will switch to novice-friendly language if they sense that their actions are not being understood. Additionally alternative descriptions like “the box from the last step” and deictic expressions like “give me that one” that can be disambiguated with context become possible with shared space and history as the task moves forward.

Now, communication between an expert and a novice during physical tasks is primarily about giving and receiving procedural guidance which heavily revolves around referring to objects and describing actions to be done on them [93]. Observations of communicative actions during physical tasks [87, 114, 94, 351, 197] reveal that experts and novices engage in four crucial communication processes (Outlined in table 2.2): referencing [44, 147], communicating technique [94], acknowledgement, and outlining strategy [69, 53].

Table 2.2. Key communication processes collaborative physical tasks rely on.

Key Communication Processes During Physical Tasks	
Referencing	The process of referring to target (task) objects in the environment, and having that reference understood by others.
Communicating Technique	The process of describing actions or manipulations to be performed on those target objects and having that understood by others.
Acknowledgement	The process by which collaborators communicate a confirmation of their understanding of a message or of an action being performed successfully.
Outlining Strategy	Collaborators also engage in communicative action that serves to provide context for individual actions, discuss, or plan the task going forward.

2.3 Technology-Mediated Collaborative Physical Tasks

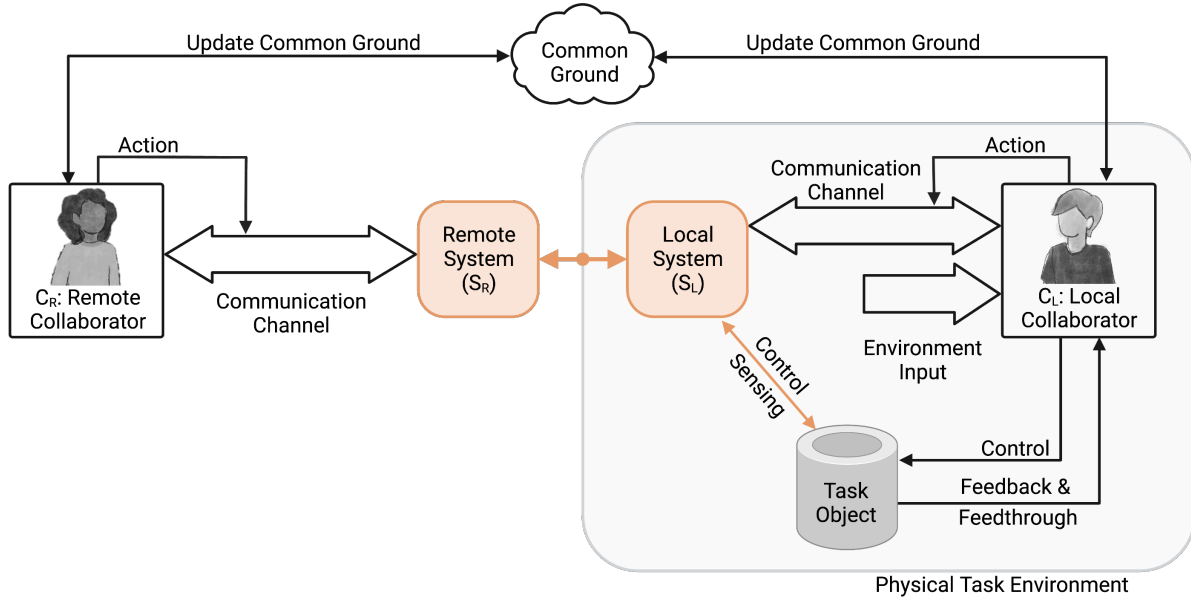


Figure 2.2. The impact of distance: A conceptual model of the mechanics of technology-mediated collaboration during physical tasks.

Advancements in technology have allowed us to circumvent the constraints of natural human communication [337] and partake in synchronous collaboration across distances in meaningful ways. While a shared physical space affords people symmetric and easy access to information (Fig. 2.1), remote collaborators lose access to information about each other and geographically separated collaborators in particular lose access to information about the task environment [283]. Fig. 2.2 models this experience of technology-mediated communication and depicts the collaborative system in orange, with the system (S_R) used by the remote user (C_R) shown as separate from the system (S_L) used by the local user (C_L).

Notice that the interpersonal communication channel through which coordination occurs in collocated settings is replaced by individual communication channels between each collaborator and the system they interact with. Additionally, collaborators removed from the task environment no longer have the ability to directly control or receive feedback or feedthrough from task objects. This means that the remote system (S_R) becomes the sole source of information

its user (C_R) uses to generate the environment awareness [73, 78] required for the object-focused communication [51, 54] the collaboration relies on. The local system (S_L) typically *senses* some input from the task environment that can be passed on to the remote user. And while not always necessary, it could also have the ability to *control* task objects (for example, remote collaborators could control a robot arm to manipulate objects in the task environment). The local system's interaction with task objects is thus depicted as a bi-directional arrow.

In general, the design space of the features that can make up a collaborative system is driven by the collaborator needs it caters to and the technology that enables it. These features, in turn, impact the ease with which users can partake in the communication task as well as their situated experience during a collaboration session (Fig. 2.3). We discuss these aspects in the remainder of the session.

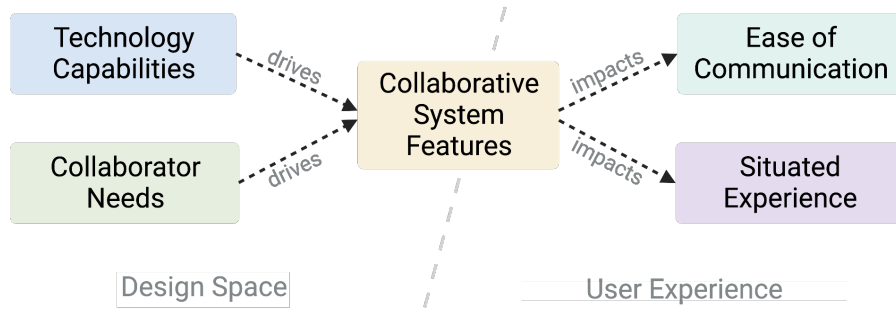


Figure 2.3. Every collaborative system is designed based on the collaborator needs and technology capabilities available, the features of this system in turn impacts the ease with which collaborators can communicate and the situated experience associated with it.

2.3.1 The Impact of Distance on Communication Bandwidth

When collocated, people enjoy a dynamic establishment of shared context, a natural recognition of non-verbal cues, and an intuitive interchange of focus between their communication task space (with mutually viewable objects) and their fellow collaborator(s) – all which add to the ease of establishing common ground and the richness of the communication in collocated settings. These communication benefits, to a large extent, are due to the richness and amount of verbal and non-verbal cues that can be shared and interpreted by collaborators [174, 337]. In

other words, the extent of information exchange that being in the same physical space allows for makes face-to-face communication a “high bandwidth” setting.

The bandwidth metaphor, which categorizes a communication medium based on how much information can be transmitted and received by the five human senses or input streams [340] has been integral to some of the earliest work on mediated communication and is most notably seen in variations of the media richness theory [58]. Technological constraints like computational capabilities and network throughput limit how much information can be relayed and the repertoire of actions a user is afforded by a communication channel. For example, a remote collaborator is only aware of aspects of the environment that the technology is designed to transmit. Similarly, a local user using an audio-only system cannot sense and understand non-verbal gestures even if the remote user is interacting with a video conferencing system, forcing the collaboration to rely on verbal communication.

In general, technology-mediated communication typically comes with a reduced view of the task environment, diminished cues from collaborators, and limited ability to use gestures [297]. This added layer of distance and technology induced constraints disrupt the ordinary flow of information extraction and situated cognition that is taken for granted in collocated settings [226] – it introduces seams and discontinuities in the collaborative experience that change how people communicate as they are forced to learn new ways of working that accommodate the technology [140]. For example, Gutwin found that people started working more independently and waited longer to take their turn to cope with technology-related delays in the communication [120]. When comparing face-to-face and video-mediated communication, Doherty-Sneddon et al. [69] also found that video-mediated communication had more formal turn-taking with longer turns, fewer interruptions, was less spontaneous, and non-verbal gestures were harder to convey and tended to be ignored.

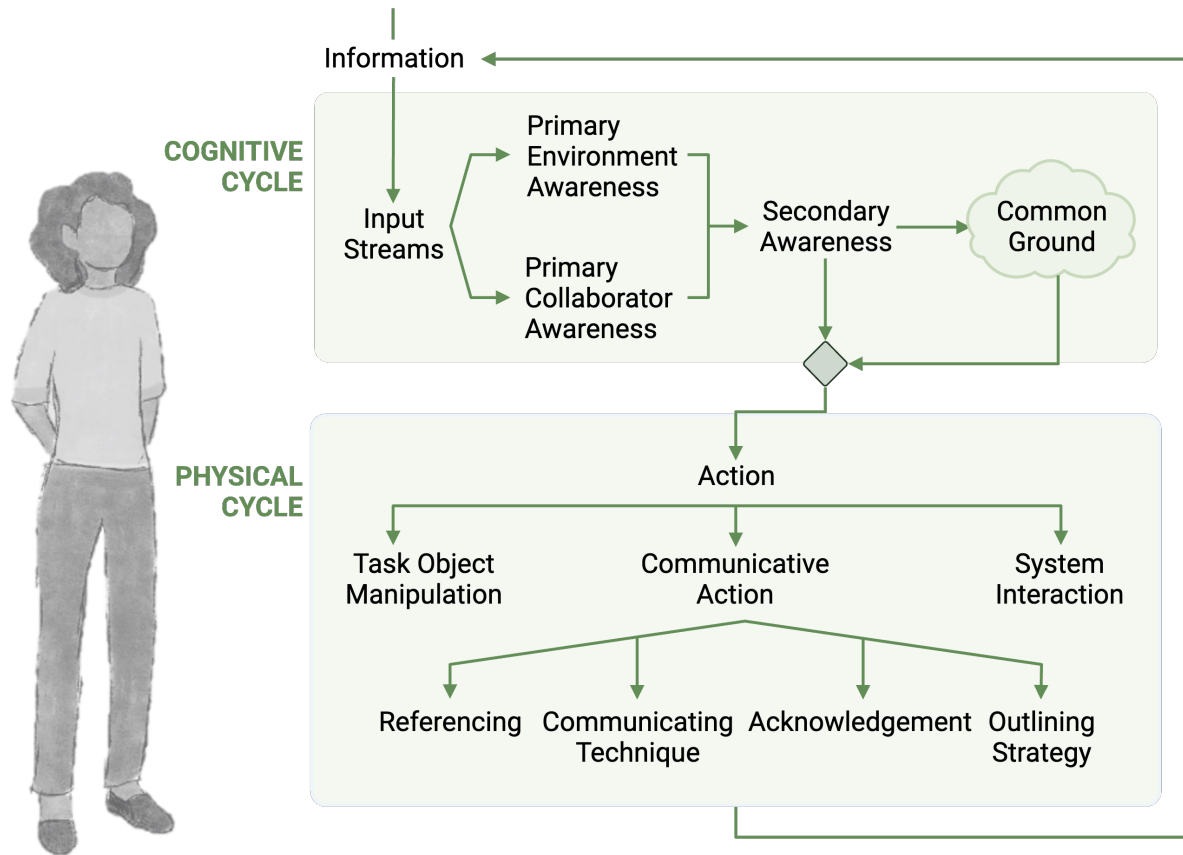


Figure 2.4. Information from both the communication and environment channels is sensed by the five *input streams* (visual, aural, tactile, olfactory, gustatory) and used to build awareness and common ground in the cognitive cycle (top), which in turn influences collaborator decisions and the actions they take in the physical cycle (bottom). The cognitive cycle drives collaborators’ information needs and the physical cycle specifies what actions need to be supported.

2.3.2 Collaborators’ Communicative Needs

Facilitating mechanisms for rich interpersonal communication during remote collaboration is a non-trivial feat, and building effective systems requires careful consideration of the individual user needs the remote or local system should cater to. While the exact user needs will vary with the specific scenario and task the system is designed for, the cognitive-physical cycle (discussed in Sec. 2.2) that outlines how collaborators make sense of and act on information can serve as a guide for the key information and communicative needs collaborative systems need to design for. This process is also visualized in Fig. 2.4 – The cognitive cycle drives collaborators’ information needs and the physical cycle specifies what actions need to be supported.

- **Action Needs:** When collaborating across distances, each user can engage in *task object manipulations* that drive the domain task forward, *communicative actions* that move the communication task forward, or *interface interactions* that control their experience with the system. Designing support for communicative actions, in particular, is especially crucial as no communication can take place without it.
- **Information Needs:** Collaborators' ability to take actions, however, fundamentally revolves around their ability to obtain and make use of information pertinent to the task at hand. Now, not all elements of the task environment or collaborators are equally important for a physical task. For example, it is often more important for collaborators to be aware of state of the task objects than it is for them to be able to see each others faces during physical tasks. Designers should, therefore, determine what information is necessary for each user to build the awareness required for the communication processes physical tasks rely on [78, 174].

2.3.3 The Impact of the Medium on Support for Communication

The communication mechanisms collaborators can partake in is heavily dependent on the specific medium or technology employed for each system and how effectively it can convey different types of information and user actions. It is however important to note here that people are very good at adapting their communication styles and modifying the mechanisms they use to suit the medium at hand [314]. Beyond a surprisingly low threshold, the communication bandwidth and the technology don't impact the successful completion of a task [176]. For example, experts successfully guide novices through complex medical procedures over an audio call today [291]. What changes, however, is the *ease* of partaking in communication [226] and the perceived quality of the collaboration.

The extent to which a technological system allows its users to partake in communication processes is fundamentally determined by two factors or dimensions – the effort required to engage in communicative action, and the effort required to achieve the awareness needed to be

able to take that action. In this framework, we divide the level of support offered to individual collaborators into three broad categories across both dimensions (Fig. 2.5). We assume a verbal channel that allows collaborators to talk to each other in real-time is a baseline level of support systems provide for both communicative action and achieving awareness. Systems can also enable communication processes through dedicated features that provide enhanced support [73].

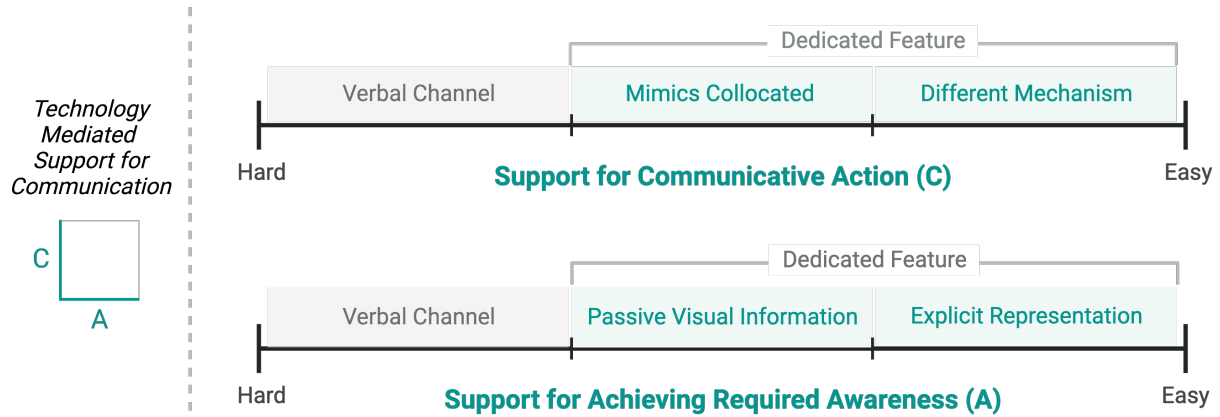


Figure 2.5. Dimensions of support for communication with technology-mediated collaboration.

SUPPORT FOR COMMUNICATIVE ACTION

- **Mimics Collocated Mechanisms:** One way to ensure a high level of support for different communication processes is to design interactions that enable users to partake in communicative actions that feel natural and intuitive. These techniques, therefore, aim to *mimic the communication mechanisms used for it in collocated settings*. For example, being able to gesture at or point at task objects with one's hand allows remote users to refer to using deictic expressions that can feel easier than having to verbally describe its location to the local user.
- **Provide Different Mechanisms:** While it is undeniable that it allows for extremely rich communication and might even have aspects that cannot be replicated [226], despite being treated as a gold standard [141], face-to-face interactions are not perfect. Moreover, what is possible in one medium might not be possible in another and different types of technology

can mediate communication and present information differently. The unique affordances of technology, then, can be leveraged to improve support for communication in new ways that go beyond what is possible in collocated settings – a concept explored by Hollan and Stornetta in *Beyond Being There* [132]. For example, pointing to a cluster of objects in collocated settings can introduce ambiguity about which object is being referred to. However, inferring the target object using gaze tracking and modifying how the pointing gesture is represented to the remote collaborator can remove this ambiguity [280] and make it easy for experts to refer to things more easily. Therefore, the design space of features that support communicative action also includes those that facilitate it through mechanisms that are different and theoretically easier than collocated interactions.

SUPPORT FOR ACHIEVING AWARENESS

- **Passive Sources of Information:** Features that relay information can be *passive* sources of information where visual information is captured by the system and relayed to the user without additional computational effort. Users typically have to process the information to isolate the aspects relevant to the task at hand.

Given the importance of visual information for communication during physical tasks [174, 315], providing features like video feeds of the task space is a key method collaborative systems use to support their users' need for information. While extremely useful in enhancing the level of support for communication, video feeds are passive sources of information as users have to comprehend where different task objects are, or where their collaborators are looking by further processing what the video is showing.

- **Explicit Representations of Information:** One way users fail to achieve the awareness they need even when the system transmits it is when the information is presented to them in a form that makes it hard for collaborators to glean the details relevant to them [121]. Since it is the system that transmits the information between collaborators, it can choose to alter both what is mediated and how it is presented to the user.

Taking advantage of the discontinuity and separation between the remote and local systems to provide information in a manner that reduces the amount of comprehension and information processing users have to engage in is another opportunity to help collaborators *go beyond being there*. Dedicated features that provide information, therefore, can do so through *explicit representations* that extract aspects relevant for collaborators and present it to them in an easy to understand manner without additional distracting peripheral information [283]. For example, visual indicators for collaborators' gaze [244] or view [240] are examples of explicitly representing information that collaborators have usually have to gauge from a view of the task space.

2.3.4 The Impact on Collaborators' Situated Experience

Our ability to make sense of the what we are experiencing comes from an active and pragmatic engagement with the world around us [217], and the importance of the environment around us for cognition is underscored in theories like the situated cognition or the embodied-embedded hypothesis [341, 167]. A core claim with these theories states that how we think and act in a situation cannot be uncoupled from the environment that it occurs in. In modifying the environment collaborators operate in, technological systems also impacts impacts users' situated experience during the collaborative task (Fig. 2.3). For example, we constantly structure the environment we work in and partake in epistemic actions to reduce our cognitive load in a given scenario [169]. Additionally, the tools we have available to us fundamentally change how we engage with the task at hand [168].

The specific capabilities of the collaborative medium and the perceptual and interaction richness it offers, therefore, will define how collaborators interact with the task environment and each other. Moreover, different technologies will allow for different mechanisms users can use to offload and reduce their real-time cognition. Designing effective systems with a particular technology therefore requires a thorough understanding of the capabilities or affordances it brings as a collaborative medium. Affordance refers to interaction possibilities of with respect to an

artifact or environment [106, 109]. This concept of affordances – most notably applied to design by Norman [221] – has been used to design intuitive systems based on what the technology offers in terms of the user experience.

2.4 The Capabilities of XR as a Collaborative Medium

Unlike screenbound interfaces, collaborative systems that leverage ubiquitous and emerging technologies like XR allow users to act within the interface itself – fundamentally changing how users encounter the collaborative space [339]. In this section, we define the various capabilities XR offers as a collaborative medium capable of supporting a myriad of complex interactions and group them based on what they offer in terms of the user experience. This organization is influenced by the key capabilities of MR as a collaborative medium identified by Szalavari et al. [289] and is summarized in Table 2.3.

Present Multi-Sensory Virtual Content (1a-e): A primary capability of MR is that it can present content and stimuli over the full range of human input streams including two or three-dimensional (2D or 3D) visual content, audio signals, haptics, and virtual smell or taste cues [274]. Presenting 3D visual content, in particular, allows for full depth perception of a virtual object and enables users to perceive it in a similar manner as they would a physical object. Additional stimuli, like haptics, can further increase the realism of virtual objects being presented to the user.

Modify User’s Experience with the Real World (2a-e): MR allows users to have varying levels of visibility and engagement with the virtual content. This includes the ability to augment or diminish the real world with information overlays or by occluding aspects from the users view. It is also possible to completely block out the real world, as with VR devices today. Additionally, MR can modify how reality is perceived by choosing to represent aspects in ways that are not completely true to its nature. For example, spatial relations between objects can be preserved without preserving the absolute distance between them. Moreover, the content

Table 2.3. XR capabilities grouped according to what it offers in terms of the user experience.

Technology Capabilities of Mixed Reality	
1. Present multi-sensory virtual content	(a) Present 2D visual content (b) Present 3D visual content (c) Present audio signals (d) Present haptic feedback (e) Present virtual smell and/or taste
2. Modify user's experience with the real world	(a) Augment the real world with virtual content real (b) Remove or diminish aspects of the real world (c) Warp the presentation of reality (d) Block out the real world (e) Enable the experience of remote locations
3. Enable spatial interfaces and understanding	(a) Spatially distribute virtual content around the user (b) Spatially anchor virtual content to the real world (c) Detect and track objects in the environment (d) Track user location and orientation within the environment
4. Enable human understanding	(a) Understand natural human language (b) Detect user's non-verbal behavioral and social cues (c) Detect user's physiological and emotional state
5. Provide independence of experience	(a) Personalize virtual content and user interface (b) Allow freedom of movement within environment (c) Allow freedom of perspective within environment (d) Provide ability to modify environments

MR presents can also be used to recreate the experience of being in a real physical space that is geographically separated from the user (for example, through 3D reconstructions or digital twins).

Enable Spatial Interfaces and Understanding (3a-d): MR allows us to move beyond screen-bound experiences and offers the opportunity to distribute the content presented to the user in the 3D space around them. This goes beyond just visual content - for example, users can experience spatial audio in MR environments. Content can also be anchored to particular points in the real world, making the digital-physical experience more seamless and allowing the

interface to use the context of the environment the user is in. This environment understanding MR technology offers is often enabled by building real-time updated maps using a process called Simultaneous Localization and Mapping (SLAM) and most MR devices today detect surfaces like a floor or table with relative ease. A more specific form of this understanding includes detecting and tracking the presence of various situation-relevant objects and their movement in space (often with additional motion capture algorithms, or tracking devices mounted on the objects). MR devices can also keep track of the user's location within the environment in real time (typically with head tracking and indications of where the user is looking).

Enable human understanding (4a-c): MR technologies offer the possibility of using a deeper understanding of human interactions and social behavior to drive the design of an experience. Technology can capture subtle aspects of user behavior that the interface can react to. For example, a users speech patterns can be used to provide additional experience cues, their body language and actions can be used as proxies for intent, and detecting their facial expressions or physiological states could drive emotion-based design.

Provide independence of user experience (5a-c): MR allows for tailored experiences catered to the individual users allowing for different virtual content to be experienced even within the same physical space [289]. Users can also experience a freedom of movement within the 3D MR environment. In addition to the freedom of perspective this offers, MR also allows users to view a location from multiple perspectives - for example, they can view it from a first person perspective, a dollhouse perspective, or even from the perspective of a third entity. Users might also be able to modify aspects of their own or remote environments.

2.5 A Native Model for XR-Mediated Remote Collaboration

Fig. 2.6 highlights how the spatial and perceptual differences of XR-mediated virtual content changes the collaborative experience during physical tasks.

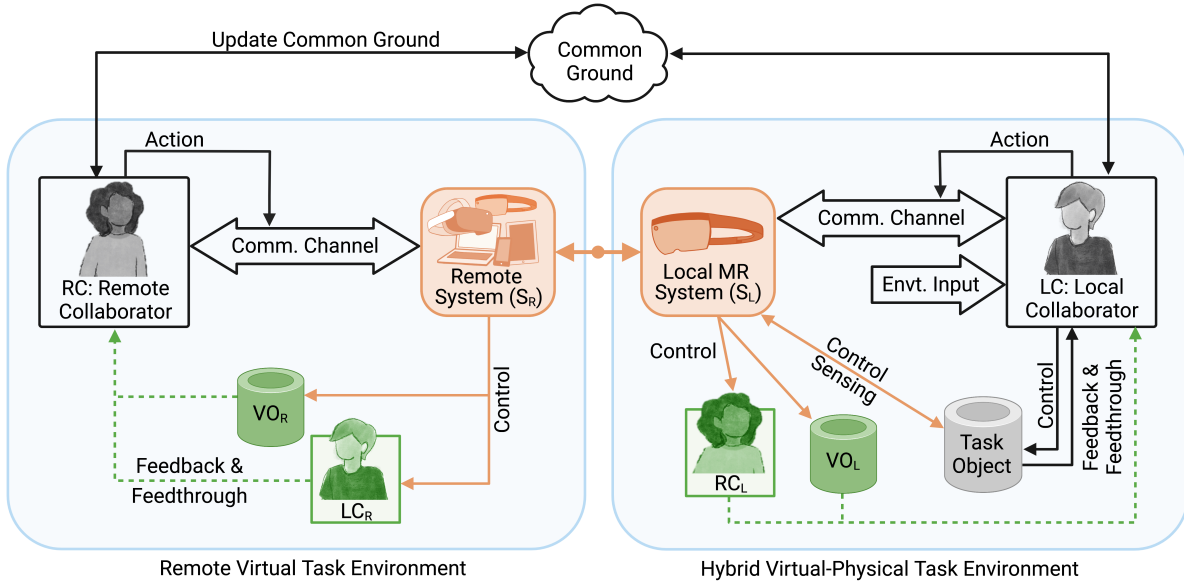


Figure 2.6. A conceptual model for XR-mediated collaboration.

2.5.1 A Hybrid Virtual-Physical Space for Local Users

MR has been identified as the medium of choice for collaborators in the physical task environment [26, 289] because it allows users access to the physical objects while providing digital content in-situ and because it allows for behaviors that are much more like face-to-face communication when compared to traditional screen-bound collaborative systems. Because of this, we fix MR as the technology the local system is built on (Fig. 2.6, Right). Additionally, MR also allows for the representation of collaborators in ways that feel like they are in the same physical space as the local user. We also integrate this by depicting a representation of the remote user (RC_L) in the local users space. The combination of immersive virtual content and physical objects in the local user's space make up a virtual-physical hybrid task environment

Compared to traditional screen-bound experiences, immersive and spatial technologies like Mixed Reality (MR) and Virtual Reality (VR) pair especially well with the spatial nature of physical tasks and facilitate more natural interaction and communication behaviors [26]. MR's ability to understand and present information within the environment grants users instructional and gestural cues from remote experts while maintaining environment context. This makes

it an effective communication medium for novices and has been shown to allow for efficient communication, increased feelings of presence, and decreased cognitive loads among both local and remote collaborators [323, 18].

Notice how the virtual objects and the representation of the remote user are both controlled by the system (depicted by the uni-directional arrows from the system to it). Any manipulation of virtual objects or interactions with representations of remote collaborators is facilitated through the system - the user loses access to them without the system. However, since they can be perceived in ways similar to physical objects, we include this experiential feedback and feedthrough using dashed arrows.

2.5.2 A Virtual Task Environment for Remote Users

The physical environment the remote collaborators are in typically has no significance to the task at hand. This then doesn't constraint the type of technology used in remote spaces – we depict this flexibility for the remote system in Fig. 2.6 (Left). However, the inclusion of immersive and mixed reality technologies mean that the remote user could also interact with virtual objects and remote collaborator in ways similar to physical, colocated settings. This is depicted through the addition of virtual objects (VO_R), a representation of local collaborator (LC_R), control arrows from the system, and experiential feedback and feedthrough arrows to the remote user.

THE SITUATED EXPERIENCE OF VIRTUAL TASK ENVIRONMENTS

Since the physical environment of the remote collaborator is of little importance to the task, we don't include any physical objects on the remote side. However, the addition of immersive virtual content fundamentally changes the experience for the remote user and is called the virtual remote environment. In general, the freedom of technology choice for remote experts means that there is substantial variety in the situated experience offered by the *virtual task environments* and can be defined by 3 key factors or dimensions:

- **Coverage of the Physical Task Environment:** The virtual task environment can offer different levels of visual coverage of the physical task environment the local user is in. For example, remote collaborators can be immersed in a replica or mirror of the physical task environment, or the environment can be something completely different.

Many physical tasks have an *active task space* where a bulk of the action or object manipulations that drive the task forward occurs. For instance, the surgical field and the car engine are the active task spaces for surgery and car repair respectively. Virtual task environment can offer remote users a view of just the active space, more than the active space (partial space), of the full task environment.

- **Spatial Presence in the Virtual Environment:** It is important to note that unlike the hybrid local environment, an immersive remote environment is not a mandatory aspect of the collaborative experience for it to be considered MR-mediated collaboration. The virtual environment for remote users can be non-immersive (if the remote system is on

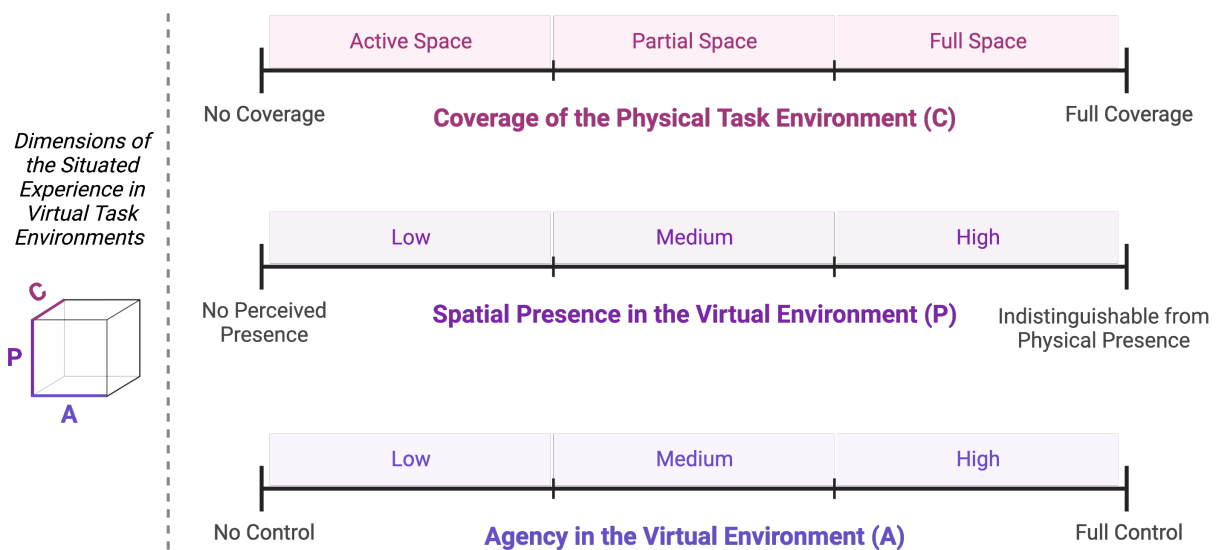


Figure 2.7. Dimensions of the situated experience provided by the virtual task environment or the interface for remote experts. A virtual task environment can have varying levels of visual coverage of the physical task environment as well as varying levels of spatial presence and agency remote users can experience within it.

a tablet or desktop computer, for example), or can be on the virtuality continuum with technology ranging from augmented reality to fully immersive virtuality.

Given that, the remote system can offer its user varying levels of spatial presence in the virtual task environment. While spatial presence can be influenced by a plethora of factors including technological immersion and realism of the virtual content, we use the term to refer to the user's subjective feeling of being physically present within the virtual environment or the space rendered by the medium [273]. This most closely resembles the concept of self-location or the place illusion commonly associated with virtual reality [275, 276].

- **Agency in the Virtual Environment:** Remote collaborators can also have varying levels of agency in the virtual environment. Agency refers to the sense of control a user has over their virtual environment. Virtual task environments with low agency are those where the remote user is a passive viewer of the virtual content around them. A high level of agency would mean that they have a lot of control over their surrounding environment and can change or modify the virtual content around them.

2.5.3 Asymmetric Experiences of the Mutual Collaborative Space

Unlike collocated settings where all collaborators share the same physical space, collaborators fundamentally operate in their own spaces with technology-mediated collaboration. However, technology also allows them to share aspects of their spaces with each other. For example, a remote collaborator can virtually share the physical task environment with the local user if they are immersed in a replica of it, and a local collaborator may be able to see a representation of the remote user interacting with virtual content in their space. This shared ground is called the mutual collaborative space in our model. However, since this mutual collaborative space is facilitated by the collaborative system, collaborators are still fundamentally operating in their own space and can experience both their own space and the mutual space independently and differently. For example, the remote collaborator can walk around the virtual replica of the task

space without any intervention from the local collaborator, making their experience of the space different from the local users.

Additionally, in traditional technology-mediated collaboration, remote collaborators have the option to control physical objects in the local environment through the system. XR democratizes this control over each others environments by allowing both remote and local collaborators to modify and create virtual objects in each others spaces - fundamentally changing the perceived environments collaborators operate in and possibly impacting how others think and act during the task. In our model, this control would be an action taken by, say, the remote collaborator, which is sensed and transformed into action within the local space via the collaborative system.

2.6 Design Goals for Remote Guidance

One of the key characteristics of remote collaborative physical tasks is the difference in roles the collaborators play. Remote collaborators are typically domain experts who are there to use their tacit knowledge to provide the help novices physically in the environment need to successfully complete the task. It is, therefore, important to consider how the different skills experts and novices bring to the table can be combined in the most effective manner regardless of the collaborative medium. Additionally, while novices are perceptually located in the real environment, remote experts are perceiving the workspace through technology. This asymmetry in their access to the physical environment means the shared task environment is dependent on how much of it is effectively shared with the remote user and impacts how each collaborator experiences a collaborative sessions. Designers, therefore, need to be cognizant of how these asymmetries impact the collaboration and design to bridge the gaps it brings.

When it comes to experts, **the primary design goal for the remote system is to make it easier to provide procedural guidance.** More specifically, the system should help experts make inferences about the task environment and the novice to know what guidance needs to

be provided and when [176] – and it should do so in a manner that allows experts to use their specialized knowledge to establish shared context, make spatio-temporal references, and steer the domain task with the least amount of overhead.

When it comes to novices, **the primary design goal for the local system is to make it easier to understand and act upon procedural guidance.** For guidance to be effective, novices need to receive it when their preconditions for understanding it have been met [176]. The skill gap between experts and novices means that designers should also consider communication mediums that can present the guidance experts provide in a way that bridges both the gap between the communication and task space as well as any gaps in domain knowledge. For example, a system that allows experts to use domain jargon to refer to an action while providing cues (like an animation) that contextualizes or expands on it would reduce the overall effort required for task guidance.

2.7 Conclusion

This chapter introduced a conceptual framework for remote collaborative physical tasks that outlined the collaborator needs a system should be designed for and the dimensions of support technology can provide for the communication task. It also introduced a native model for XR-mediated collaboration that highlighted the impact of the technology on the user experience and discussed the dimensions of the situated experience with virtual task environments. The next two chapters focus on the design of XR interfaces for novices and experts and evaluate them in realistic time-critical settings to uncover how the visual and spatial capabilities of XR fare in realistic scenarios with the constraints of today’s technology.

Acknowledgements

The contents of this chapter, in part, are currently being prepared for submission for publication of the material by Janet G. Johnson, Nadir Weibel, and Mark Billingham. The dissertation author is the primary investigator and author of this material.

Chapter 3

Evaluating Mixed Reality Guidance for Novices in Time-Critical Settings

Thesis Question: Do spatially localized instructions enhance novices ability to comprehend and execute procedural guidance during time-critical physical tasks when compared to screenbound instructions?

This chapter demonstrates the utility and feasibility of Mixed Reality guidance cues for novices during time-constrained emergencies through the design and evaluation of a Mixed Reality application that aids unskilled bystanders with performing CPR. A user-centered iterative design process highlights a number of key considerations for designing alongside the limitations of technology today. When evaluated against screenbound instructions in a realistic setting, spatially localized instructions improved response time, accuracy, and task performance for novices.

3.1 Introduction and Motivation

Being able to comprehend procedural guidance for a physical task and act on it in a timely manner is especially important during emergencies or time-constrained settings. However, the need to react fast coupled with the lack of knowledge required to complete the task makes this setting one that is particularly stressful for novices. Mixed Reality (MR) could make it easier for novices to understand and act upon procedural guidance by integrating information directly onto the task environment and translating information found in traditional instructions (like text and pictures) into spatially localized visual cues. The hands-free nature of MR headsets

also adds to it being a compelling medium to support procedural guidance for time-sensitive high-pressure tasks.

We don't typically process information that's spatially distributed around us, and the lack of domain knowledge with novices could make presenting information in a spatial manner confusing, especially when compared to more familiar presentations like 2D instructions on a paper or a digital screen. Additionally, despite the promise of the medium, today's Mixed Reality headsets have technological constraints that could make using it challenging. Moreover, Mixed Reality is a relatively under explored area which means that we don't currently have any standards or guidelines in place for effective designing guidance cues. Given these uncertainties, **are spatially localized instructions truly better than those on a two-dimensional screen for procedural guidance in time-constrained settings?** Do the technical and ergonomic constraints of today's MR headsets reduce the effectiveness of any in-situ instructions MR may provide? What is the easiest way to direct novice's attention to areas and objects of interest in the task environment? The research presented in this chapter aims to answer some of them by designing and evaluating novice-focused MR guidance cues for time-critical and stressful situations.

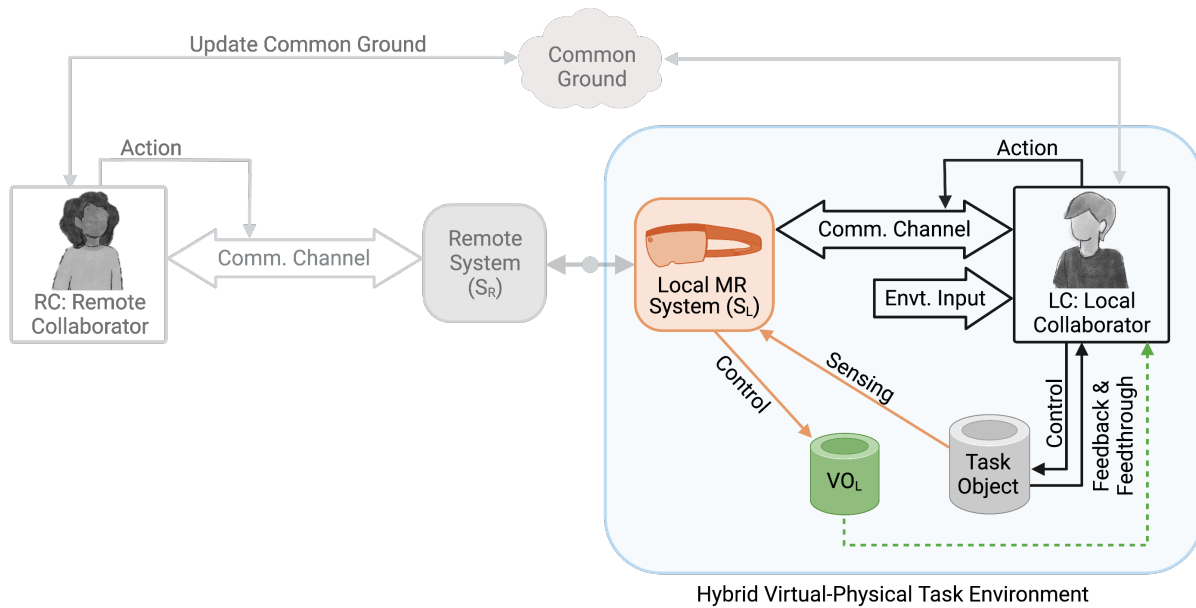


Figure 3.1. This chapter focuses on the hybrid virtual-physical task environment that includes spatially-localized virtual objects that leverage MR's understanding of real-world objects.

In order to do this, we focus on exploring the utility and feasibility of MR in a realistic setting where novices are confronted with a time-sensitive activity and need to react quickly – bystander-initiated Cardiopulmonary Resuscitation (CPR) in the event of an out-of-hospital cardiac arrest. Performing CPR usually requires trained individuals on the scene. However, given the importance of starting CPR at the earliest possible moment to increase the odds of survival for the victim, it is crucial for people around them to provide this care. Unfortunately, even when standardized step by step instructions are available, most bystanders do not attempt resuscitation due to panic or fear of failing, often at the cost of the victim’s life.

The physical nature of this task along with the need for skilled action by novices in a stressful scenario make it a setting that is well-suited to study the impact of MR guidance in time-critical settings when compared to traditional media. For these reasons, we chose to design and evaluate a real-time MR resuscitation aid that helps bystanders not trained in CPR to intervene and follow a correct procedure to resuscitate a person. We first engaged users in an iterative multi-step design process that focused on leveraging the visual and spatial affordances that MR provides. Then, to truly understand the effects of the additional affordances that MR provides over traditional media, we developed an application that incorporated the resulting MR guidance cues, *HoloCPR*, and evaluated it in a between-subjects study that compares it with traditional 2D, screen-bound instructions.

Overall, this work makes two critical contributes to further our understanding of the implications of MR as a medium to provide procedural guidance:

1. Our in-depth design process clearly outlines a number of challenges in terms of designing MR interfaces in the today’s technological environment, and proposes practical strategies to overcome these challenges for time-critical situations in MR.
2. Our evaluation demonstrates that MR can be used by novices to respond to time-critical emergencies and that its use significantly reduces reaction time and increases accuracy.

3.2 Background and Related Work

3.2.1 Evaluation of MR Guidance in Realistic Settings

Visualizing information in the space around the user has seen many applications in fields like manufacturing, assembly, entertainment, education, military, and even medicine, for both training and real-time guidance [311, 272, 346]. However, over the years there has been very little research focusing on user-based interface design approaches. Until 2004, out of all the publications in primary venues for AR, only 14% of articles were HCI-related and only 8% of those contained any user-based experiments [288]. A more recent review of publications reporting on AR assembly tasks applications found that while 39% of the articles deal with assembly guidance, only 8% focused on effectiveness evaluation and only 4% conducted usability evaluations [330].

Despite an overall lack of user testing, pioneering work since the 1990s has enabled MR or AR to advance to today's standards. In 1993, Feiner and colleagues [83] described an augmented reality system that provides explanations of, and assistance with, complex 3D tasks. In the same period, Caudell et al. [40] devised a system to aid assembly of complex aircraft parts at Boeing, digitizing vast amounts of instructions on a personal screen. A few years later, Whitaker and colleagues [336] developed a method to track and register information on real objects, allowing for virtual annotations that describe different pieces of complex systems.

Following up on these initial prototypes, more recent studies focused on comparing and evaluating different approaches targeting procedural task guidance. Sarupuri et al. [264] explored the use of AR to assist pallet racking and pick up tasks, and observed a significant increase in task performance. Uva et al. created a system, the SAR Workbench [311], intended for maintenance procedures, and Funk et al. [91] used in-situ projected instructions for manual assembly and evaluated its long term usage in the industry. Results suggest that AR was useful in helping untrained workers learn, but decreased performance for experts. Tang et al. performed a more extensive study comparing the relative effectiveness of providing instructions on paper,

a 2D laptop interface, static images on Head-Mounted Displays, and full AR on assembly operations [296]. When used for an assembly task based on Duplo blocks, they found significant reduction in error rate with the use of full AR. Their results also indicate decreased mental effort, showing that AR could complement and help reduce the user's overall cognitive load. Henderson et al. [130] focused on using AR during the psychomotor phase of a procedural task and conducted user studies comparing the AR system to one based on a standard LCD screen. They found that participants were faster and more accurate with the LCD screen, but preferred using the AR system. Recently, Blattgerste et al. [29] compared in-situ instructions on Epson Moverio (a wearable AR system based on a glass-like support), Microsoft HoloLens, a smartphone, and paper. Participants were given an assembly task and took less time when using instructions on paper, but committed less errors using HoloLens.

Despite increased attention to user studies that compare MR (and the HoloLens) with standard interfaces, it is not entirely clear whether providing MR procedural guidance can truly help enhance novice's physical task performance and their procedural adherence. It is also unclear whether these mixed results stem from the limitations of technology today. Additionally, to the best of our knowledge no work has investigated the use of MR to support novices during time-critical emergencies. Understanding the potential of this medium to help novices better understand procedural guidance, therefore, requires designing guidance cues for today's technology and studying it in a realistic setting.

3.2.2 Supporting Bystander-Initiated CPR

Each year, approximately 350,000 people in the United States experience an out-of-hospital cardiac arrest, but only 12% survive the event, making it one of the top killers in the US [214, 220]. Basic Life Support (BLS) is CPR combined with the use of an Automated External Defibrillator (AED). Performing BLS immediately after a victim collapses plays a major role in the increase of survival rates in the event of a sudden cardiac arrest. It has been shown that chances of survival fall by 7-10% for every minute without resuscitation and if not

provided within 10 minutes, the survival rate is close to zero [182]. However, it usually takes 8 minutes or longer for skilled medical personnel to arrive on the scene. Initial care in the first minutes after a sudden cardiac arrest, therefore, critically relies on the action of bystanders – or people, not part of the organized emergency response system, and often without a professional healthcare education, who happen to be near the victim.

CHALLENGES TO BYSTANDER-INITIATED CPR

Bystander-initiated CPR in the US was found to double or triple chances of survival, with survival rate increasing to 49-75% when an AED is used [218].¹ Despite evidence that bystander-CPR is crucial for both survival and health improvements, administration rates remain very low, reaching only an average of 30-40% [220]. Studies show that CPR provision increases in bystanders that have CPR training, and that carefully designed interventions can help increase CPR training [112]. For instance, a national program in Denmark that included widespread CPR training, dissemination of CPR education kits, and improvement in dispatch-assisted bystander-CPR, resulted in a significant increase in post-CPR survival over 10 years (1-year survival increased from 2.9% in 2001 to 10.2% in 2010) [342]. However, it has been observed that bystanders hesitate to perform CPR even when trained, and a significant increase in providing CPR is only observed in bystanders with an active CPR training, one that has occurred within the past five years [112].

Unfortunately, the complexity of resuscitation guidelines and the cost of CPR training decreases the likelihood that the general public is able to maintain an active CPR training. More importantly, it has also been shown that panic may influence readiness to act in an emergency situation irrespective of prior CPR training [112]. The most commonly cited barriers are fear of harming the victim or performing CPR improperly. Given the importance of bystander-CPR and its potential to save lives, having additional aids on site is an important step towards improving CPR delivery and providing real-time resuscitation aids could increase the confidence of the

¹Note that this varies markedly across US counties

bystanders, overcoming the barriers created by panic and the ability to initiate CPR.

EXISTING CPR AIDS FOR BYSTANDERS

The typical response to improve guidance for bystanders during CPR is to increase availability of checklists and alike to guide the provisioning of CPR step-by-step. These checklists and step-by-step guides are sometimes included in first-aid kits, or are made available in public spaces (e.g. installed on the walls where AED devices are also located). While increased availability is good, these guides tend to be text-based, and following them is hard, especially in emergency situations. The use of mobile phones has been strongly encouraged for bystanders responding to a cardiovascular emergency [173]; while this makes instructions more readily accessible than checklists, bystanders still have to read and interpret the instructions, potentially increasing their cognitive load in an already stressful environment. Traditional media like paper and two-dimensional screens also require the user to switch focus between the medium and the victim, reducing the promptness with which they can respond. There is therefore a need for real-time contextual assistance systems to guide high-quality resuscitation.

When it comes to Mixed Reality, the rare applications focusing on CPR tend to be designed for training purposes and also tend to lack user testing. Park et al. [232] created a system for CPR training that provides interactive feedback using projections. Like many other CPR training programs, this AR system is dependent on sensors embedded within a mannequin that cannot be directly translated into a system to be used in real events. HeartiSense [180] is also an AR system for CPR training; the authors conducted user tests using the system to validate its accuracy. However, their participants were expert emergency responders with prior CPR training, and as such the results cannot be translated to similar interfaces for novices.

3.3 Designing Mixed Reality Guidance Cues

Successfully guiding bystanders, especially if untrained in CPR, through the time-critical and stressful task of resuscitation, requires an interface that allows for the intuitive comprehension

of the information presented to them. To understand how novices would interpret and piece together the information that their augmented environment would now offer, we employed an iterative user-centered design approach [223]. This approach proved crucial in outlining the design elements of our prototype MR application. In this section we detail our design process and the elements that emerged from it. We also introduce the resulting prototype, HoloCPR.

3.3.1 The Iterative Design Process

The goal of the first phase of our research was to design a MR prototype that would effectively exploit the visual and spatial representation of instructions enabled by MR, as well as the cues that would prompt the transitions between instructions (Fig.3.2, Left). Our iterative design methodology continuously involved novices who contributed to the implementation of intermediate prototypes. To implement our MR prototype we used Microsoft HoloLens (Fig. 3.2, Right), a standalone head-mounted display that renders virtual objects (which Microsoft calls holograms)² into the user's environment. We chose this device because of its self-contained nature, the Unity-based SDK, and the availability of many development libraries which supports

²<http://www.microsoft.com/hololens/developers>



Figure 3.2. Envisioned MR prototype for expert-guided CPR. Left: chest compression instructions, and correct positioning of the hands displayed on top of a resuscitation mannequin. Right: Microsoft HoloLens head-mounted display; note the clicker taped to the head-mounted display as it was used for our experiment.

rapid prototyping, allowing us to focus on the user experience and design of the guidance cues.

We recruited 16 participants (7 females), with an average age of 23 years. Participants' background were in Computer Science (4), Cognitive Science (4), Public Health (1), Biology (1), and Other/Undeclared (6). All of them had no or limited experience with MR or HoloLens, and 5 of them had been trained in CPR. We engaged the participants in a four-phase iterative design process based on (1) user testing, (2) participant interviews, (3) rapid prototyping, and (4) prototype development.

1. **User Testing:** In each iteration participants used the latest prototype on the HoloLens to try and resuscitate a simulation mannequin. They were instructed to wear the HoloLens and follow the steps indicated by the application. Each session lasted between 10 and 15 minutes.
2. **Participants' Interviews:** After testing the prototype each participant was interviewed to shed light on the obstacles they faced. We employed semi-structured interviews [56] to balance eliciting specific feedback on our prototype, and encouraging the participant's own view on MR. Topics and ideas that emerged during the interview were analyzed by our team and discussed prior to proceeding to prototyping.³
3. **Rapid Prototyping:** A rapid prototyping session supported by sketches, low-fidelity paper prototypes, and a whiteboard representation of the different entities of the envisioned system (Fig.3.3) helped design the next iteration based on the observed user challenges.
4. **Prototype Development:** The resulting prototype was finally implemented using Unity 3D and deployed onto the HoloLens to be tested with the new iteration. Development continued until we reached saturation and no new or missing design element emerged.

³When feedback from one participant did not result in enough material to warrant a prototyping phase, feedback from up to 3 participants was analyzed before moving onto the prototype phase.

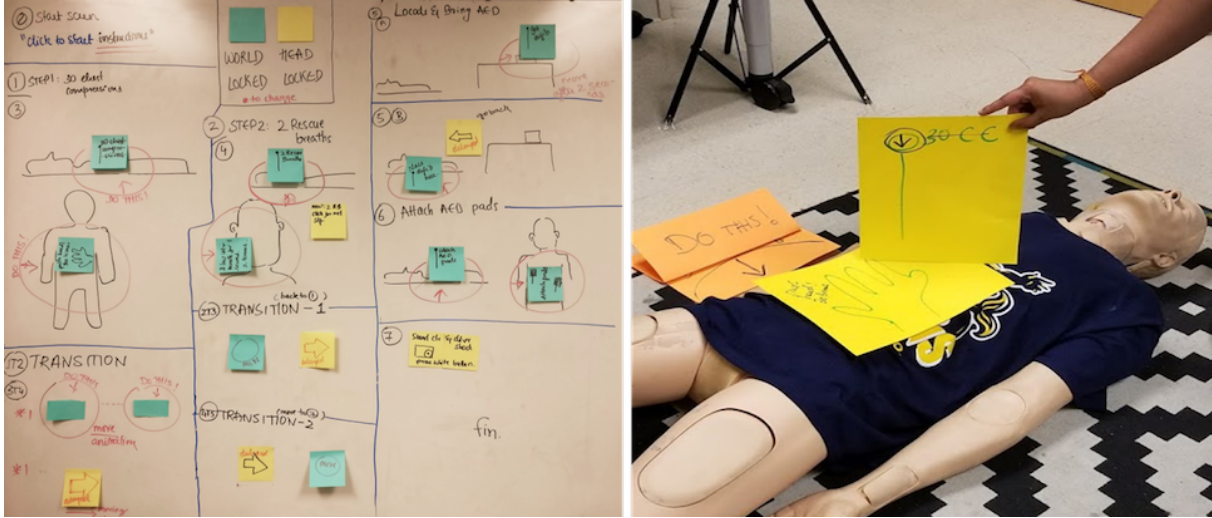


Figure 3.3. Prototyping with HoloLens and our HoloCPR mannequin. Left: Our iterative prototyping environment on a whiteboard. Right: Physical projection of the 3D representation using paper prototyping.

3.3.2 Design Challenges and Proposed Solutions

We hypothesize that mixed reality interfaces will overcome some of the problems of providing real-time guidance on traditional media. Through our design phase we were able to uncover a number of challenges that we did not anticipate, but resulted in key design elements that we eventually integrated in our application, including a particular attention to current technical limitations of MR systems, as well as users' unfamiliarity with the medium. In the remainder of this section, we describe the design elements that resulted from our iterative prototypes and discuss how they contribute to the more general development of MR interfaces for time-critical situations.

LACK OF FAMILIAR DESIGN PATTERNS FOR GUIDANCE

Over the years two-dimensional interfaces have benefited from design patterns that emerged from common usability research –e.g. users expect interfaces to follow a sequential direction (usually top-down, left-right) to scan for and read instructions [309]. On the other hand, MR does not inherently offer users a way to intuitively scan the interface for information.

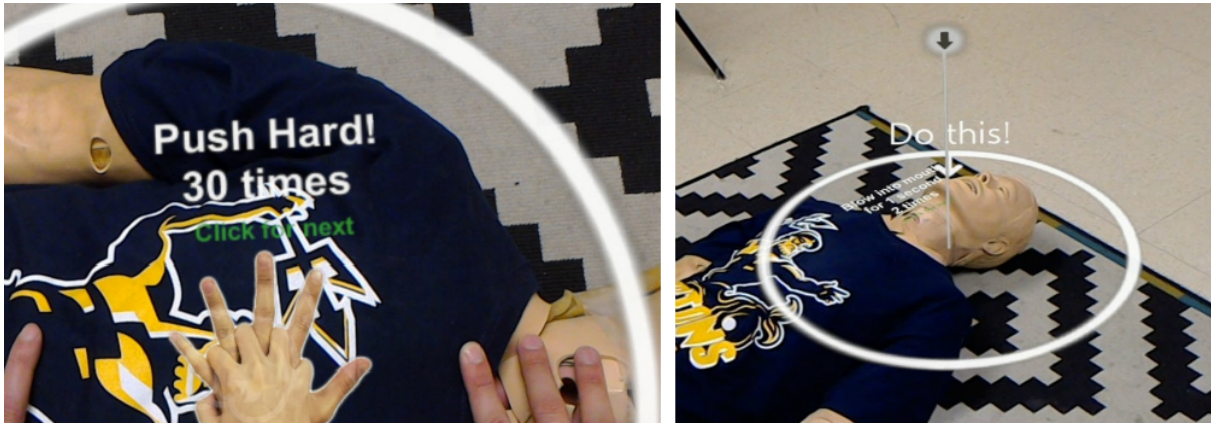


Figure 3.4. Instructions for cardiopulmonary resuscitations, as seen by the wearer. Visible are also the Circle of Attention and the Vertical Pole (right) in the middle of it.

Besides some basic guidelines,⁴ designers are left with a mostly unknown interaction space. This lack of familiarity with design patterns was particularly challenging in our scenario where users had to interact with and consume information rapidly, in a stressful situation, and with information that did not necessarily appear in the same position all the time. In the case of time-critical emergency situations, MR instructions are spread in space (rather than on a specific 2D interface), and their position depends on the user's relative location with respect to the victim and the environment.

In order to support this situation, we needed to explicitly guide user attention to the task at hand and help users establish patterns in the interface so that they could easily find the information they needed without prior training. To guide users in the MR space, we introduced a composite visual element that we called the *Circle of Attention*. The Circle of Attention is composed of a white *circle* with an *arrow* on its perimeter, textual prompts, and any number of additional images. Figure 3.4 shows its representation applied to rescue breaths.

In addition to the circle of attention, often it was important to guide users away from their current gaze position (e.g. while focusing and looking at the victim's chest), in case they needed to attend to a different instruction. We used a *Red Arrow* that followed the users' gaze to direct

⁴<http://developer.microsoft.com/windows/mixed-reality/design>

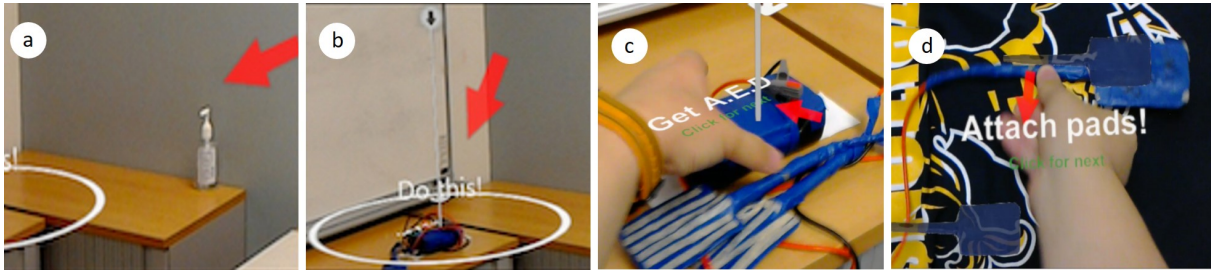


Figure 3.5. The “Red Arrow” guides users by indicating where the important things are located and where they should act. From left to right, the prototype shows how to locate first the AED (a, b) and then how to grab it and position the AED’s pad correctly on the victim’s chest (c, d).

them towards the instruction they needed to attend to. Figure 3.5 illustrates how the red arrow helps users find the AED pads elsewhere in the room and apply them correctly to the victim.

NARROW FIELD OF VIEW

A narrow Field of View (FoV) on the HoloLens causes holographic content to frequently be cut off. [P8] noted how *“the frame cut off weirdly at some parts, which made it difficult to know whether or not there is something supposed to be projected into a certain surface”* and [P4] pointed out how *“it was a little hard to see all of the instructions, so I had to look around to actually see them all.”*. Because of the limited FoV, digital content could be invisible even if it is in the user’s natural FoV (see example in Fig. 3.6).

Although this is mostly a limitation of the current technology, we wanted to mitigate the limited FoV and compensate it by encouraging users to use less eye movement and more head movements when they needed to change their gaze attention. To accomplish this and allow stationary users to be aware of information outside of their FoV, we integrated a *Vertical Pole* at the center of the Circle of Attention (see Fig. 3.4) that stretches upwards vertically. When the area of instructions is in the user’s natural FoV, but vertically outside the HoloLen’s FoV, the vertical pole helps users direct their gaze to the center of the circle, towards the needed information.

In addition to the Vertical Pole, we designed the *arrow* on the perimeter of the Circle of

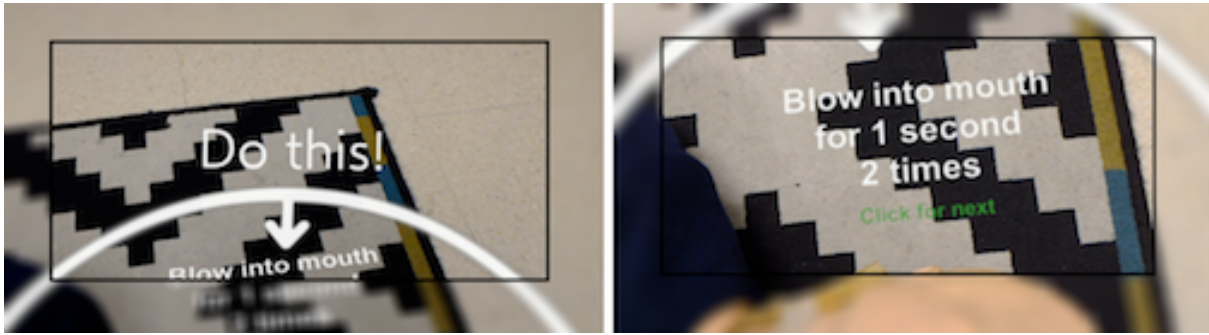


Figure 3.6. In this image, we can see how the “Circle-of-attention” points the user to where the important information is (users move their gaze from what they would see in the left picture to the right one). The figure also shows how the holograms in the area outside the black rectangle are not visible for the user, given HoloLens’ narrow field-of-view.

Attention to move around and always be diagonally opposite to the user. It always points to the center of the circle, prompting them to look towards the center to find more instructions.

CLOSE RANGES AND ERRATIC MOVEMENT

Current technical limitations inherent to many MR devices cause visual aids to disappear when users get too close to objects or people in front of them. Extensive movement also causes those visual aids to shift and makes focusing on instructions more difficult. [P15] mentioned that “*moving around with the HoloLens made it a little harder to read the instructions*”. Since performing CPR means that users would both be very close to the victim and move erratically, we needed to be able to help users hold their attention and minimize discomfort during extensive rapid movements.

The circle of attention described above also served to help users retain their focus on the instructions, even when in close range to the digital content and when moving extensively. In addition, the constant presence of the circle of attention helped mitigate the effect of *disappearing* holograms.⁵ The appearance of the circle of attention and the arrow on its perimeter, as soon as users move far enough from the objects or people in front of them, helps users to immediately *re-capture context* and proceed with their intended interaction.

⁵E.g., when users lean in to breathe into the victim’s mouth, the head moves very close to the instructions, and digital content disappears.

Finally, interaction during erratic movements (e.g. during chest compressions) proved to be problematic. While textual information always looks jittery, we found that adding enough thickness to the border of our circle of attention helped users to keep the focus and their attention in the correct place, and move back to that particular position if their body or gaze moved during the erratic movements.

MANAGING HIGH USER EXPECTATIONS FOR INTELLIGENT SYSTEMS

One thing that was extremely clear from our interactions with all the participants was that they had very high (and often wrong) expectations from the MR experience. Expectations were especially focused on the *smartness* of the device, and how it would automatically assess the situation. In particular the majority of our participants expected the device to automatically recognize their actions and guide them accordingly.

While somewhat natural for an emerging and mostly unknown technology, we were surprised by how this expectation recurred in many of our design sessions. In order to mitigate these expectations, we added a *clicker* taped on the side of the head-mounted display (see Fig. 3.2, Right) and asked participants to explicitly click to move to the next stage of the CPR. However, their expectation was so severe that even after explicitly telling participants that their actions were not recognized and to use the clicker, we received feedback like “*I waited for it to move by itself*” [P3] and “*I did expect a little for the headset to tell me what to do, even though I was warned not to*” [P10]. The only way to avoid these misunderstanding, was to add a textual instruction “Click for next” in green in front of the users to remind/prompt them to do so. We selected green because while clearly visible, its appearance is less salient than the white texts and images used for task-related instructions.

3.3.3 The Resulting Prototype: HoloCPR

Our iterative design process yielded key observations in terms of design, interactions, and visual components. We integrated these observations from partial prototypes into a complete

MR application to support the real-time delivery of CPR. HoloCPR was built as a HoloLens application using Unity 3D⁶ and exploited the prototypes developed in our design phase. In this section we describe the overall goal of the application, as well as its envisioned use. In the next section we then describe how we evaluated its effectiveness.

HoloCPR has two main components, the first one is focused around the physical resuscitation, and the second one assists with the placement of an AED. In the event of a cardiac arrest, users place the HoloLens on their head and activate the HoloCPR application. When users look at the victim, the system indicates the correct placement of the two hands in the center of the victim's chest and instructs the user to perform 30 chest compressions (see Fig 3.2, Left). When they are done, the system guides the users to the correct placement of their hands around the mouth of the victim, and instructs them to blow into the victim's mouth 2 times (see Fig. 3.4). If an AED is available nearby, HoloCPR instructs the users to stop CPR and guides them to the location of the AED and back to the victim using a red arrow in the center of their visual field (see Fig. 3.5). It then instructs them to place the AED pads in the correct position on the victim's body to provide defibrillation.

Participants' feedback and expectations outlined how automatic recognition of chest compressions (including rhythm and force applied), mouth positioning (including blow frequency and length), as well as automatic AED localization are key elements needed for an effective deployment of HoloCPR in the real world. However, this chapter focuses on the current implementation of HoloCPR and investigates visual and spatial cues in Mixed Reality, and advanced recognition features are not currently in the scope. As a result current interactions between different steps in the resuscitation are performed through direct interaction by the user (e.g. "click to go to the next step"), using a HoloLens clicker taped on the side of the device (see Fig. 3.2, Right).

Based on the needs we observed during our design phase, we decided to divide the standard instruction in each step into three groups: *what*, *where* and *how*. We then used the

⁶<https://unity3d.com>

where to provide positional context for the instructions, and used text and images to convey the *what* and *how*. For example, to help users perform chest compressions (see Fig. 3.7), they needed to know what to do (chest compressions, 30 times), how to do it (push hard, use this hand form) and where to do it (in the middle of the victim's chest). HoloCPR represents this particular interaction using two sets of text and an image depicting the hand form as well as the position (see Fig. 3.2, Left). A similar approach was taken for all other HoloCPR instructions.

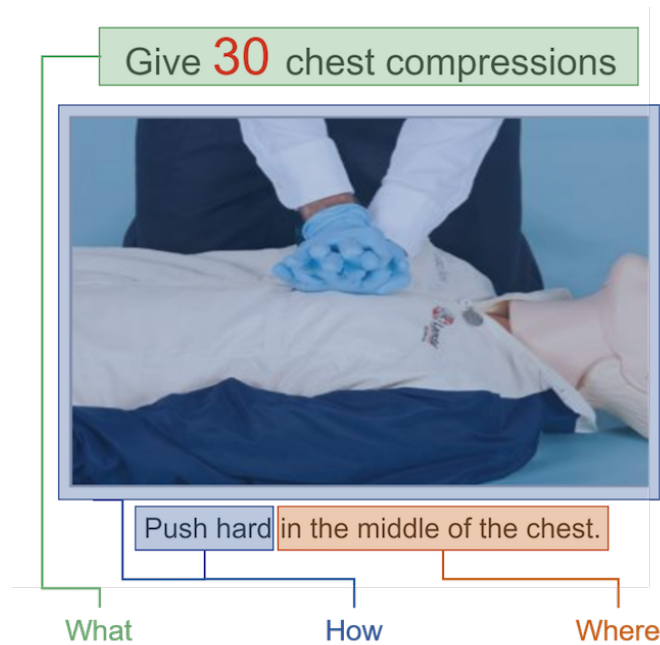


Figure 3.7. Dividing instructions into “What”, “How” and “Where”

HOLOCPR IN ACTION

In order to better illustrate how HoloCPR works, in the remainder of this section we focus on a scenario, based on Avi, a 28 year old graduate student who uses HoloCPR to resuscitate a victim. The goal is to direct Avi towards the instructional point of interest where she would find textual prompts and images that guide her through her actions. The scenario outlines both the use of HoloCPR and how the visual aids that resulted from our user-centered design helps her to proceed through the CPR experience. Figure 3.8 shows a schematic representation of the central elements of the scenario we illustrate below.

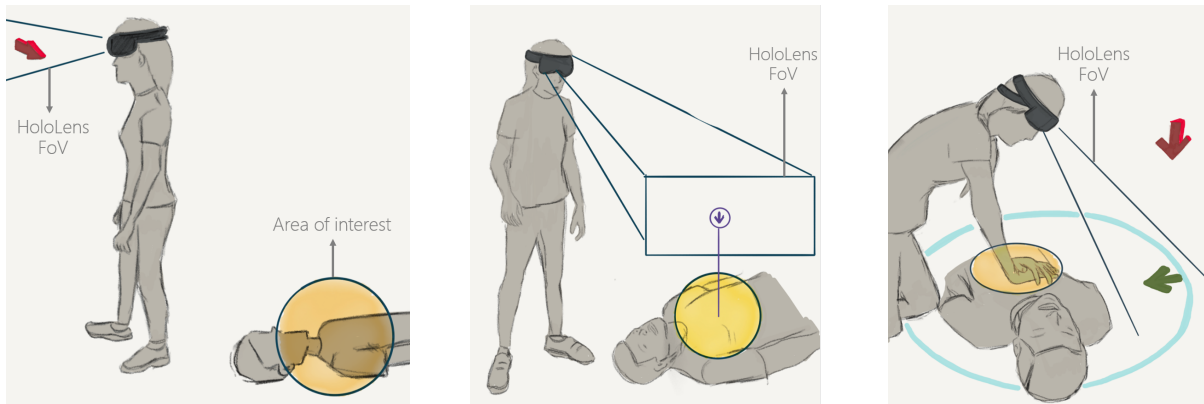


Figure 3.8. HoloCPR Scenario. Left: User faces away from the area of interest; Center: Area of interest in the user's natural FoV but not in HoloLens' vertical FoV; Right: Area of interest at close range and in the user's natural FoV but not in HoloLens' FoV

Avi enters her office and sees a person on the floor who is not breathing. After rapidly assessing the situation she decides to perform CPR, but she does not remember all the steps. She decides therefore to make use of the HoloLens and the new HoloCPR application from the first-aid kit that has just been deployed in the building. When Avi initiates resuscitation guidance, the victim's chest is not in her natural field of view, but she immediately notices a red arrow that points towards the upper body of the victim (Fig 3.8, Left). She also registers a white circle, big enough for her to stand in that moves away from her almost as soon as she sees it. The slight delay in movement serves as the first cue in understanding its purpose. Avi instinctively looks towards the direction it moved to and notices that the red arrow is also pointing in the same direction, waiting for her to follow it. Now she can see that she needs to rapidly approach the victim's chest to start resuscitation. Just as she feels like there aren't any instructions (because it is outside the HoloLen's FoV), she notices a vertical pole in the middle of the white circle in front of her, and she immediately traces its length vertically (Fig 3.8, Center). Avi sees the white circle again, now with additional information that instructs her to do 30 chest compressions by placing her hands on a digital representation of two hands projected onto the victim's chest.

She immediately kneels down beside the victim, places her hands on the victim's chest, making sure to interlock her fingers as shown (see Fig. 3.2, Left) and begins chest compressions. As she performs chest compressions, Avi can no longer see the instructions (Fig 3.8, Right). She moves her eyes to look down at where she just saw the digital representation of a hand, but she doesn't see it anymore. Avi however remembers registering an arrow on the perimeter of the white circle that seems to move around and always face her. It also seems to always point to the center of the circle, encouraging her to look towards it. When she does, she finds the instructions again. Although the rapid movement due to the chest

compressions makes the HoloLens feel heavier on her head, she is able to focus on the only thing that isn't jittery, the white circle. She continues counting through her compressions. When she is done with 30, she presses the clicker to move to the next step...

3.4 Evaluating MR Guidance against Traditional Media

To assess the effectiveness of HoloCPR we conducted a between-subjects experiment in a realistic lab setting. Participants were asked to perform CPR on a simulation mannequin by following the steps shown by HoloCPR. We measured the time it took for them to go through the simulation, as well as the accuracy with which they performed the procedure. We compared HoloCPR with a control condition that simulated the traditional CPR experience using standard 2D instructions for guidance on a separate support. The instructional medium used for guidance was the only independent variable.

We posit that the additional visual and spatial affordances of MR could provide multiple benefits and lead to a decrease in the time it takes to translate instructions to action, especially for non-skilled users, enhancing the promptness with which bystanders can respond to the victim. In the event of a sudden cardiac arrest, spatially localized instructions can offload the burden of interpreting location specific instructions to the system itself (for example, correct placement of hands for CPR or retrieving and attaching AED on the victim's chest). More specifically, we hypothesise that:

- H1** *In comparison to traditional media, MR will reduce the response time needed to begin resuscitation in the event of a sudden cardiac arrest.*
- H2** *In comparison to traditional media, MR will reduce the time it takes to read, interpret, and act upon an instruction.*
- H3** *In comparison to traditional media, MR will improve procedural adherence.*

3.4.1 Evaluation Procedure

PROTOTYPE USED

We based HoloCPR’s resuscitation guidance on published and recognized guidelines taken from the Adult First CPR/AED Reference from the American Red Cross.⁷ For the sake of our evaluation we chose to prompt participants to perform only 2 rounds of chest compressions and rescue breaths before they moved on to 1 round of simulated defibrillation. This amounted to a total of 7 instructions listed in Fig. 7.5 (left).

To allow us to measure reaction time and the time it took for participants to transition between instructions, the interface presented one instruction at a time and was designed to include a “startup page” that prompted users to click to initiate resuscitation guidance. In addition, to simulate traditional CPR guidance and measure time-to-task, in our *control condition* we used a tablet interface that showed a 2D representation of the CPR/AED guidance on an Apple iPad Pro. We carefully designed the tablet interface for the control condition to mimic as much as possible a paper-based checklist. The only difference for participants was the *start* button to begin resuscitation guidance and the *next* button to go to the next page. Instructions consisted of title, two lines of text and a picture, all from the CPR guidelines. At the bottom was the title of the next step along with the *next* button.

PARTICIPANTS

We recruited a total of 42 participants (12 female), graduate and undergraduate students at the University of California San Diego, across multiple departments (i.e.: 23 from Computer Science and Engineering, 6 from Cognitive Science, 4 from Biology/Medical School, 2 from Electrical and Computer Engineering, 2 from Mechanical and Structural Engineering, 2 from Bioinformatics, 1 from Nanoengineering, 1 from Chemistry, 1 from Economics), with ages spanning from 18 to 27 years (average 22.7 years). None of the participants had been exposed to HoloCPR and only 10 had been trained in CPR. All but 6 participants had never used a HoloLens

⁷<http://www.redcross.org/cpr>

or any mixed reality device before. All participants were familiar with the general use of a tablet and touch screens. Participant assignment to the two conditions balanced gender, background, and CPR experience. All participants were unaware of the task and the instructional media they would use.

EXPERIMENT DESIGN

Setup — The experiment was conducted in a room equipped with a simulation mannequin, a mock heart rate monitor showing on a large screen, a mock AED device, and a computer where participants would fill out the quick pre and post-task questionnaires (see below). The mock heart rate monitor was remotely controlled and could be set to flatline and trigger an alarm to indicate a “crash” (cardiac arrest). We instructed participants to start resuscitation of the mannequin (either with HoloCPR or with the 2D checklist, depending on the assigned condition) when the alarm would start. A video camera recorded the activity in the room for further analysis.

Introduction to the HoloLens — To introduce participants to the HoloLens and acquire basic familiarity with its weight, the concept of holograms, and the clicker, we asked them to play RoboRaid⁸ (a mixed reality 3D first-person shooting game) for 12-15 minutes. This also served to familiarize the users with the position of the clicker. Participants were kept unaware of any HoloCPR specific interface details. To account for the possible fatigue introduced by this procedure, and ensure that any variables introduced by playing the game remained constant across conditions, all participants in both conditions were required to complete it.

Experimental Procedure — Once the participants completed the necessary training, they were taken into the experiment room where they were asked to fill out a pre-study questionnaire asking for basic demographic information. For both conditions, a researcher took up the role of a caretaker and introduced the simulation mannequin (henceforth called the victim). Participants were told the victim was recently moved back home from an ICU and that they would be alone with him for a short period of time while the researcher stepped out to “run an errand”. They

⁸<https://www.microsoft.com/en-us/hololens/apps/roboraid>

1. Give 30 chest compressions
2. Perform two rescue breaths
3. Give 30 chest compressions
4. Perform two rescue breaths
5. Locate the AED
6. Attach the AED pads to the mannequin's chest
7. Defibrillate the mannequin



Figure 3.9. Left: CPR Instructions for chest compressions and rescue breaths. Right: A researcher and participant (C) in the experiment room with a mock AED device (A), a simulation mannequin (B), a mock heart rate monitor (D).

were alerted of the heart rate monitor and the instructional medium available to them (HoloLens or iPad) and were instructed to attend to the victim should anything go wrong. No additional instructions were given.

Participants in both conditions were ready to initiate resuscitation aid before the researcher left the room. HoloCPR participants kept the HoloLens on and saw the prompt "Click to start". Control participants were given the iPad with the start page already opened and ready to go. They were, however, instructed to not initiate resuscitation aid unless they needed to (i.e. when the victim was in need of their help). Once they were left alone in the room the heart rate monitor was set to indicate a cardiac arrest at a random time within the first two minutes. The alarms from the heart rate monitor were kept going until participants reached the very last step of the resuscitation to help build and sustain a sense of urgency.

After all steps were accomplished we asked participants from both groups to fill out a post-study questionnaire about their experience. Participants reported their subjective opinion on the importance of the aids used with respect to performing CPR (step-by-step guide using a tablet, or HoloCPR using MR). We also inquired about difficulties or confusion they experienced during

resuscitation, and performed a semi-structured interview that elicited more in-depth feedback on the interface, as well as specific feedback on their performance.

DATA ANALYSIS

All the sessions were recorded using an external camera and basic logs were collected on the HoloLens and the iPad. Videos were coded using the ChronoViz analysis tool [89]. We then used a Python script to extract and analyze the data from the videos. Of the 42 participants we excluded 9 from quantitative data analysis (4 in the HoloCPR, 3 females, and 5 in the control condition, 1 female) who did not complete the whole experiment due to technical issues (5) or because they misunderstood the task (4). In the next section we report and discuss the results of our experiment and the answers collected through the post-study interview and questionnaire.

3.4.2 Results

Data extracted from the videos of the experiment was coded for response time, performance, and procedural adherence, with the goal to compare the control (i.e.: paper-like step-by-step on tablet) and treatment (i.e.: mixed-reality with HoloCPR). To assess the statistical significance of our results we used a 2-sample t-test ($\alpha = 0.05$). In this section we report on the outcomes of our analysis and discuss results with respect to our original hypotheses. Figure 3.10 summarizes our results.

RESPONSE TIME

HoloCPR users react faster – We compared the time it took for participants to attend to the victim, more specifically, the time it took them to start resuscitation once a cardiac arrest on the mannequin was triggered. HoloCPR was significantly faster than the control condition ($p < 0.01$). On average, HoloCPR allowed participants to respond to the victim in 7.04 seconds while tablet users took 11.96 seconds. To better understand what contributed to this difference, we broke this response time into two phases.

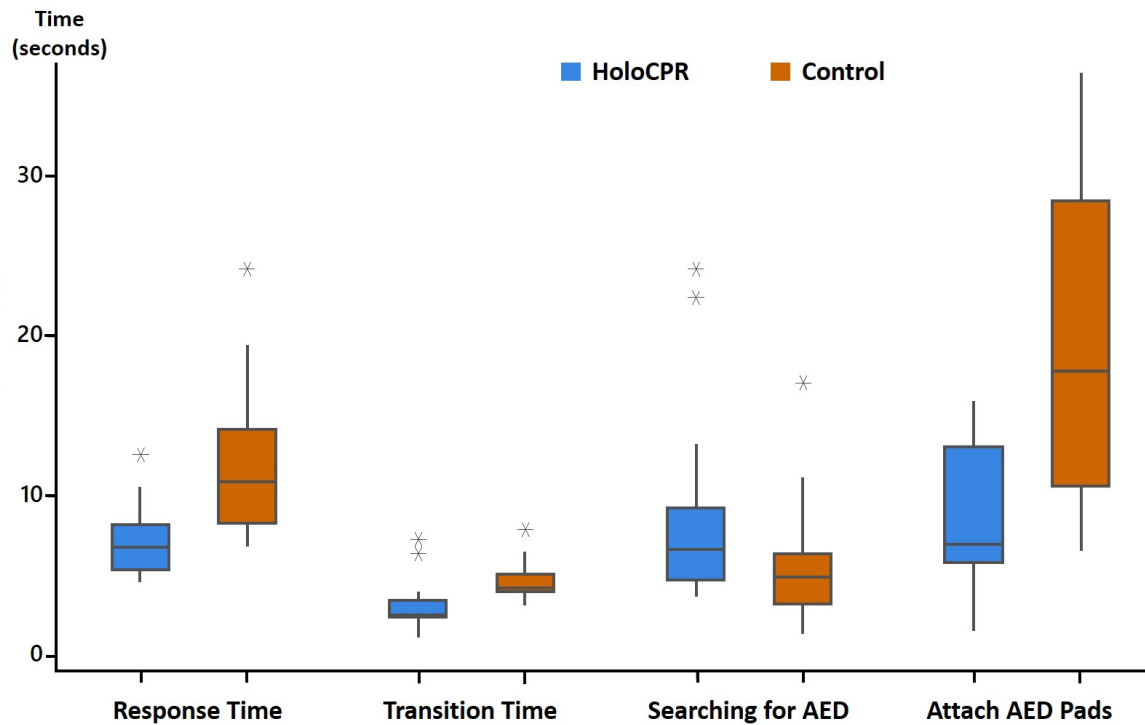


Figure 3.10. Response time, transition time, time spent searching for the AED, and time taken to attach AED pads across both conditions.

1. **Guidance Initiation Time** was defined as the time it took for participants to initiate resuscitation guidance on their respective devices (“click *start*”) once a cardiac arrest was triggered. HoloCPR users clicked *start* within 2.96 seconds on average while tablet users used 3.7 seconds on average. While HoloCPR did show a trend towards shorter initiation times, our comparisons showed no statistical significance between the two conditions ($p = 0.175$).
2. **Initial Reaction Time** was defined as the time it took for participants to start CPR once they initiated resuscitation guidance. This time the difference was statistically significant ($p < 0.01$) with HoloCPR users taking an average of 4.08 seconds vs. control users taking 7.25 seconds.

PERFORMANCE AND TRANSITION TIME

HoloCPR does not impact time spent searching for AED – We no significant difference in the time participants spent searching for the AED ($p = 0.12$), with control participants taking 5.77 seconds on average and HoloCPR users taking 8.61 seconds on average.

HoloCPR users are faster while attaching AED pads – On the other side, HoloCPR users were faster ($p < 0.01$) in attaching the AED pads on the victim's chest. On average HoloCPR resulted two times faster than the control condition (8.98 seconds vs. 19.41 seconds on average).

HoloCPR users transition faster to next steps – Transition time was measured as the time it took participants from the end of one step to the beginning of another. This generally included clicking *next* as well as interpreting the instructions before moving on to the next step. HoloCPR users were significantly faster ($p < 0.01$) than the control condition, and took an average of 3.17 seconds to transition to the next step (vs. 4.65 seconds).

PROCEDURAL ADHERENCE

HoloCPR helps stay on track with the procedure – While all users were instructed to follow the 7 steps indicated in our simplified procedure, 3 of the control participants performed additional rounds of chest compressions or rescue breaths, with a total of 12, 10, and 8 steps instead of 7. For HoloCPR only 1 participant deviated, performing 2 additional steps.

Hand position is often correct with HoloCPR – Both the control condition (using a picture) and HoloCPR (using a representation of digital hands) instructed participants to perform chest compressions with one hand on top of another and their fingers intertwined. However, we observed that only 12.5% of the control participant started CPR with the correct form. For HoloCPR this number tripled with 41% of the HoloCPR users performing the chest compressions correctly. Out of the participants that started with the wrong form, none of the participants in the control condition corrected their form whereas 30% of the HoloCPR users did.

HoloCPR does not help with correct AED Pad Placement – We observed that 18.75% of the control participants and 17.65% of the HoloCPR users placed the AED pads in the wrong

position. A correct position was one where users places one pad on the upper right side of the victim's chest and the other pad on the left side of the chest.

SUBJECTIVE FEEDBACK

Besides the quantitative measures outlined above, we also report on subjective feedback from our users. The vast majority of the HoloLens participants (82%) reported that the technological aid was important (score of '5' or '4' on a 5-point Likert scale), while only half of the control participants (56%) reported that the technology was important (score of '5' or '4'). When asked about ease of use of the interface (with '1' being easy to use and '5' being hard/confusing), HoloLens participants were mostly neutral (37.5% selected '3', 25% selected '1', easy, and 25% selected '4', difficult). The majority of the control participants, in contrast, reported the interface to be easy to use (54% selected '2').

Participants using the tablet interface commented on its simple and intuitive use – *“All of the buttons were big so it was easy to know what to do on the tablet”* [P58], *“The directions were simple and easy to follow, the pictures helped a lot!”*[P52]. However, non-CPR trained tablet users recalled feeling uncertain about how to put in practice the steps indicated: [P20], for instance, reflected *“I felt like I was losing time because I had to keep reading the tablet for the next step. The instructions were easy to follow though because of the pictures and diagrams but I wasn't sure if I was doing it correctly”*. A CPR-trained tablet user [P25] also mentioned *“I felt I wanted more instructions regarding how to perform chest compressions or the rescue breaths – technique is important in these maneuvers”*.

On the other hand, users expressed positive feedback for HoloCPR's interface – [P48] commented *“Showing the positioning of where to press for CPR and where to position the AED was helpful because I didn't have to spend extra time figuring that out myself”* – non-CPR trained users noted how *“the system told me what to do - otherwise, I would have no idea”*[P32] and how *“the spatial awareness was key. Telling me *where* and *how* I had to hold my hands and telling me where the CPR device was, made it super intuitive”*[P41]. HoloCPR Participants

were also critical, for instance [P38] commented on how “*to put on the HoloLens in such an emergency situation could take some time*”, and [P46] noted that “*when the arrows would try to guide me to something, it would’ve been nice to know ahead of time what I was being guided to*”.

3.4.3 Discussion

Overall we have been able to observe a trend towards better response times, accuracy, and performance when HoloCPR was used for resuscitation guidance. In particular this is true for all of our hypotheses (H1: reduced reaction time; H2: reduced interpretation and faster action; H3: improved accuracy). Our analysis suggests that mixed reality guidance reduces the reaction time needed to respond to a cardiac arrest (faster reaction and reduced initiation time), and confirms H1. Performance and transition time are mostly reduced as well, so in general, hypothesis H2 is also supported. Although we saw a tendency to spend more time looking for AED, our results suggest that mixed reality reduces the time it takes to read, interpret, and act upon an instruction. Finally, we can say that H3 is partially supported as well: HoloCPR helps positioning hands and improves accuracy of the steps to follow, but does not help with AED more than the traditional guidance does. In general, however, HoloCPR did not decrease accuracy with respect to the control condition.

We also observed a higher degree of procedural adherence from participants using mixed reality guidance. However, our observations could be the result of attention bias – a phenomenon that occurs when users focus on highlighted areas at the cost of other areas [349]. More extensive studies are needed to understand exactly how much procedural adherence is due to this phenomenon and discuss how the extreme reliability on cues in augmented reality displays influence its use in emergency time-critical scenarios. In our study we observed interesting episodes of extreme reliance on the mixed reality interface, especially during our iterative design session. For example, a participant performed chest compressions when directed, even though the indicated location was not on the victim, but under a table, due to a calibration issue. Others followed and waited for the *red arrow* even when the AED was in their field of view. Our

experiences confirm existing literature [349] which points to a need to design mixed reality interfaces to sustain users attentional focus, without encouraging unwarranted reliability on the technology.

All in all, the results of our study support our initial intuition that the additional visual and spatial affordances of mixed reality decrease the time it takes to translate procedural guidance into action, especially in a time-constrained environment. In overlaying information directly onto the area of interest, HoloCPR reduces the need to switch focus between the instructional medium and the task environment (the victim, in this case). Spatially distributing information (especially in a non-textual manner, like showing a digital representation of hands with interlocked fingers on the victim's chest) also reduces burden of reading and interpreting instructions, instilling a higher sense of confidence in users.

3.5 Conclusion

This chapter designed and evaluated an MR application that provides novices with the procedural guidance required to perform CPR during an emergency to study the impact of MR-mediated guidance for novices during time-critical emergencies. Our results indicate that MR's visual and spatial capabilities makes it easier for novices to understand procedural guidance in time-critical and stressful settings when compared to guidance presented on traditional media. In the next chapter, we design and evaluate an immersive interface for remote experts that aims to reduce the effort it takes for them to provide MR-mediated guidance to novices.

Acknowledgements

This chapter, in part, includes portions of material as it appears in *HoloCPR: Designing and Evaluating a Mixed Reality Interface for Time-Critical Emergencies* by Janet G. Johnson, Danilo Gasques Rodrigues, Madhuri Gubbala, and Nadir Weibel in the Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare, ACM, 2018. The dissertation author was the primary investigator and author of this material.

Chapter 4

Evaluating An Immersive Interface for Experts in Time-Critical Settings

Thesis Question: How do the visual and spatial capabilities of Extended Reality meet remote expert requirements during collaborative physical tasks and impact their ability to provide effective MR-mediated guidance in time-critical settings?

This chapter provides an understanding of how the visual and spatial capabilities of XR impacts the ease with which remote experts can provide MR-mediated procedural guidance for novices performing trauma surgery. A participatory design study highlights their domain-specific needs and the technological constraints that together define the resulting XR-mediated telementoring system. A feasibility study that evaluated the use of this system for combat casualty care procedures confirmed that XR-mediated collaboration can effectively facilitate the successful completion of a wide variety of realistic procedures. However, a qualitative analysis of collaborator behaviors revealed that despite the information and interaction richness XR offers, referencing real-world artifacts during physical task guidance is still difficult from a distance.

4.1 Introduction and Motivation

Presenting rich and contextual information that help novices understand and act on procedural guidance in time-critical settings when compared to traditional media confirms MR's suitability for physical tasks. Virtual Reality (VR), on the other hand, has the potential to make providing MR-mediated guidance feel more intuitive – especially because it would allow us to immerse remote experts in a recreation of the task environment and enable interactions that feel more natural and embedded in the physical space.

While VR devices that offer immersive and spatial experiences have come a long way and could be on the brink of being ready for large-scale consumer applications, there are still a lot of constraints to their use. While we don't fully understand how the differences in how we perceive virtual content impacts user experiences, today these constraints primarily stem from technological limitations of the current technology. These include the computational complexity involved with recreating remote spaces, and the bandwidth limitations for real-time communication. Together, they impact how much of a remote space can be recreated, the fidelity and therefore utility of these recreations, and the type of communicative features that can go with it. Even though the technology is developing at a rapid pace, these technological constraints are likely to remain a reality for some time to come. Leveraging the unique capabilities of this medium, therefore, requires understanding how to design effective interfaces that fulfill collaborator needs alongside the constraints that come with it.

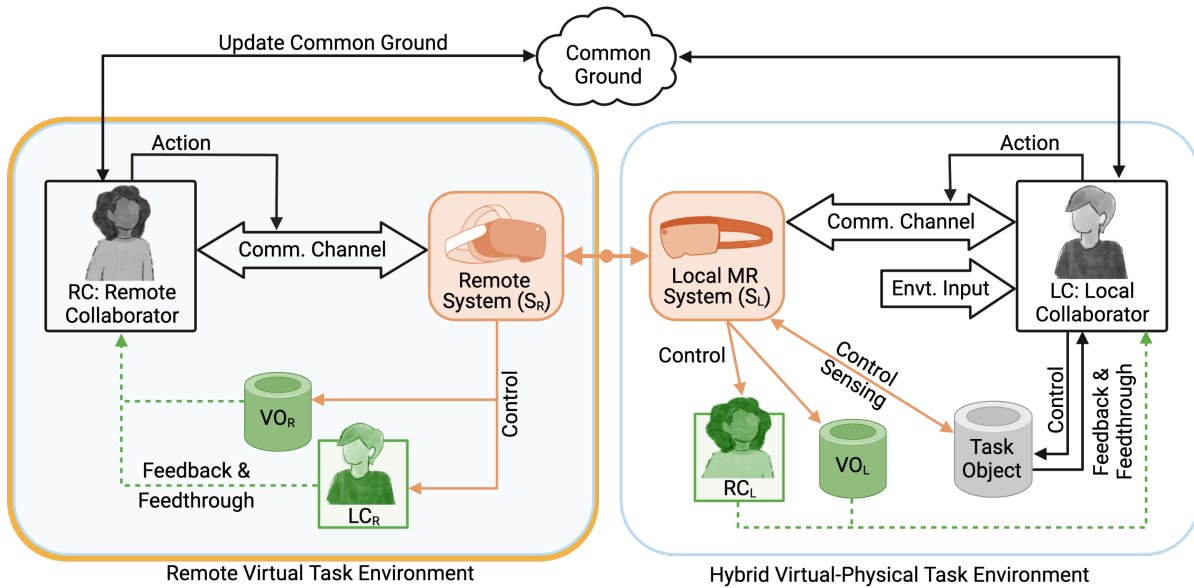


Figure 4.1. In this chapter we focus on the choice of immersive technology to provide a virtual task environment for remote experts (highlighted) and evaluate its impact on their ability to provide MR-mediated guidance in time-critical settings.

In this chapter, we use a participatory design approach to create a VR interface for remote experts in an effort to understand how they envision using the additional immersive and

spatial capabilities this medium brings to the practice of telementoring in a time-critical setting – trauma surgery. Our approach focused on building an interface that attempted to fulfill experts’ domain-specific needs while working with the practical limitations of the technology. We also report on experts’ use of the resulting interface during a feasibility study where they provide MR-mediated guidance to novices in a complex and realistic setting and discuss the benefits and challenges associated with it.

4.1.1 Supporting Telementoring for Trauma Surgery

Traumatic injuries – whether from natural or man-made disasters – happen suddenly and have a high degree of severity that requires immediate medical attention to prevent mortality and morbidity. However, expert trauma surgeons are often not on-site in resource strapped regions, rural medical centers, or when large-scale disasters overwhelm local hospitals. In these situations, where timely access to medical expertise is challenging, local surgeons or medics rush to perform life-saving interventions to stabilize patients before initiating a transfer to a specialized surgical facility. When that is not possible, telementoring where an expert surgeon uses a collaborative system to guide a novice becomes the best way to perform specialized trauma surgeries [13, 260].

The most common way medical professions provide surgical guidance today is through audio calls where remote experts verbally instruct a novice. The advances in technology have also led to telementoring systems where the expert surgeon can annotate a live video feed of the surgical site to provide instructions to the novice surgeon [31, 216]. Unfortunately, both audio and traditional audiovisual systems have limited ability to impact the physical care required for trauma procedures and constrain the level of surgical consultation that can be offered [101, 31]. Moreover, these systems typically underperform when compared to settings where collaborators are side-by-side during physical tasks [175] primarily because collaborators can see and understand each other and the task environment, making it easier to communicate through a combination of verbalization, gestures, and actions [108].

When it comes to existing XR-mediated collaborative systems for surgery, a collaborative

system called STAR [196, 260, 259] provides novices with a head-mounted display through which they can view an expert's annotations and instructions on the patient body. Here, the expert surgeon sees and annotates the patient view on a flat monitor which makes it quite difficult to communicate, say the depth of incisions, and forces experts to learn a new way of working. Leveraging XR technologies to create immersive interfaces for remote experts could expand the type of gestures and collaborative behaviors they can engage in and offer more intuitive ways to incorporate the 3D nature of complex surgical guidance.

4.2 Designing an Immersive Expert Interface

In order to explore the possibilities and limitations immersive technology brings to XR-mediated collaborative physical tasks, we engaged seven domain experts (4 surgeons, and 3 operating room technology specialists) from the Naval Medical Center in San Diego, California in multiple participatory design sessions. The goal of these sessions were to understand how domain experts envision using the interaction capabilities XR offers to fulfill their domain-specific needs and provide MR-Mediated guidance to a novice in a time-critical setting.

4.2.1 Participatory Design and Prototyping Sessions

The domain experts participating in our study were engaged in multiple sessions. We first started with a design thinking exercise where they walked around a room that was divided into different areas. These areas reflected possible stages of communication during a surgery (establish context, pre-operative, procedure, post-operative), and as the experts walked around, they wrote down what they envisioned they would need to be successful. Experts then participated in multiple roleplaying [287] and bodystorming [229] sessions and worked the research team (that consisted of technology and design researchers) to refine and bring their ideas to life. During these sessions, experts interacted with, and provide feedback on prototypes in a mock operating room and expert space we set up in a lab (Fig. 4.2). In all of these sessions, a researcher or another expert acted as the novice.



Figure 4.2. Model operating room (left) and expert space (right) set up in the lab for iterative design sessions.



Figure 4.3. Design sessions with domain experts, technology researchers, and design researchers.

We started with low-fidelity sketch prototypes using a custom-made AR prototyping tool [102] so as to reduce the impact of technological constraints on their thinking, and created prototypes with increasing fidelity over time as a way to introduce the technology constraints and understand the design options that could work with them.

4.2.2 Expert Needs, Envisioned Features, and Workarounds

Here, we discuss the different collaborator needs during remote trauma surgery and report on how experts envisioned using the additional capabilities the technology offers to fulfill them. We also discuss how the technological constraints shaped the final design of the expert interface.

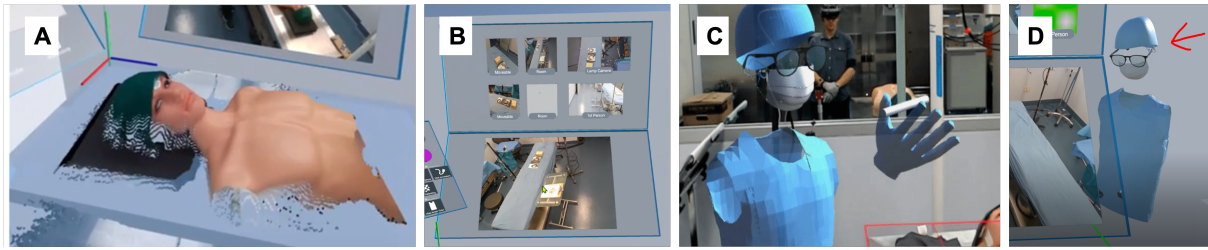


Figure 4.4. Features that provide remote experts with a view of the task environment and collaborators.

3D Reconstructions of the Task Environment

One of the defining capabilities of VR is that it can present its user with 3D visual content and emulate the experience of being in a real physical space. Experts envisioned using this capability to have an interface where it felt like they were in the task environment, performing a real surgery – they felt that the additional depth perception and being immersed in the task space would make the experience feel more natural and intuitive when compared to traditional media.

Current depth cameras and the computational complexity involved with creating a 3D reconstruction of a remote space, however, limited the region of the task environment that could be covered. On understanding this, experts emphasized that the most crucial portion of the task environment to cover would be the surgical site. They expressed that this region is where they need the most information from, and watching the novices perform actions within it in a spatial manner produced a more embodied experience when compared to a video feed of the same region. Moreover, having a patient body that felt visually realistic could also help them tap into their tacit experiences better - something they envisioned would be especially helpful if they were to annotate on the body.

The final interface included a depth camera attached to the surgical lamp in the novice's environment to recreate the surgical site for the expert (Fig. 4.4-A).

Video Feeds of the Task Environment

Another benefit of a VR interface is that it has the ability to supplement the 3D content it presents with additional information in more traditional formats, and allows for "free-floating" 2D interfaces to surround the user. Experts envisioned using this capability to include additional 2D video feeds to their interface as a way to work around the limited region a 3D reconstruction can cover. They also commented on the higher visual fidelity 2D videos offer (albeit without depth) and found that crucial to aid their situational awareness. Experts opted to include different types of video feeds to their interface, with each type serving a different purpose.

- *A first person view of the novice:* As experts used the 3D reconstructions of the surgical site, they found that its grainy nature did not provide them with as much detail as they would have liked. A first person view (FPV) of the novice from a camera on their MR headset was a design option that could provide them with a better view of the surgical space as the novice performed the surgery. Additionally, in being tied to the novice, this view also provided the experts with an accurate sense of exactly which parts of the task environment they were seeing, providing experts with a better sense of the novices actions, even outside of the surgical site. Being tied to the novice also allowed experts to direct them to move towards artifacts they wanted to see better - experts envisioned directing novices to move closer to parts of a surgical site to get a better view of the underlying anatomy, something that was difficult to see from afar or through the depth camera since once an incision was made, the blood and general lack of light reaching it made it hard to see details.
- *Video feeds from stationary room cameras:* While the FPV gave experts some sense of the operating room, experts found it difficult to get a full view of the space since it was a moving view that could be quite jittery. The experts however, mentioned that it was also important for them to be aware of what was around the surgical site. Stationary cameras positioned in the corners of the room helped provide this additional context which included

a view of the tools and equipment available to the novice and other people (like surgical assistants) in the space. Experts also found it useful to have different angles of the patient body.

- *A moveable camera:* In general, experts did not want to miss out on any information from the task environment. Experts found that the 3D reconstruction of the surgical site could be occluded when novices were performing actions that required them to be in the line of sight of the depth camera, and the jittery nature of the FPV made it fatiguing to glean details from. Therefore, a camera on wheels was introduced to the system – allowing for an up-close, detailed view of the task environment where experts needed it in the moment.

While the different views provided the information richness experts wanted, it could be difficult for them to attend to all of them equally at once. However, experts emphasized their need for all the information these different views would provide. Therefore, in the final interface, we opted to provide a 2D panel that housed and labelled all the different views. This panel was designed to be placed in front of the 3D reconstruction and allowed for experts to choose one view to expand on that would primarily stay in their field of view when they were looking at the 3D reconstruction of the surgical site (Fig. 4.4-B).

Collaborator Avatars and Hands

Experts also envisioned having avatars to represent the collaborators to each other. A avatar representing the expert user could be presented to the novice surgeon, and an avatar representing the expert could be placed in their immersive interface. These representations of collaborators would make it seem as though the novice was in their interface with them and could increase the sense of co-presence as they performed the surgery together. In our design sessions, people looked at the avatars of their collaborators as they spoke even though the audio itself was not spatialized.

Experts also envisioned being able to share their hand gestures which could help with



Figure 4.5. Features that help remote experts communicate technique.

referring to different artifacts on the patient body by pointing towards it, and it could also help experts communicate how to use different instruments as well enact certain maneuvers that novices could then mimic to complete certain actions by standing in the same location as the expert (relative to the patient) creating a “ghost hands” effect for themselves[[CHI’21]] (Fig. 4.4-C&D).

Annotating the Surgical Site

Communicating technique is a key aspect of providing guidance for a physical task, and surgical techniques tend to be quite complex and requires frequent references to different anatomical aspects. Expert surgeons therefore found it crucial that they were able to make these references in a way that was easily understood by the novice. MR’s capability to provide information in-situ was seen as beneficial and experts envisioned being able to draw on the 3D reconstruction of the surgical site to annotate it so that it could show up at the same spot for the novice. Experts prototyped using this to draw on the patient body to indicate where an incision needs to be made that novices could trace along with a scalpel. They also imagined using this to indicate different anatomical artifacts within the surgical site once it was opened up and also used them to communicate the technique required for different procedures (Fig. 4.5-A).

Experts also thought it was crucial that these annotations should be of varying thickness, and had the option of different colors that would easily show up on different skin tones or over blood. The final interface included the ability for experts to communicate using annotations and a control panel that they could use to modify the color and thickness of it (Fig. 4.5-B).



Figure 4.6. Sketch of the features that could help novices identify tools.

Guidance for Surgical Instruments

The names for the specialized instruments physical tasks require tend to very domain specific and surgery is no exception. Surgery, in fact, is known for having eponymous names for instruments that are obscure for any outside of a particular sub-speciality. Experts predicted that being able to direct a novice to pick up the right instrument could be quite hard since they would not be familiar with the names that are used for it, even if they had some surgical training. This is especially true for more specialized procedures like neurosurgery.

During the design process, experts envisioned being able to make it easier to direct novices to the right instrument in a few different ways. They envisioned being able to point at a video feed with a laser pointer and to have that laser point be replicated in the novices space. However, it can be quite difficult to see small surgical instruments from afar on a video feed, so the experts and researchers explored other ways of making it easy for the novice to understand which instrument was being referred to. XR technologies also include tracking capabilities to support its environment understanding which meant another option was to have the tools in the novice's space be tracked – this would allow experts to refer to it by saying the name of the instrument, or picking up a similar instrument in their own space which could initiate visual indicators for the novice that highlights the instrument for them in their operating room. Sketches of these design options are shown in (Fig. 4.6).

While being able to refer to the instruments and get a sense of what instruments were in

the novice's operating space was important to experts, they felt that they could verbally describe the instruments to the novice since they just needed to visually search for it among a set of finite instruments available to them. Therefore, in the final version, this feature was not implemented.

Pre-made Video Clips for Complex Procedural Guidance

Experts found it difficult to verbalize or draw about some complex surgical procedures and actions, and envisioned having pre-made video clips for different procedures that could make it easier for the novices to understand maneuvers, especially when they had no prior familiarity with the procedure or experience that they could use as a reference. The final interface, therefore, provided a library of video clips that experts could choose from and had that video clip show up for in the novice's mixed reality space (Fig. 4.5-C).

Haptic Feedback for Experts

Surgery is a very tactile task, and surgeons often used their sense of touch to feel for different anatomical structures. They often rely on how certain body features feel to confirm an anatomical landmark – like finding a soft spot to make an incision for cricothyotomy. Experts envisioned being able to use haptic feedback along with the 3D reconstruction to help them get a better sense of the patient's specific anatomy. However, accurately sensing and replicating pressure and other haptics can be quite challenging with technology today. Therefore, the final interface only included visual elements.

Control Others' Spaces

In our prototyping sessions, the representation of the experts hands and the annotations frequently came in the way of the surgical site for the novice. It was, therefore, important to have a way novices could hide the digital content in their workspace or control the interface in general. However, experts were also acutely aware of the type of the cognitive load and stress a novice would experience performing an emergency trauma procedure without prior experience. Plus, the novice's hands would typically be holding instruments or within the patient body (making

it hard for them to control the interface). Using voice commands could also be difficult since the novice could be in a noisy environment. Experts therefore envisioned having an additional control panel in their interface that they could use to control the novice's interface based on their verbal requests.

4.3 Evaluating Collaborative XR for Casualty Surgery

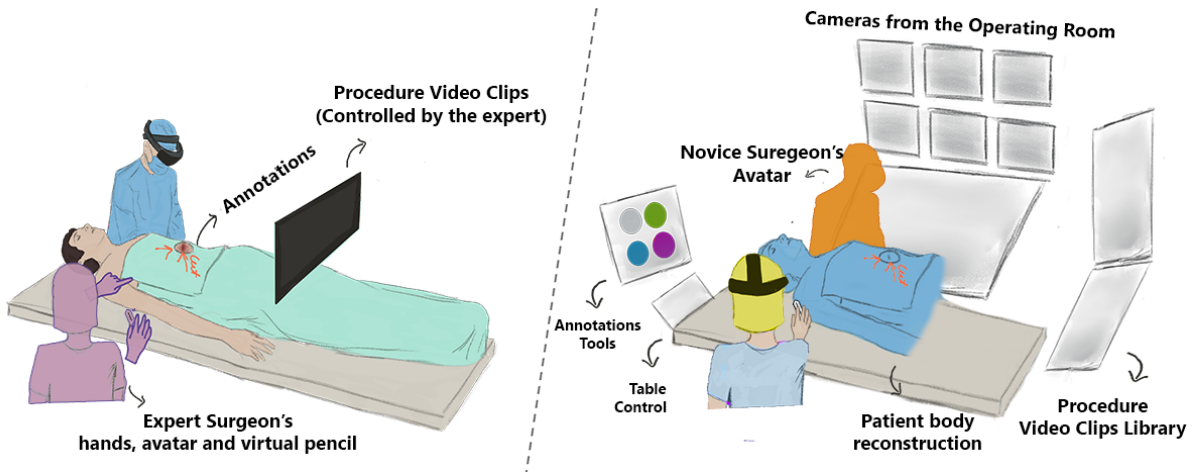


Figure 4.7. An artistic rendition of the features integrated into final design for both the remote expert and the local novice.

4.3.1 Feasibility Study with Perfused Cadaver Models

After our design sessions, the final interface was implemented and then deployed at the Naval Medical Center in San Diego, where it was evaluated in a feasibility study for combat casualty care procedures using a perfused cadaver model. The purpose of this study was to understand whether XR-mediated casualty care successfully support complex surgical procedures, especially given the limitations of the technology. In the study, the final system was evaluated against audio-only telementoring set up (which is the current standard of care for emergency trauma telementoring) during a 10 month period where 14 novice surgeons and medics completed 69 different procedures (38 with XR, and 31 with audio) with guidance from 3 critical care surgeons and 3 neurosurgeons. The procedures performed ranged from relatively simpler trauma

procedures to complex and specialized neurosurgical procedures. The expert surgeons performed surgical procedures with both the XR system as well as just audio. Once completed, video recordings of the surgical site (without audio) were subjectively evaluated by expert surgeons who were blinded to the condition. A full report of the study and the procedures are presented and published separately [291]. In this section, we report on the key results of this feasibility study and provide some context for it through our qualitative observations of 8 sessions (which included multiple procedures).

32 of the 38 procedures performed with a collaborative XR system were performed correctly, and the system also allowed for the successful completion of neurosurgery procedures when none of the 5 novice surgeons had performed any cranial procedures before. Therefore, despite the technological challenges, the additional visual and spatial capabilities of XR supported the completion of multiple procedures in a realistic setting, showcasing the feasibility of this interface in supporting time-critical surgeries. The blinded expert scores suggested that the XR capabilities trended towards demonstrating improved accuracy for the procedural steps with MR-mentored novices (4.12 for MR, 3.59 for audio) and timeliness (3.92 vs 3.47). Surprisingly, there were no statistical differences between the blind scores for the procedures completed with the XR interface and those that were audio-guided [291]. In the remainder of this section, we attempt to uncover why.

4.3.2 Qualitative Observations of Interface Use

The experimental nature of our system meant that all of the components were not yet at a stage where it was fully ready for wide-spread consumer use. The various technological challenges associated with the system are discussed in details in a paper by Gasques et al. [101]. The two key challenges experts faced during the study was that (1) the 3D reconstruction was grainy and lower fidelity than they would have liked, making it difficult for them to make out the details they needed; and (2) the annotations and representations of their hands had some offset in the novices space, making it hard for them to achieve a high level of accuracy in terms

of the placement of these communicative cues for the novice. These challenges, including the additional time required for calibration steps and the novelty of the system could be the primary reason the timeliness of the procedure didn't improve as much when compared to audio guidance. However, we wanted to dig deeper into the nuances of this effect.

During the surgical sessions, three researchers made observations about the user interactions with the system and the verbal communication between collaborators in order to better understand how experts used the visual information and communication aids provided by their VR interface and its impact on the collaborative task as a whole. The experts and novices also filled a simple questionnaire about their experience with the system after each session. Due to a loss of data due to constraints in our collection procedures, we only have the data for 31 procedures. In this section, we present the major themes from our preliminary qualitative analysis of this data.

Experts and novices found the XR system helpful despite technological constraints

Experts cited the fidelity of the 3D reconstruction, the recalibration steps required, and inaccuracies with the annotations as the main things that hindered their flow of providing guidance. Despite these technical challenges and the constraints they placed on the communication, experts were still able to provide effective guidance for a number of complex procedures with the VR interface and were impressed with what it could achieve saying *"To do those 5 procedures in 40 minutes, especially 2 of which he's never done before is pretty great"*. Novices too thought it was *"pretty neat"* for them to be able to see guidance cues directly on the surgical site. This system was the first immersive system to ever be used in a realistic setting for surgeries, and therefore showed great potential for MR-mediated surgical guidance system that provided multiple information and communication aids in an immersive interface.

When asked whether the added visual information helped them tell when the novice needed assistance, experts' average rating with the VR interface was 4.4 (out of 5) as compared to 3 with audio. Experts ranked the first person view as the most important feature that impacted

their ability to provide guidance (4.6 on 5), with the 3D reconstruction of the surgical site and the ability to annotate on them being the second most important (3.6 for both), then the room cameras (3.4). Experts did not find the additional procedure clips particularly useful and provided a score of 1 for it. Experts also mentioned they benefited from their ability to toggle between the different video feeds, allowing them to focus on the information they deemed most important in the moment. Novices deemed the annotations that communicated landmarks and incision positions to be the most helpful *“despite being off by 1 cm”*. This information needed to be communicated verbally in the audio condition.

It was hard to achieve situational awareness through the 3D reconstruction

The lower fidelity of the 3D reconstruction of the surgical site meant that experts and novices spent a lot of time in conversation to confirm either what they were both seeing or to check in about whether an action had been performed correctly. For example, novices constantly asked their expert *“do you see what I’m seeing doing now?”* and the expert often asked for confirmation of what they were saying with phrases like *“looks like your straps are through, is that correct?”*.

However, these confirmations were also not always easy. Experts often switched to video feeds to get this information, and asked novices to move around or closer to certain artifacts so that they could get a better view from their FPV, especially when the details were not clear from the 3D reconstruction. For example, when they wanted to get a better view of the chest tube insertion site, the expert directed the novice to go closer saying *“OK umm...so go ahead and look at the chest tube site...a little closer with your hololens, I can’t see [it] in the hologram”*. Experts also used verbal descriptions of what they were seeing to confirm that this was indeed the state.

Representation of hands and annotations help with communicating technique

Experts regularly used their virtual hands to communicate how a particular action needed to be performed. For example, experts used the representation of their hands to show novices

how to use a clamp or retract a chest tube. This allowed them to use more deictic phrases in their communication without extensive prior verbal descriptions. For example, one expert corrected a novices placement of an instrument by demonstrating the right placement with their hands and saying “*I would turn it a little bit more like this*”. Experts also used annotations to communicate a certain technique. For incisions, they drew annotations on the patient body in the direction that the incision needed to be made. They also drew out anatomical structures or explained the motions for a certain procedure by drawing abstract representations of it mid-air.

However, the limited field of view of the HoloLens meant that novices did not always see the experts hand gestures or annotations. Experts often talked as they gestures and drew and would then realize that the novice did not have all the context required to understand their supplementary verbal instructions. This led to the need for repair strategies in the communication, and over time, experts started bringing attention to their gestures by explicitly telling the novice to look up.

Video feeds acted as the primary source of information

Experts primarily used the video feeds (especially the FPV) as their primary source of information, especially because the lower fidelity of the 3D reconstruction made it hard for them to glean details from it. Additionally, since the novice frequently needed to bend over the surgical site, the reconstruction was occluded from the view of the camera and therefore incomplete. We observed that experts primarily kept their awareness of the task status up-to-date through the video feeds available to them and used the 3D reconstruction of the surgical site as a communicative aid, shifting their attention to it when they needed to gesture towards a landmark or annotate on it. Once the expert had performed a communicative action, they would look back at the videos, especially the first person view, to confirm that the novice understood what they meant and get a sense of what they can and can’t do. If the novice needed additional instructions at this point, the novice would typically opt to verbally instruct them, narrating through the actions that needed to be performed. For example, when a novice was having trouble finding a

specific space between the patients ribs, the expert watched the FPV to provide additional verbal information, giving the novice more nuanced feedback about how to place their fingers based on where they currently were.

Switching between information sources led to a loss of information

Since experts were heavily reliant on the video streams for visual information, they switched between the different feeds often depending on what they deemed most useful in that moment. They also constantly switched their attention between the 3D point cloud and the videos. This meant that experts spent time searching for the optimal source of information, especially when the novice was occluding parts of the surgical site in multiple videos. Moving between different pieces of information sometimes caused experts to miss pieces of information that impacted their awareness of the current state of the task. For example, when a novice was trying to show the expert an instrument, the expert mentioned they could not see what they were showing them clearly and begins switching between different video feeds looking for a better view. In the meantime, the novice repositions the movable camera on wheels to be in direct view of it and asks “*Can you see this better?*”. The expert, however, doesn’t notice that the view of the movable camera has changed and is trying to understand what the novice is saying based on the room cameras and mentions that they still can’t see it, which prompted the novice to explain it verbally.

Referring to anatomical landmarks required a lot of verbal confirmation

Referring to different anatomical landmarks is a key aspect of providing guidance to a novice, and this formed a bulk of the communication between the novice and the expert. Experts regularly used the communication aids available to them to refer to different landmarks on the surgical site and patient body. For example, experts used their hands to show where on the throat the novice should make an incision, or drew small dots and lines to annotate the different points of interest on the patient body and then verbally explained what the action involving them were.

Experts also used hands and annotation together – using the hands to gesture to the general area they wanted to pay attention to, and then drawing annotations to reference more specific artifacts within the region.

However, annotations and gestures often appeared a little off where they were originally intended for the novices. This offset made it quite difficult for experts to use it to refer to precise locations on the patient body or within the surgical site, especially once an incision was made and opened up. They often asked for confirmation of where the annotation appeared for the novice before continuing. For example:

Expert: *“What did I just draw now?”*
Novice: *“You just drew on the groin line.”*
Expert: *“Was it exactly on the crease?”*
Novice: *“Yeah, pretty much fallen on the crease”*

When the annotations’ offset was too off, experts often spent time verbally describing where they intended for it appear. They then asked for the system to be re-calibrated and tried again. For example, when the novice described a landmark incorrectly based on where they saw the expert’s drawing, the following interaction took place:

Expert: *“when I look it looks close to the sternum”*
Novice: [re-calibrates the system]
Expert: [redrawing the annotation] *“There. Can you show me where that is? Where do you see me drawing?”*
Novice: *“Right now I see you right at the sternal notch”*
Expert: *“Perfect”*

In cases where the re-calibration wasn’t fully successful, experts simply verbally described the landmarks and then asked the novice to point at them to confirm their understanding, verbally correcting the novice when necessary. Over time, the additional communication and frustration that they experienced because of these inaccuracies led to a reduced use of annotation, where they primarily used communicative aids to refer to different general locations on the body and communicated the specifics verbally. Experts also sometimes annotated in mid-air to clarify their rough annotations on the body. For example, one expert circled a few rough areas on the

body and then drew above it and directed the novices attention to it to give “*a spatial relationship between things*”.

References to surgical instruments relied solely on verbal communication

Referring to the surgical instruments required is another key aspect of providing surgical guidance. Since the system itself did not explicitly provide support for it, experts and novices primarily used the video feeds for the communication required for it. For example, when a novice wasn't sure which suture to use and was looking through the different options on amongst their instrument kit, the expert used their FPV to help them pick the right one using verbal descriptions like “*the long stringy thing*” when the novice was unfamiliar with the proper term for it. Experts also used the video feed to correct novices when they picked up the wrong instrument saying things like “*use the bovie, not the mosquito clamp*”. We also observed novices occasionally showed the instrument they had on top of the surgical site in hopes that the 3D reconstruction would include it for the expert saying “*do you see what I have?*”. However, the metallic nature of surgical instruments made it difficult to be reflected in the point cloud.

Experts often wanted to get a sense of the different instruments available to the novice ahead of time so that they could refer to them more directly during the procedure where the time spent looking for instruments needed to be minimal. For example, when an expert asked them for a description of the instruments, the novice placed their kit in the line of site of a camera saying “*this is what I have*”. Experts also explained that they will be spending some time gathering supplies and used the novices FPV to get a sense of the instruments as the novice walked around the space and verbally identified and explained the equipment to be used in the procedure.

The information richness and technology constraints leads to communication fatigue

While the VR interface provided experts with a lot of information and in a way, fulfilled their need for extensive visual information about the novice's operating space, getting the exact information they needed at a given moment took a lot of effort (like switching between the 3D

and 2D sources, and looking for the optimal video feed). They also needed to confirm what they were seeing with verbal information. The communicative aids also required experts to perform a lot of physical motions. The toll this takes on an expert was evident in some of our sessions as we noticed experts placing their hands on their head to indicate strain and shrugging or rubbing their shoulders as though they hurt after drawing a series of annotations. In one of our pilot sessions, an expert verbalized this saying *“If my head keeps moving up [looking at the interface with video cameras] and down [back to the body], it will break”*.

Over time the loss of trust in some of the features because of the technological limitations also led to them resorting to do most of the communication verbally and experts often indicated this shift saying things like *“I’m just going to talk you through this”*. A novice reflected on this saying that the guidance was verbal *“90% of the time”*. Now, the verbal communication itself was quite effective since novices were still people with medical training and a lot of the referencing required could be made unambiguously using the medical term for it (especially with respect to anatomy, instruments sometimes needed additional verbal descriptions since their names are not part of general medical training). However, this still takes communicative effort and causes fatigue.

4.4 Implications for Designing Remote Expert Interfaces

Through our design sessions, we were able to better understand the kind of support remote experts needed to facilitate the completion of a complex task. And the feasibility evaluation and our qualitative observations from it gave us a sense of how the different features available to experts impact the collaboration needed to help novice medical personnel perform trauma surgeries in a realistic setting. In this section, we discuss the key takeaways from both these endeavours and outline the implications of it for the design of future remote expert interfaces that supports their ability to provide effective guidance.

Having the additional features an immersive interface allows for - like combining different

types of visual information and providing communicative aids that feel more intuitive enabled expert surgeons to provide effective and rich guidance for a large number of emergency trauma procedures. Despite the largely experimental nature of our system and the various technological constraints that came with it, the feasibility study showcased the immense potential of such a system to extend surgical expertise to remote and expertise strapped regions. However, the qualitative observations of how remote experts used the different elements of their interface to support their information and communication needs also highlights the different gaps we currently have and outlines a pathway for the future research needed to build rich collaborative systems that leverage technological capabilities well.

MR-Mediated guidance is effective for novices and can support guidance for complex tasks in time-critical settings

While the experiential challenges collaborators faced due to the limitations of technology made providing MR-mediated guidance difficult for experts. Now, surgery is a highly standardized, albeit complex, physical task with domain specific jargon that basic training covers. Therefore, in our feasibility, it was not difficult for experts to switch to verbal communication when the technology failed, as they could use relatively unambiguous terms to refer to anatomical landmarks. However, when it came to communicating complex maneuvers, referring to aspects of the anatomy that novices were unfamiliar with, and guiding them to pick up instruments that had eponymous names that only those trained in a certain sub-specialty would understand, experts resorted to using the MR-guidance, making it especially useful for true novices.

Moreover, most of the difficulties we noticed came from the remote expert's ability to either get the information they need, or from their difficulty in being able to provide accurate annotations and hand gestures. But when they did use these communicative aids in the absence of technology constraints, novices were able to follow the guidance with ease.

Information is crucial for remote experts, but different parts of the environment serve different information and communication needs

Surgery is a complex task rife with unknowns - surgeons constantly use their awareness of the current state of the task and the specific anatomical characteristics of the patient to assess what to do next. Trauma surgery in emergency scenarios, in particular, can be challenging because they often don't have an extensive medical history of the patient and also can't always predict the issues and complications that they have to deal with - which means that they very heavily rely on their ability to process what they are seeing and feeling in the moment to make crucial informed decisions to reduce the casualties.

In our design sessions, experts emphasized the need for this information and opted to have different sources of information when one seemed insufficient. Experts (rightly) thought it was crucial for them to understand all the details were very reluctant to make design choices that would mean they would miss out on any information from the environment. Experts used this information to stay aware of what was happening in the task environment and used this information to ground their communication with the novice. Outside of general task status awareness, experts needed this information to know *where* different instruments and anatomical landmarks were to then be able to demonstrate the actions that included them in the right location.

However, in our feasibility study we noticed that experts did not use the information from all parts of the novice's operating room equally. Experts spent most of their time focused on the surgical site where a bulk of the action occurs, and it was here where they seemed to need the most details from. They used the different sources of information to get as much information as they could about this space and constantly asked novices for verbal confirmations and more information as the task moved forward. While experts constantly kept themselves aware of what was happening in the rest of the environment, they primarily did so to get additional context about what instruments and equipment the novice had to be able to plan the task and direct the novices attention to it. When the conversation shifted to focus on the regions outside of the task space, it was primarily to clarify something, or to bring a certain instrument to the surgical site,

after which experts switched their attention back to the surgical site.

While we do need a more in-depth analysis and more research to understand the nuances of the different purposes of visual information in complex physical tasks, future collaborative XR systems can take the varying needs into account when designing interfaces.

Information richness alone does not make object-focused communication easier

Despite their ability to annotate on the surgical site and share their hand gestures, experts were not able to use these features well without information from the environment. Our observations of the feasibility study sessions highlighted that experts needed to where everything in the surgical site or the operating room was to be able to, say, annotate over it to provide guidance for the action that needed to be performed.

However, despite the extensive amount of information the VR interface provided, it was difficult for experts to get the information they needed for two main reasons (1) technological constraints that reduces the effectiveness of the information sources (like the 3D reconstruction being grainy), and (2) information overload that was evident by experts constantly flipping through videos to look for the ideal view.

Therefore, future research for collaborative XR systems should aim to increase the effectiveness of information sources like 3D reconstructions and also aim to decrease the different types of information provided to reduce the information overload and make it easier for experts to get the information they need.

The difficulty of technology-mediated referencing is often underestimated

Our design sessions revealed that despite understanding the need to reference to anatomical landmarks and instruments, experts focused on making it easier to communicate the technique required to novices, opting to not include features that could help with referencing because they believed it was easy to rely on a verbal description of objects and have the novice visually search for the object that matched it in their space. This belief is quite natural because the ease of being

able to refer to objects and artifacts in a physical space makes it feel like referencing is a trivial portion of the communication.

However, our observations of the feasibility study highlighted just how difficult this seemingly trivial conversational piece becomes in technology mediated conversation. In fact, referencing is the key aspect of the cooperation required for physical tasks that is fundamentally tied to the environment. And while it is indeed quite trivial when you have the information needed for it and the necessary representations to support it, communication breaks down pretty quickly in the absence of support for either – and we notice experts needing extensive verbal communication to clarify their references or reverting back to verbal information to communicate it. Future collaborative XR systems must therefore make it easier for experts to make references to the task environment.

4.5 Conclusion

The visual, spatial, and immersive nature of XR, especially VR, opens up the design space of expert interfaces during collaborative physical tasks by offering a more natural way of viewing remote environments and expanding the interaction possibilities within the virtual task environment. In this chapter, we co-designed an immersive interface with expert surgeons to understand how they envision leveraging the additional affordances XR offers to make it easier for them to provide contextual MR-mediated surgical guidance during an emergency. An evaluation of this interface showed that an immersive environment can help remote experts provide rich procedural guidance for novices. However, more in-depth qualitative observations of the interface use highlight that experts achieving the environment awareness required for seamless communication is still a challenging process that requires experts to process a lot of visual information. Referring to instruments and anatomical landmarks is the object-focused communication process most impacted by this as it heavily relies on experts' awareness of the

task environment. There is, therefore, a need for research towards mechanisms that enhance the support for referencing by either making it easier for experts to get the information they need or facilitate referencing mechanisms that don't rely on environment awareness. In the next chapter, we take a closer look at existing collaborative XR systems in the literature to better understand how they support collaborative physical tasks and discuss the gaps in the literature that provide an opportunity to make object-focused communication easier.

Acknowledgements

The research presented in this chapter was funded by the US Department of Defense, under the Joint Program Committee 6.7 research program (grant no. DP_67.2_17_1_17_J9_1650). The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, the Department of the Air Force, the Department of Defense, or the US government.

Cadaveric specimens were provided by and approved by the Department of Medical Education/Anatomical Services of the University of California, San Diego. Training of NMCSDBSTC staff to perfuse cadaver specimens was performed by MAXIFI surgical simulations¹ using their proprietary blood solution (Simblood).

This chapter, in part, includes material that appears in *ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring* by Danilo Gasques, Janet G Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Yifei Zhang, Wanze Xie, Xinming Zhang, Konrad Davis, Michael Yip, and Nadir Weibel in the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; and in *I Want to Be a Surgeon! Role Playing for Remote Surgery in Mixed Reality* by Thomas Sharkey, Janet G. Johnson, Danilo Gasques, and Nadir Weibel in Workshop on Interactive Systems, 2019 CHI Conference on Human Factors in Computing Systems. The dissertation author was one of the primary investigators of this material.

¹<https://www.maximumfidelitysurgicalsolutions.com>

Chapter 5

A Review of XR-Mediated Remote Collaboration for Physical Tasks

***Thesis Question:** How do existing collaborative XR systems facilitate communication between remote collaborators during physical tasks, and what are the research gaps concerning support for task object referencing?*

This chapter presents a review of the field of XR-mediated collaboration based on the conceptual framework introduced in chapter 2. We analyze the communication focused features 155 existing systems include to provide an overview of the design space and the current level of support collectively offered to facilitate the extensive coordination physical tasks require. An in-depth discussion of the virtual task environment for remote users and the support for referencing it offers highlights the research gaps that could drive future design explorations that reduces the joint effort required to partake in collaborative physical tasks.

5.1 Introduction and Motivation

Our research in the previous two chapters highlight the experiential challenges of using collaborative XR systems for remote guidance during physical tasks. A key takeaway from this research has been that object-focused collaboration and more specifically, referencing, is the major communication process that is fundamentally tied to the task environment. Being able to reference artifacts in the task environment from a distance in a manner that is easily understood by the local collaborator remains challenging even with the additional support XR offers. In general, the limitations of technology today coupled with the information overload and lack

of design patterns for XR make it challenging to support collaborators’ ability to comprehend technology-mediated information.

In this chapter, we aim to systematically understand how the field of collaborative XR as a whole leverages the technology to mitigate the difficulties of being geographically separated during collaborative physical tasks. We do this by systematically analyzing existing collaborative systems based on the conceptual framework developed in Chapter 2. The overall objective of the review is to better understand the gaps that could help us further the research towards easing object-focused communication. However, our review analyzes more features than just those designed to support referencing for completeness – primarily because a single feature often provides support multiple communication processes regardless of what it was designed for. We then present an overview of the design space the communication-focused features we see today occupies, and then dive into a more in-depth discussion about the level of support collectively provided for referencing task objects during physical tasks and outline key research gaps for it.

5.1.1 Review Methodology

We collected papers through a systematic search on Google Scholar, the ACM Digital Library, and IEEE Xplore. The search terms used were a combination of (1) “mixed reality”, its synonyms “augmented reality” and “extended reality”, and their abbreviations “MR”, “AR”, “XR” AND (2) descriptive terms for “remote collaboration” and “physical tasks” including “collaborative”, “synchronous”, “cooperation”, “remote”, “guidance”, “telementoring”, “spatial tasks”, and “real world tasks”. Papers were chosen to be included in the final dataset if any of the following criteria was satisfied: (1) the paper presented an XR system for collaboration; (2) the system used MR for the local user in a manner where virtual content or cues were overlaid onto the real world; (3) the collaboration setting was for tasks where at least one collaborator worked on real objects in their physical location; (4) the collaboration goal was to complete the task – as opposed to for training or education, and (5) the papers were published in the year 2000 or after. The PRISMA [230] flow diagram for our article screening is shown in Fig. 5.1. Appendix A lists

all the papers analyzed.

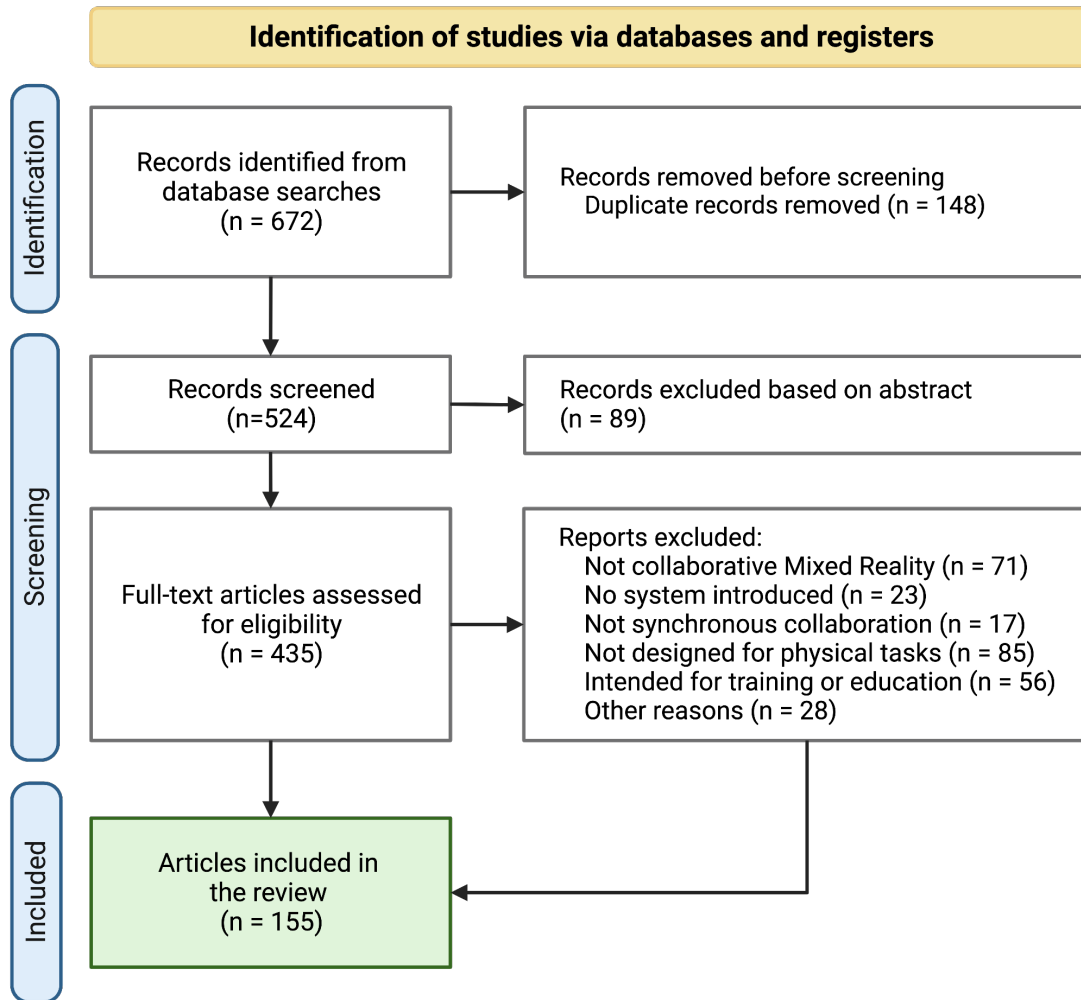


Figure 5.1. The PRISMA flow diagram for article selection and screening.

While our systematic compilation of this corpus provides an in-depth view into the research space, we consider these set of papers to be a representative rather than an exhaustive list since there may be papers that did not end up in our list using our search terms, or were at the edges of our inclusion criteria. However, our focus is to outline the current state of the design space and we believe these papers provide a representative set of the features being developed for MR-mediated physical collaborations.

ANALYSIS APPROACH

We used a multi-step analysis process in our review. The first step was an open coding approach to tag each paper on the type of technology used, the type of collaboration and task space it supported, the domain the system catered to, and the local and remote user features that the system offers. We then reviewed the tags from the initial coding to ensure consistency — merging tags, or modifying and removing duplicates as needed. Next, we tagged each feature with the collaboration needs (from Sec. 2.2) it supported as well as the mixed reality capabilities (from Table 2.3) it exploited. Then, we grouped these features into broader sets when appropriate, merging the tags for the needs it supported and the capabilities it exploited if needed. This coding was done using the qualitative research tool Dedoose¹. We also marked groups of features according to the level of communicative support (from Sec. 2.3.3 it offered as well as the situated experience it provides (from Sec. 2.5.2) when relevant.

Additionally, we used a “best intent” approach for our analysis. This meant including papers where the local user used any form of augmented or mixed reality that allowed for digital content to be overlaid onto the real world like tablet-AR or viewing the task environment through a phone, and not just papers that used MR headsets. The best-intent approach was also used when coding the needs a feature supported — all the aspects that were coded weren’t always laid out directly in the paper. Therefore, if it seemed reasonable (based on what other similar features), the features were coded leniently. The primary reason for this choice was because we are looking at what each feature affords across the field in general. So, for example, even though all 2D video feeds of the task environment did not really capture the local collaborator, 2D video feeds in general *can* capture local collaborators. Therefore, all 2D video feeds were coded for it. This rule of best intent was only broken when it would provide a false sense of what a feature could provide. This happens when multiple similar features were grouped together in the analysis to make the final grouping or feature type we list here. In that case, we retained

¹Dedoose Version 9.0.107, web application for managing, analyzing, and presenting qualitative and mixed method research data. Los Angeles, CA: SocioCultural Research Consultants, LLC www.dedoose.com

the numbers of those groups when coding and populating the table with the design space. For example, audio channels included both spatial and non-spatial versions. And therefore, when marking the feature for using the “spatially distribute content”, we only marked it for one of the channels (which included spatial audio). Similarly, facial or physiological cues included only 3 features that explicitly provided awareness of facial expressions and detected non-verbal cues, while 3 provided secondary information by detecting physiological cues. Another example is gaze and head pointer were grouped together, but there were only 4 features with head pointer, so the cells that correspond to the needs it serves reflects this (head pointer can tell you orientation, but is not an explicit source of gaze).

5.2 The Design Space of Collaborative XR Systems

The number of collaborative XR systems in research have grown through the years with its popularity rising substantially over the last few years (Fig. 5.2, Left). Of the 155 papers analyzed, 112 did not cater to a task within any specific domain – they were built to explore or study guidance systems in general. 13 papers focused on tasks in the healthcare domain, 29 focused on tasks related to maintenance and manufacturing, and 3 focused on other domains (Fig. 5.2, Middle). Additionally, a majority of the systems (141) tackled the mentoring scenario where a remote expert guides the local novice to complete a task, 2 papers focused on peer-to-peer collaboration, and 14 allowed for the role of the collaborators to be flexible (Fig. 5.2, Right).

Our analysis revealed 28 distinct types of features across the 155 papers reviewed. These features catered to both remote and local users – some for one of them, and some for both. We grouped these features under three broad themes based on the primary collaborator need they satisfy. 43.5% of the features we encountered were designed to facilitate some form of communicative action between collaborators, 39.3% were a source of information about the task environment for the remote user, and 17.2% provided one or both collaborators with an awareness of each other. In this section, we describe these features with a few examples.

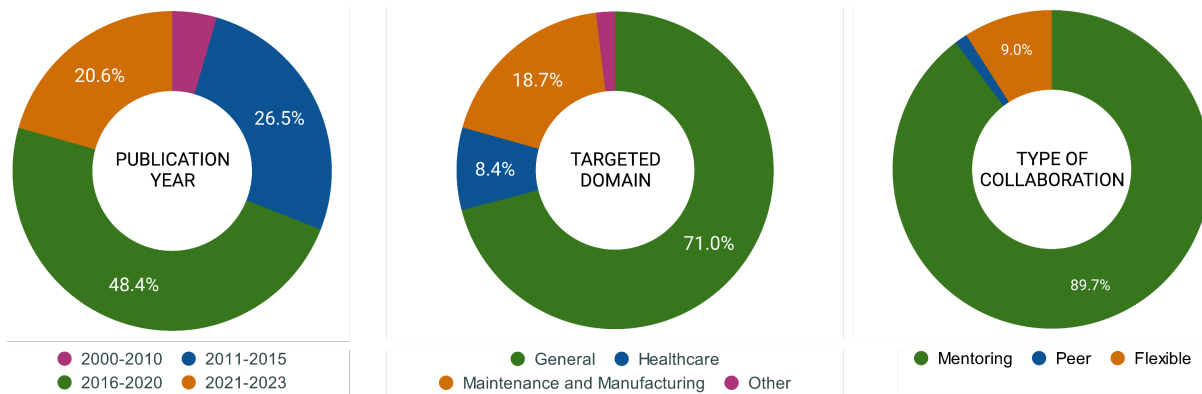


Figure 5.2. Distribution of papers presenting systems designed for MR-mediated guidance for physical tasks across the years (left), based on the primary domain it focuses on (middle), and the collaboration style it aims to support (right).

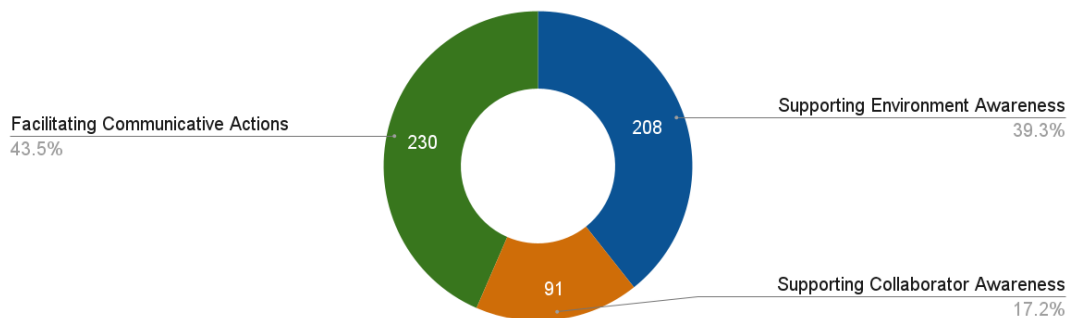


Figure 5.3. Distribution of features based on the primary collaborator need they satisfy.

The conceptual framework introduced in chapter 2 offers us the theoretical scaffolding to understand both the collaborator needs a system needs to support as well as the dimensions of support that can be provided for it. It also outlined the capabilities of XR as a collaborative medium. Therefore, in this section, we also contextualize the various features existing systems provide to support communication during remote physical tasks and present an overview of the design space to provide an understanding of how the different capabilities of MR were used to support various collaborator needs. A visualization of the design space the features in our dataset occupy is in Figure 5.1, and provides an understanding of how the different capabilities of MR were used to support various collaborator needs. The top half depicts the needs and type of support for it each of the features reviewed support, and the bottom half depicts the XR

Table 5.1. A visualization of the how the design space for collaborative XR is currently populated.

			Supporting Environment Awareness										Supporting Collaborator Awareness										Facilitating Communicative Actions																				
			First person View 2D Video Stream 360° Video 3D View 3D View + 360° Video Differently Scaled 3D View Choose Type of View Control Own Perspective Control Others Perspective Choose View Scale System Chosen View										2D Video Stream 3D Capture of Collaborator Collaborator Avatar Frustum View Indicator Gaze Cues or Head Pointer Location Representation Action Representation Facial or Physiological Cues Audio Channel Virtual Replicas of Task Objects Representation of Hand Gestures Fixed Environment Annotations Freehand Environment Annotations Virtual Pointers Highlight or Navigate to Task Object System Visualization for Technique																														
Total Number of Features			683	55	44	20	51	4	5	8	16	1	2	2	2	4	17	8	9	19	2	5	155	16	48	39	59	43	17	6													
Target User			657	55	44	20	51	4	5	8	16	1	2	2	2	4	17	8	9	19	2	5	155	16	48	39	59	43	17	6													
Features for remote user			223																																								
Features for local user												1				2	3	17	7	1	26	2	1	3	155	2	2	1															
Information and Communicative Needs	Environment	Verbal Channel	155																					155																			
		Passive Source	179	55	44	20	51	4	5																																		
		Explicit Representation	Presence of task objects	16																					16																		
			Spatial information of task objects	0																																							
	Collaborator Awareness	Passive Source	117	44	20	28	4											2	4											15													
		Explicit Representation	Collaborator's location	34																																							
			Collaborator's pose or orientation	36																																							
			Collaborator's view	78	55																																						
	Collaborator's gaze		29																																								
	Communicative		Communication Process Supported	Collaborator's facial expressions	3																																						
		Collaborator's action		1																					3																		
		Secondary awareness		2																					2																		
		Type of Mechanism		Referencing	206																																						
	Communicating technique		168																																								
	Acknowledgement		0																																								
	Outlining strategy		0																																								
	Mixed Reality Capabilities	Present multi-sensory virtual content	2D visual content	337	55	44	20		4											2																							
			3D visual content	369			51	4	5																																		
Audio signals			158																																								
Haptic feedback			2	1																																							
Virtual smell and/or taste			0																																								
Modify user's experience with the real world		Augment the real world	438																					3	17	7	1	26	1														
		Diminish aspects of the real world	0																																								
		Modify the presentation of reality	60																																								
		Block out the real world	75		20	51	4																																				
		Recreate remote locations	82			51	4	5											4																								
Enable spatial interfaces and understanding		Spatially distribute content	392		20	51	4	5											4	21	11	12	33	2																			
		Anchor content to the real world	261																					3	17	7	1	26	1														
		Detect and track objects	20																																								
		Track user location and orientation	53																					2	4	21	11	12		3													
Enable human understanding		Understand natural language	0																																								
		Detect non-verbal cues	47																																								
		Detect physiological state	2																																								
Provide independence of experience		Personalize content and interface	0																																								
	Allow freedom of movement	55			51	4																																					
	Allow freedom of perspective	104		20	51	4	5	8	16																																		
	Ability to modify environments	152							8		1	2																															

capabilities each feature exploits. The shade of the colored cells depict the frequency of the feature - or the number of times the feature has been seen across the 155 papers. A darker shade indicates a higher frequency. Note that a single feature can often support multiple collaborator needs and can make use of multiple technology capabilities. This shows which features are provided in the remote system and which ones are in the local system.

5.2.1 Facilitating Environment Awareness

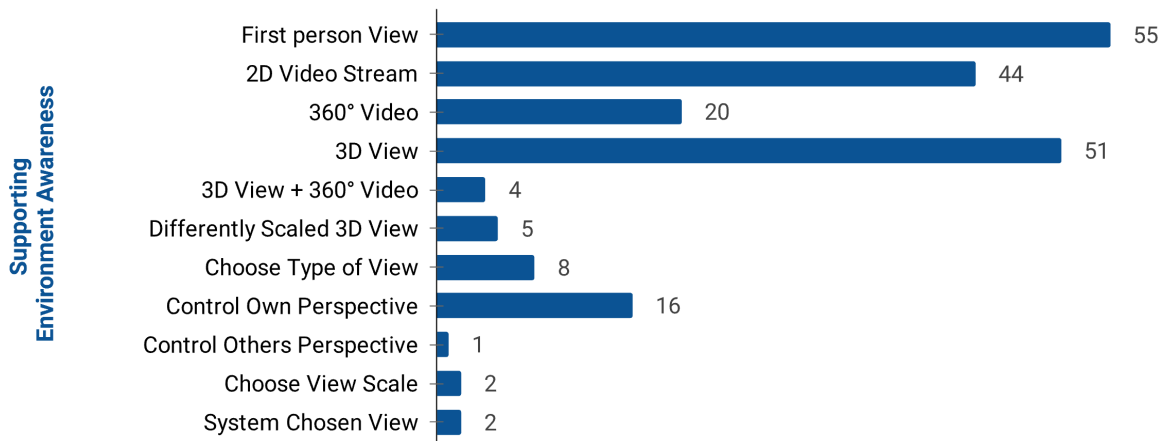


Figure 5.4. Features that support environment awareness for remote physical tasks.

A view of the task environment supports general environment awareness for the remote collaborator and helps build the common ground necessary for effective communication. It is not surprising that it is a heavily explored class of features with every paper implementing at least 1 of them for remote users to have a view of the physical task environment. A total of 11 different types of features cater to providing environment awareness in today's systems (Fig. 5.4).

The most common way collaborative systems support an expert with building common ground and situational awareness is through a (shared) view of the environment [92]. Prior research shows that for physical tasks, remote experts benefit most from a view of the workspace, task tools, and the novice's hands as opposed to one that focused on a view of the novice and/or their face [107, 108]. This shared view supports environmental awareness and improves commu-

nication efficiency by allowing for less reliance on verbal verification [209, 174]. Collaborative MR systems typically provide a view of the task environment through video streams, 360 videos, 3D reconstructions, or a combination of these features [63].

- **First Person View:** A popular method of providing remote users with a view of the environment is to stream a 2D first person view (FPV) that captures the perspective of the local user. These views are captured from a camera on the local user's headset [187] or mobile device [235], or from a separate camera attached to them [5]. In addition to providing a view of the environment, FPV's also allow the remote user to get a sense of the local user's field of view.
- **2D Video Streams of the Task Environment:** 2D videos streamed from external cameras that were continuously streamed to remote users is another heavily used feature that supports their need for environment awareness. These videos could provide a view of different parts of the task environment, with some systems only providing a view of the active space [318] and others providing a view of the full task environment. 2D video streams were used even when remote users were in VR [325], and some systems also included multiple video feeds that captured different parts of the environment [101]. One system combined a physical surface and a virtual plane where the video was rendered in VR to provide remote users with haptic feedback as they interacted with the video [319].
- **360° Videos:** Another way systems have shared the local environment with remote collaborators is through a 360° live video of the space with the remote collaborator being immersed in it through a VR headset [302]. While being in an immersive medium allows users to move around freely, 360° videos can also be experienced through traditional screens like in *360Anywhere* [281].
- **3D View:** Researchers have increasingly explored immersing helpers in 3D reconstructions of the task environment to allow remote users to engage with it like they would if they

were physically present in the task environment and bring the collaborative MR setting closer to face-to-face collaboration [26]. Systems chose to either recreate the full remote environment [100], just the active task space[354], or something in between [313]. While commonly experienced through a VR headset [265], some systems used other devices like a textured model on a flat screen [293] or stereoscopic displays that use active polarization [14] to display the 3D view.

Of the 52 features that were classified as 3D views – 28 were live reconstructions typically captured using depth sensors in the task environment [98], 21 were static recreations that were captured before the start of the collaborative session [300] or digital twins or pre-existing virtual models [225]. Unlike live reconstructions where the local collaborator can be captured as part of the view, static reconstructions don't offer any form of collaborator awareness. However, they do typically allow for better visual fidelity as they can be processed to eliminate visual noise. Some systems also combine a live reconstruction with static regions to provide a higher visual fidelity with real-time updates of aspects of the task environment deemed crucial [308].

- **3D View + 360° Videos:** Some papers also provided a combination of 360° videos and 3D reconstruction of the task space to provide full coverage of the task environment for the user without the added complexity of having a full 3D reconstruction [303, 301].
- **Differently Scaled 3D View:** Some systems provide remote users with 3D views of the task space that retains depth information but are scaled differently than in the real world. For example, world-in-miniature views are scaled down versions of the reconstruction [307], and views like the "miniature mode" [241] immerse the user in a larger-than-life reconstruction of the task environment.
- **Controlling View, Perspectives, and Scale:** Remote users are typically passive viewers while interacting with views of the task environment without any real agency within it.

However, some systems did allow remote users to control their experience to a certain extent. For example, remote systems allowed their users to choose between different modes of viewing the task environment during the collaboration (like switching between a 360° view and 3D reconstruction of the task space [301]). Remote collaborators could also explicitly change their perspective of the task environment like in *TeleAdvisor* [119] where experts could move a robotic arm that controlled a workspace camera and therefore their view of it. Immersive interfaces sometimes enabled remote users to switch between following their collaborator’s perspective or experiencing the environment on their own [237]. Systems also explored allowing users to change their own scale within a 3D view of the task environment – for example, a feature called “God mode” allowed remote users to become bigger and feel like a giant [241]. This freedom of scale and perspective that XR provides, facilitates an interaction with the task environment that is impossible in the real world. Lastly, systems like *SceneCam* [250] used the environment understanding capabilities of XR to automatically choose an optimal view for the remote user based on the local user’s location within the task environment.

5.2.2 Facilitating Collaborator Awareness

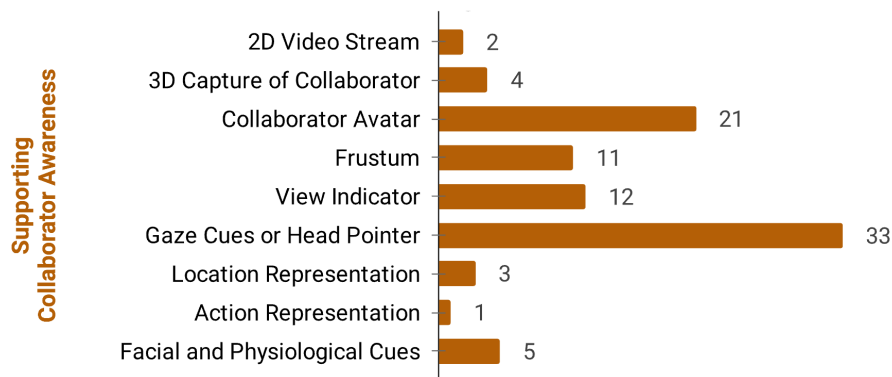


Figure 5.5. Features that support collaborator awareness for remote physical tasks.

Support for collaborator awareness is primarily achieved across 9 features that cater to both the remote and the local user (Fig. 5.5). A majority of these features provide collaborators

with explicit sources of information about each other.

- **2D Video Stream:** While collaborators are sometimes captured within the view of 2D video feeds meant to provide a view of the task environment, a few systems also provided users with video feeds that focused on each other [338].
- **3D Capture of the Collaborator:** Some systems provided their users with a real-time 3D capture of their collaborators, with 2 of these papers only having the functionality available for remote users. An example of this feature can be seen in *Loki* [307].
- **Collaborator Avatar:** Many systems used an avatar of some form to represent collaborators to each other, allowing users to comprehend each other's location and orientation within the task environment [324]. These can be full-body avatars [64], a representation of just the upper body [193], or just a head [300]. Not every system provided this feature to both users at the same time, with 17 of them having a representation of the local user in the remote user's interface and vice versa. While avatars typically don't capture the real user, they are a way to explicitly represent their location and orientation in the collaborative space in an anthropomorphic manner and are typically seen when the user engages with immersive technology.
- **Frustum:** One way collaborative XR systems provide an explicit representation for collaborator awareness is with a 3D visual cue commonly referred to as a Frustum that has its apex at the collaborator's eye or head, and a base that is an indication of their field of view. These serve as a combined cue that communicates a user's location, viewing direction, and view. An example of this can be seen in *CoVAR* [238, 241], and the system presented by Bai et al. [18].
- **View Indicators:** Systems also provide their users with a standalone visual cue that is an explicit representation of their collaborator's field of view with respect to the task

environment. These indicators can be 2D thumbnails [104], or a rectangle overlaid on the users view of the task environment [199].

- **Gaze Cues and Head Pointer:** Many systems also provided users with an indication of each others eye gaze [117, 7]. In addition to letting users know where the people they are working with are looking, users can also use their own gaze to refer to things they want to direct their collaborator's attention to. Researchers have also explored providing explicit cues when users focus on a certain region [144] or when both collaborators' gaze overlaps [145]. EyemR-Talk also allowed for certain speech phrases to turn on gaze cues for the remote user. Additionally, the Mini-Me system [239] redirects the gaze cue to indicate the right location in the task environment for the local user when the remote user's scale changes. A few systems also provided users with visual head pointer cues that indicated their collaborators viewing direction [313, 320].
- **Additional Features that Aid Collaborator Awareness:** Collaborative XR systems also explored a few other explicit representations that would support collaborator awareness. For example, *SceneCam* [250] provided a 2D representation of the local user's position and orientation within the environment, Rasmussen et al. [248] also included audio cues that indicated where the remote user was, and Bai et al. [18] included a 3D arrow that always pointed in the direction of the collaborator's avatar. 2 systems explicitly captured and represented users' facial expressions [85, 25] and others like *Empathy Glasses* [194] also included representations of their heart rate and galvanic skin response in an attempt to help build an empathic connection among collaborators. Additionally, one system provided the local user with audio and visual notifications of the remote user's interface actions [49].

5.2.3 Facilitating Communicative Actions

Support for actions that collaborators can use for key communication processes like referencing, communicating technique, acknowledgement, and providing context is especially

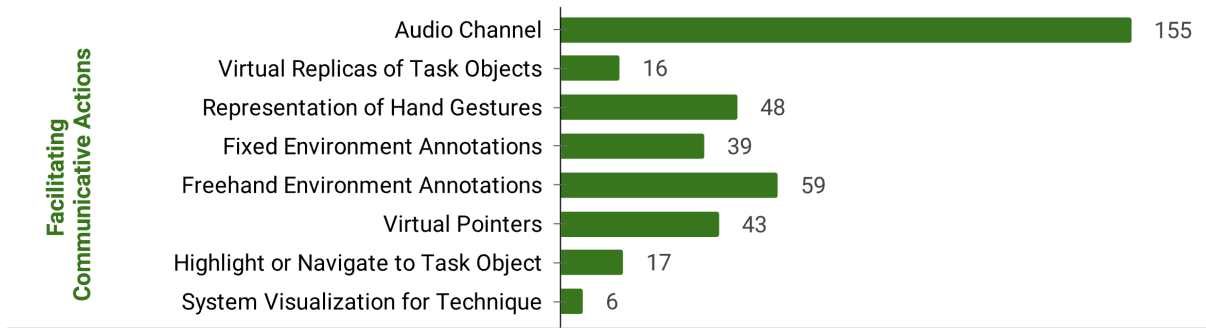


Figure 5.6. Features that support communicative actions for remote physical tasks.

crucial in systems that support remote physical tasks – as without it, no communication can take place. The majority of features encountered in our dataset aim to support at least one communication process and can be categorized under 8 different types of features (Fig. 5.6)

- **Audio Channel:** Even though not all the papers implemented an audio channel as a feature, they all facilitated verbal communication between users in some way (like having them separated by a screen in the same room) - making verbal communication the primary way collaborators communicated. However, only one paper [348] exploited the spatial audio capabilities of MR for it.
- **Virtual Replicas of Task Objects:** Demonstrating an action and describing how to manipulate 3D objects can be quite difficult during remote collaborations. To help with this, collaborative systems often provide remote users with virtual replicas of physical objects in the task environment that they can use to demonstrate actions with [324]. While most systems facilitated this through the use of pre-made 3D models, a few systems allowed remote users to segment out objects from their reconstruction of the task environment that could then be used to convey technique [355]. While not the primary function, virtual replicas also typically tells the remote user that the local user has access to them – making it an explicit source of information that communicates the presence of at least a subset of task objects in the physical environment.

- **Representation of Hand Gestures:** Given how important the use of our hands are for pointing and communicating object manipulations, researchers captured user's hands and overlaid a representation of them onto their collaborator's workspace whether it was experienced through an immersive medium or one on a screen. Examples of this feature can be found in systems like *HandsIn3D* [136], *BeThere* [279], and *3D Helping Hands* [298]. One paper allowed for the representation of the remote user's hands to be "stretchable" where it was anchored to the local collaborator's shoulders but the fingertips remained where the remote user's fingertips were with respect to the environment [186]. 15 systems represented the captured hands as is (they presented a view of the user's actual hands) using either depth cameras [10] or by filtering out the background from a video feed [138], while others used other 3D models to represent a collaborator's hands [322].
- **Fixed and Freehand Environment Annotations:** Another relatively common feature that is typically offered to remote users is the ability to add annotations to the task environment. These annotations allow them to refer to task objects and instruct local users of what they need to do and can be divided into two broad categories: fixed and flexible. Fixed annotations include predefined content like icons, circles, or text [81, 118, 38], and freehand annotations are those where the user can essentially sketch on the task environment [101, 205]. Of all the systems that included an annotation feature, only 2 gave the local user the ability to annotate in the remote user's collaborative space as well [307, 172].
- **Virtual Pointers:** The ability to refer to objects in the environment is crucial to the success of communication during physical tasks. This is, therefore, an action for which systems have provided dedicated support. Systems typically allow remote users to highlight specific objects in the task environment using visual cues like arrows [32], laser pointers [301, 151], or a representation of the user's VR controllers [300].
- **Highlight or Navigate to Task Object:** Researchers have also explored allowing the

remote expert to highlight specific objects in the physical task environment or initiate navigation cues that can direct local users to those objects. These include being able to add object outlines [33] or labels [41] for the local user in the task environment by clicking on the object in their view of the task environment. Günther et al. [122] also explored audio beacons attached to the task object and haptic cues as alternate navigation mechanisms for objects of interest.

- **System Visualization for Technique:** A system implemented in Oda et al. [224] allowed remote users to annotate key points on a virtual replica after which the local user is automatically shown cues to help with alignment, effectively offloading some of the burden of communicating technique to the collaborative system itself. A few other systems also offload the task of communicating technique to the system by allowing remote users to initiate video clips [101] or 3D animations [64, 32] that the local collaborator could view to understand certain actions or how objects needed to be manipulated.

5.3 The Situated Experience with Current Virtual Task Environments

Collaborative systems that support physical tasks by allowing remote experts to provide MR-mediated guidance can use a variety of technologies to create the remote user interface that makes up the virtual task environment (Fig. 5.7). In our dataset, we observed that close to half the systems opted to use immersive technology for remote users - with 43% using VR headsets [239, 157, 101] and 2.6% using AR/MR devices [251, 49] including projection-based AR [236] and near-eye displays [134]. The rest of the systems used desktop setups [199], mobile phones [279, 235], tablets [12, 184], and large flatscreens [205, 9] for remote users. A few systems (2%) also used other technology like stereoscopic displays [14] and videoscopes [271].

We also analyzed the situated experience remote collaborators could experience with the virtual task environments based on the dimensions in the framework discussed in Chapter 2

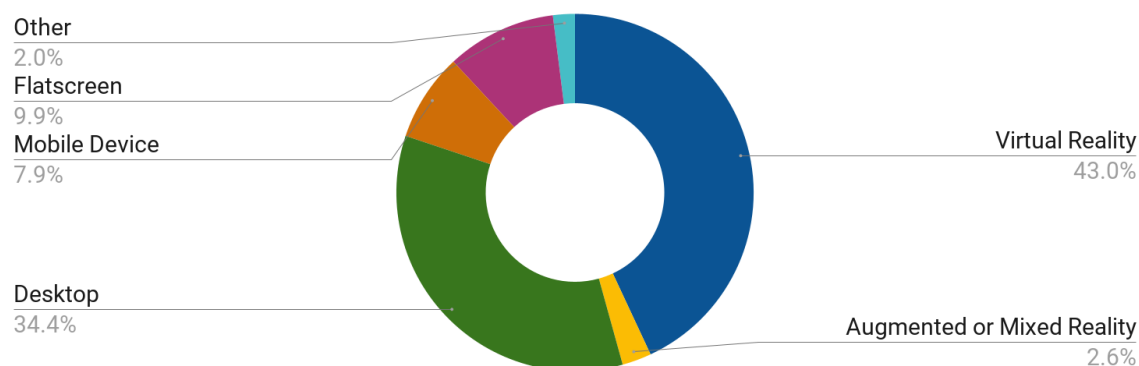


Figure 5.7. Technology used to provide remote users with a virtual task environment.

(Sec. 2.5.2). In the systems we analyzed, the situated experience of the virtual task environment is fundamentally driven by features that provide remote users with a view of the task environment. The coverage is dependent on how much of the task environment these features capture (for example, the coverage for 2D video streams depends on the camera placement in the task environment). The spatial presence is determined by both the coverage and the type of view (for example, in our database, 3D views that only cover part of the task environment presented it to the user in a virtual environment without any other substantial artifact that could impact the spatial presence, and in allowing for an immersive view with depth without a full environment, it provides medium spatial presence). The agency these features provide is dependent on whether they were accompanied by features that allow users to control their perspective (low agency), type of view (medium agency), or scale (high agency). Table 5.2 depicts how the features that provide a view of the task environment dictate the situated experience for remote users. We can see here that the current research for collaborative XR systems have collectively explored all the three possible broad categorizations for these dimensions.

The capability to situate a user in a computer-generated 3D environment is a key component of how immersive technology brings collaborative settings closer to face-to-face collaboration [26, 67]. Room-scale immersive experiences that offer high spatial presence, in particular, create a compelling sense of "being there" and this enhanced perceptual richness increases users'

Table 5.2. The situated experience current collaborative XR systems provide remote users.

Situating Experience of the Remote Virtual Task Environment	Coverage of Task Environment	Active Space	54	First Person View	2D Video Stream	360° Video	3D View	3D View + 360° Video	Differently Scaled 3D View
		Partial Space	75	55	13		7		
		Full Space	50		2	20	19	4	5
	Spatial Presence	Low	99	55	44				
		Medium	59			20	32	4	3
		High	21				19		2
	Agency	Low	20	4	6	3	5	2	
		Medium	16	3	1	6	5	1	
		High	4						4

comprehension of the virtual environment [277]. Users are, therefore, more likely to engage in more natural behaviors [278] and show higher levels of presence [34] and engagement [149] which can be especially beneficial for complex tasks [136].

As we’ve seen, collaborative systems attempt to leverage these benefits of high spatial presence by immersing remote experts in a 3D reconstruction of the local task environment, enabling them to move around the space independently like they would in the real world. These reconstructions have been shown to improve spatial awareness and social presence among remote experts while allowing for greater freedom of movement and independence of perspective [18, 187]. However, we only see 19 instances of 3D views in our dataset that are able to achieve a high level of spatial presence with most of them (32 features) only providing users with a moderate sense of being present in the environment. Moreover, a majority of the features that provide a view of the environment in today’s systems tend to only offer a low level of spatial presence. Here, we briefly dig a little deeper and discuss the tensions that arise with attempting to provide a fully immersive environment that provide remote users with a high degree of the

illusion of being in a physical space.

While 3D reconstructions theoretically allows the remote user to feel like they are perceptually in the physical task environment, they don't always live up to that promise. In practice, 3D reconstructions require specialized hardware and high network bandwidth as they are computationally expensive and difficult to update in real-time [227, 146]. In addition, real-time 3D views have a lot of noise due to systemic errors [156, 177] and more robust depth sensing and computer vision techniques are dependent on the texture and materials of objects in the task environment. For example, shiny objects or thin artifacts are hard to capture [213]. These factors compound and make real-time 3D views difficult to implement and challenging for experts' to glean details from realistic scenarios [101].

One design choice that we've seen collaborative MR systems like 3D Helping Hands [298] and Vishnu [185] make to reduce the computational burden, manage latency, and overcome bandwidth limitations, is to reconstruct just the active space of the task environment. This ensures that experts can experience the benefits of a 3D view for the part of the environment they need the most dynamic updates. However, the reduction in coverage of the task area reduces the degree of spatial presence it offers and hinders the overall effectiveness [300, 190] of the 3D reconstruction (especially because remote users get no information about the peripheral space – or the space outside of the active task space). Needless to say, a recreation of a portion of the environment also wouldn't suffice for room-scale tasks like engine room maintenance or crime scene investigation where the active space is the whole task environment.

Combining pre-scanned or static 3D recreations of task environments with live 3D reconstructions of smaller areas is another approach that reduces computational complexity without making compromises on the extent of immersion experienced. For example, Dou and Fuchs [72] scanned static portions of a room and aligned live data of moving artifacts in the space to create a noise-free 3D reconstruction of the whole environment. The main advantage of this approach is that the static portions can be processed in advance to increase the visual fidelity of the final reconstruction, leading to higher levels of realism and perceived presence

in the virtual environment [334]. Lee et al. [193] combined high-fidelity static objects that are actively worked on during the task with a lower-fidelity real-time updated point cloud of the peripheral space that experts could interact with. While prior research has shown that such a static 3D view is sufficient for remote experts to understand the task environment [100], the disadvantage is that any information represented by the static portions becomes outdated as the task goes on, especially when the locations of objects within it change.

The general difficulties and constraints that come with 3D views have prompted researchers to combine 3D reconstructions with non-3D views of the task environment to provide experts with the benefits of a real-time updated 3D view as well as spatial information from the peripheral space while conserving computational resources. For example, ARTEMIS [101]

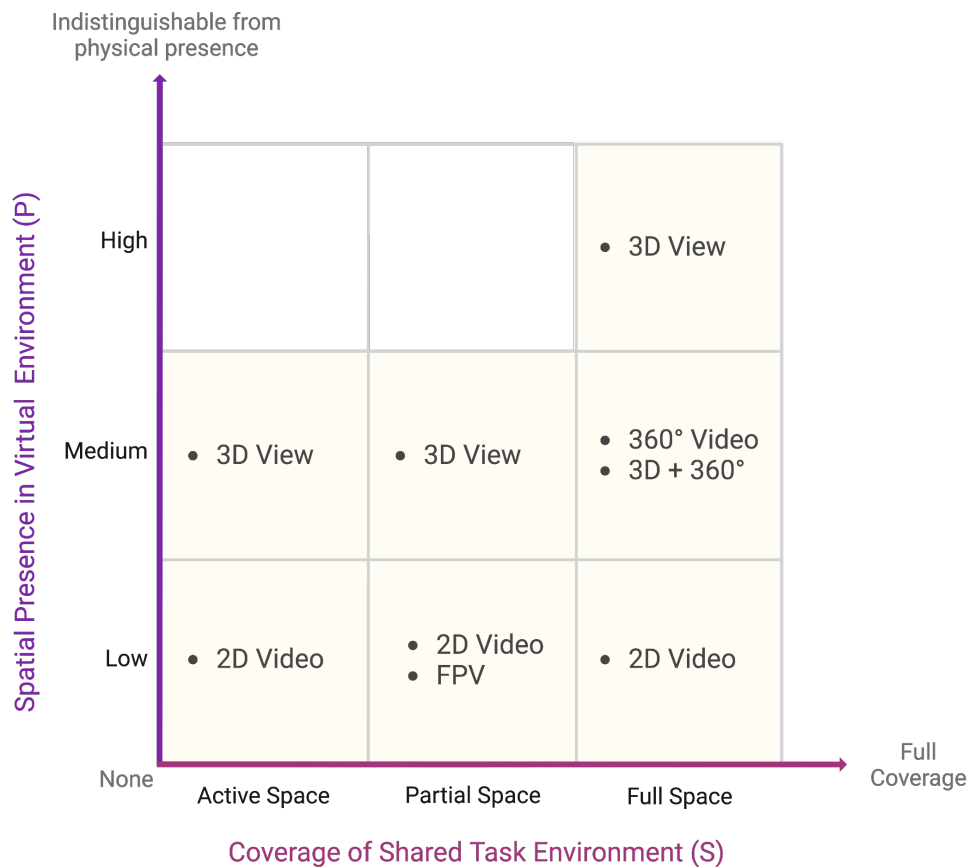


Figure 5.8. Support for referencing with current collaborative XR systems.

provided remote experts with a 3D view of the surgical site on the patient and opted to use 2D video feeds to provide information about the rest of the environment; other systems combined 3D reconstructions with 360 videos [190, 240, 300]. Our review also revealed that most collaborative XR systems simply opt for a purely 2D setup for the expert where they view video streams of a third-person perspective of the task environment, or a first-person view from a camera on the worker's HMD even when immersive technology is used for the remote expert.

A visualization of the coverage of the design space across the spatial presence and the coverage of the shared task space current virtual task environments offer (Fig. 5.8) highlights the tensions between the two dimensions and brings a key pattern in the field into focus: **Collaborative XR systems that support physical tasks have only been able to tap into the benefits of high spatial presence for the remote user when it also offers full coverage of the physical task environment through a view of the environment.**

5.4 Support for Object-Focused Communication Today

Guidance for physical tasks heavily revolves around object-focused communication which includes referring to objects and describing actions to be done on them [93]. However, describing actions to be taken on objects is fundamentally reliant on being able to refer to the objects or relevant portions of it first. The expert's ability to provide guidance and partake in object-focused communication is thus primarily dependent on the ease with which they can refer to the objects used as part of the task they want to accomplish. As a consequence, *referencing* - or the act of making a reference and having that reference understood by others [350, 44] - becomes a crucial process that collaborative systems need to support [44] as the absence of support for it severely hinders communication [73]. It is also the communication process that most ties collaborative physical tasks to the task environment. Armed with a better understanding of the different features collaborative XR systems provide and how they support collaborative physical tasks in general, this section takes a closer look at how these features support referencing.

When collaborators are physically co-present, referencing can seem like a trivial process since they share a rich visual space that facilitates the awareness required to generate these references and provide the context required to understand them [174, 54]. To be effective, remote collaboration systems need to explicitly build features that support initiating and understanding references [174, 78]. The framework for *inter-referential awareness* outlined by Chastine et al. [44] highlights the different aspects of the *referential process* that need to be considered when designing collaborative XR systems. At its core, the referential process between an expert and a novice during collaborative physical tasks is split into two sub-processes that occur sequentially and have varying needs:

1. **Making or initiating a Reference:** This is when the expert refers to an object in the task environment. For example, the expert could say “can you pick up *the red box?*”. Making a reference requires the expert to know *what* objects are present in the task environment. This awareness of the presence of task objects is crucial to the expert’s ability to refer to them, and its absence will cause the referential process to break down. We call this information the *awareness of (task) objects*.
2. **Understanding the Reference:** This is the process by which the novice understands the reference and identifies the object being referred to. Novices during physical tasks are typically unfamiliar with the task and the environment, making it difficult for them to identify an object or tool based on domain-specific names (proper nouns) even if they visually register it in their environment without additional support from experts. Therefore, when experts make a reference, they need to do it in a manner that is conducive to the novice understanding it.

In addition to verbal communication, remote collaborators or experts use various non-verbal gestures and communication mechanisms to refer to the objects required to accomplish a physical task [68]. They typically choose mechanisms that requires the least effort for them to refer to objects and for that reference to be understood by novices [51, 54] –

this means that, given the opportunity, experts will always prefer to use proper nouns, short phrases, alternate descriptions, deictic expressions, and simple gestures (like pointing) to efficiently refer to objects without lengthy verbal descriptions [234, 176]. Deictic expressions, in particular, contain words like “this”, “here”, or “that” that carry contextual meaning and refer to things, people, or locations in a manner that reduces the amount of verbal communication required for it. These expressions are also often combined with gestures such as pointing. For example, experts can point at an object and use phrases like “the blue wrench *to your right*”, “*that one right there*”, or “bring *that one* over here” to direct novices to objects of interest instead of verbally describing what and where the object is.

To employ deixis in the above example, the expert needs to know *where* the object is to point to it – they also need to share enough context with their collaborators for those receiving the reference to understand what “that” means. Therefore, being able to partake in effective mechanisms for referencing typically requires the expert to have extensive awareness of the task environment including where an object is with respect to the novice and other objects in the environment [344]. This *spatial information of (task) objects* – or information about the spatial relations between objects and novices – is fundamental to the experts’ ability to provide object-focused guidance.

Support for Communicative Action

As we’ve seen, today’s collaborative MR systems take advantage of the spatial capabilities of the medium to help remote experts use non-verbal mechanisms for referencing (Table 5.1). These features most commonly allow them to use visual cues to guide the novice’s attention to a referenced object and include capturing and showing the expert’s hand gestures [284, 328, 279, 166, 133, 321, 185, 136, 137], sharing head pointers or eye gaze cues [326, 348, 239, 238], virtual pointers [19, 163, 43, 103, 304], and annotations [228, 3, 307, 101]. However, by imitating how we make references in the real world, these features suffer from the same drawbacks, especially

in cluttered or complex spaces. Specifically, techniques like deictic gestures and expressions used in face-to-face collaborations suffer from referential ambiguity and difficulty referencing occluded or hidden objects. For example, when a remote expert points to a set of objects, there can be ambiguity about which specific object is being referred to, or the object might not be visible to either the expert or the novice.

As we've previously discussed, the specific medium used for collaboration systems can affect communication, as what is possible with one technology might not be possible with another. While supporting communication techniques present in face-to-face settings is typically regarded as gold standard for collaborative systems, exploiting the unique affordances of a medium like MR could help overcome some of the shortcomings of these techniques, potentially improving upon the classic face-to-face paradigm [132]. One of MR's affordance is that it is a spatial medium inherently aware of its environment [26, 60]. Advances in object detection and tracking have allowed researchers to capitalize on this affordance to reduce referential ambiguity by having the system do something that cannot be done in unmediated collocated settings: draw the novice's attention to *exactly* which object is being referred to by highlighting it or providing navigational cues for it. The TAC system [32] allowed experts to highlight specific objects and regions by clicking on them within their video feed which initiated either an arrow or an outline to guide the novice's attention to it. AlphaRead [41] allowed experts to specify or label objects using the video feed of the novice's environment, and Wang et al. [329] proposed a system where task objects are pre-labeled. In both cases, the labelled objects are tracked and automatically highlighted for the novice when the expert refers to them, reducing both ambiguity and the need for experts to constantly keep track of objects in the environment. However, AlphaRead requires initial input from the expert, and the TAC system does not support referencing if task objects are moved. All three systems also require the task objects to be unoccluded and within the expert's field of view for them to make a reference.

While these dedicated features that support referencing increase the efficiency with which experts can refer to objects and the ease with which novices can understand these references

when compared to verbal communication alone, they are all fundamentally proxies to deictic pointing and mimic referencing mechanisms used in collocated settings . Experts’ ability to use these features are, therefore, heavily dependent on their awareness and spatial information about the task objects [293, 345].

Support for Achieving Required Awareness

A look at the features that support environment awareness reveals that experts’ information needs with respect to referencing comes primarily from communication over the verbal channel, or through a view of the task environment (Table 5.1). However, the different types of environment views we’ve seen collaborative systems impact how easy it is for experts to glean the information they need.

A 3D view of the task environment retains the characteristics of how people experience environments in the real world the most. This technique, therefore, provides both the awareness of task objects and the spatial information required to support referencing in a manner closest to collocated settings – especially when experienced through an immersive head-mounted display (HMD) like a VR headset. However, while immersive 3D reconstructions should theoretically be similar to being physically co-present, it is not easy to achieve in practice. First, they are computationally expensive, difficult to update in real-time, and require high network bandwidth and extensive setup with specialized hardware [301, 299, 227, 146]; the quality of these reconstructions are also typically low and/or cover a small area to manage latency, reducing its overall effectiveness [300, 190]. Second, current XR HMDs have limitations in resolution, color depth, and field of view that makes the experience different from being physically co-present [26].

Both these factors mean that 3D reconstructions high-fidelity 3D reconstructions are difficult implement and they are in general challenging for experts’ to glean details from in realistic scenarios [101]. Moreover, the reduction in coverage systems typically opt for to manage its computational resources and latency hinders the overall effectiveness [300, 190] of

3D views as a source of information for referencing as experts would need to rely on verbal communication to get information about the objects or tools outside of the 3D view. While static 3D reconstructions can offer higher levels of visual fidelity, the information it offers becomes outdated as the task goes on, making it challenging for experts to update their awareness when the locations of objects within the task environment changes. Therefore, 3D views in general don't offer extensive support for remote users to achieve the awareness they need for referencing.

We've also seen collaborative systems provide 2D views of the environment either in combination with 3D views or on their own that act as a source of information about the physical task environment. Video feeds including a first-person view from a camera on the worker's HMD are the most common features in this category. While video streams support referencing (including the use of deictic references) and enhance communication efficiency [176, 93, 247], they typically only provide a partial view of the task environment. Objects are often outside the camera's view, making it difficult to build awareness and spatial knowledge [107, 93, 92, 176, 133]. More specifically, scene views that offer coverage of the full task space with 2D video streams provide a good overview, but lack enough detail to identify those objects [92]. FPVs that are locked to the worker's view tend to be very jittery and can make for a very disorienting experience for the expert [133, 293]. To overcome these limitations, some systems use multiple camera feeds [254, 250], but switching between them can be jarring [107], and remote viewers often do not have a good understanding of the layout of the physical space [93]. Researchers have explored 360°-videos as an alternative to overcome the drawbacks of video feeds without the complexity of 3D reconstructions [190, 240, 281]. However, while they overcome the field-of-view limitations, 360°-videos lack detail (like the scene views) and are still a 2D presentation with limited depth perception [301].

In general, and perhaps more importantly, any view of the environment is a passive source of information that only provides partial and incomplete support for experts to understand the presence and spatial layout of the objects in the worker's environment. This means that experts can struggle to comprehend details and object locations (something that our research in chapter

4 confirms) without extensive verbal communication and interactions with the novice. This is especially true in dense spaces where objects are difficult to make out or occluded from view [16]. Overall, today’s collaborative XR systems for remote guidance do not explicitly support the information needs for referencing.

When it comes to supporting experts need for awareness about task objects for referencing through explicit representations Table 5.1 shows us that few systems provide helpers with information about what task objects are present in the worker’s environment through virtual replicas of the objects [293, 286, 259, 321], but there is a lack of systems that provide spatial information of task objects in a manner that is not tied to the view of the task environment.

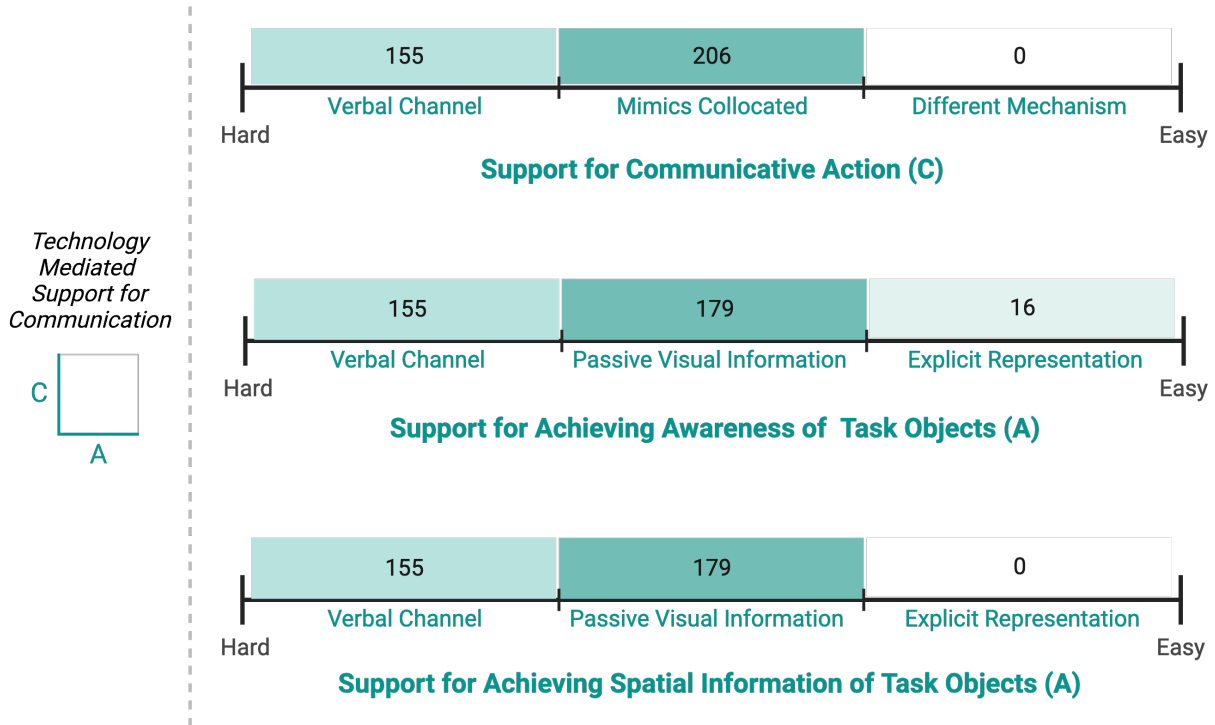


Figure 5.9. Support for referencing with current collaborative XR systems.

5.5 Conclusion

This chapter analyzed the existing collaborative XR systems that support physical tasks and contextualized this analysis based on the conceptual framework introduced in Chapter 2 to

understand the current design landscape and provide an overview of how the different features we see support the communication required for physical tasks. A visualization of the design space these features occupy (Table 5.1) showcases how the different XR capabilities support collaborators' communication needs. We also discuss the situated experience remote interfaces offer today, and take a closer look at how process of referencing is supported. In doing so, we recognize that despite exploring various levels of situated experiences for remote experts (Table 5.2), collaborative systems have still not maximized the support offered for referencing (Fig. 5.9). Moreover, these systems have only been able to tap into the benefits of having a fully immersive, room-scale environment for remote users when it also offers room-scale views of the environment (Fig. 5.8). The next two chapters in this dissertation aim to explore design options that fill these key gaps in the field and evaluates their impact on the collaborative experience for both experts and novices through controlled experiments.

Acknowledgements

The contents of this chapter, in part, are currently being prepared for submission for publication by Janet G. Johnson, Nadir Weibel, and Mark Billingham. The dissertation author is the primary investigator and author of this material.

Chapter 6

Enhancing Support for Referencing During Physical Tasks

Thesis Question: How can we leverage the spatial capabilities of XR to enhance support for referencing task objects during collaborative physical tasks?

Conversations during remote guidance for physical tasks are heavily anchored around task objects and their spatial relationships in the real world, making referencing - the ability to refer to an object in a way that is understood by others - a crucial process that warrants explicit support in collaborative Mixed Reality systems. This chapter presents a 2x2 mixed factorial experiment that explores the effects of providing system-generated guidance to task objects for local collaborators, and explicit representations for spatial information for remote collaborators. It also investigates the effects of such guidance on the remote expert's need for spatial information. Our results show that guidance increases performance and communication efficiency while reducing the need for spatial information, especially in unfamiliar environments. Our results also demonstrate a reduced need for remote experts to be in immersive environments, making guidance more scalable, and expertise more accessible.

6.1 Introduction and Motivation

Making a reference from a distance requires remote collaborators to know *what* objects are present in the task environment and *where* in the environment they are. The awareness of task objects and their spatial relationships within the task environment is therefore crucial to the ability to refer to them, and absence of this information will cause the referential process to break down. However, despite recent advances in the field, domain- and technology-specific constraints

(like limited bandwidth, field-of-view limitations, and ergonomic issues) makes it impossible for a collaborative MR system to provide all the information required to maintain the effectiveness of remote communication at the level of face-to-face interactions [174, 26, 93]. Addressing the resulting misunderstandings typically requires extensive verbal interaction [344, 345]. Our research in chapter 4 highlighted the difficulties of this for remote experts providing guidance for physical tasks and chapter 5 outlined key gaps in the existing support for referencing. Passive sources of information like a view of the environment, in particular, are typically computationally expensive to provide or don't provide an optimal level of support for the information needs for referencing, or both. The dimensions of support for communication (Fig. 5.9) point to two ways we could enhance support for referencing: The first would be to make it easier for experts to achieve the awareness required for it through an explicit representation of information, and the second would be to explore mechanisms that don't mimic those used in collocated settings.

Now, the referential process includes two sub-processes – making the reference, and understanding the reference (discussed in 5.4). During physical task guidance, it is important that the remote expert in particular has the information they need to be able to refer to task objects and guide the novice through the task at hand. However, when we look more closely, *what* object to refer to is part of their domain expertise, and they only need to know *where* the object is to be able to guide the novice's attention to it. In other words, in both face-to-face communication and traditional collaborative systems, the responsibility of referring to objects in a manner that ensures that the worker can identify them falls on the expert.

One key spatial characteristic inherent with XR is the environment awareness it comes with. Spatial coordinates of physical objects and people can be retrieved through various techniques like local sensors, indoor GPS, tags, markers, computer vision, RFID tags, etc. These tracking systems are also becoming increasingly pervasive, low cost, and accurate. For example, Tait and Billinghurst [293] use the ART system¹, Wang et al. [329] use a stereo camera and

¹<https://ar-tracking.com>

markers, and ARTEMIS [333] uses the OptiTrack motion capture system² to track the positions of different task objects as well as the collaborators. In this chapter, we leverage this environment understanding to present an explicit representation of the spatial relations between the task objects and the novice to the remote expert. In this chapter, we focus on a 2D representation as it is compatible with both immersive and non-immersive helper setups, and evaluate whether such a representation could reduce the information processing or comprehension needed to achieve spatial awareness.

Systems that use live tracking technologies could also automatically provide visual MR cues to guide novices to referenced objects without the expert having to expend effort towards crafting their reference in a way that would be easily understood. This could potentially *offload* the responsibility of the second sub-process of referencing (understanding the reference) from the expert to the system itself – allowing for a communication mechanism that doesn’t fully mimic how references are made in the physical world. While this could make it substantially easier for experts to provide guidance, collaboration is a complex process – changing the role they play in the referential process could introduce significant cognitive seams [140] or discontinuities in the experience. Therefore, in this chapter, we aim to understand the effects of partially offloading the referential process on remote guidance for physical tasks using system guidance to referenced objects.

More specifically, we explore the effects of both explicitly providing spatial information and offloading the referential process with MR guidance through a mixed factorial study. In doing so, this chapter contributes:

1. An evaluation of how *explicitly representing spatial information* supports referencing for physical tasks.
2. A discussion of how visual guidance in Mixed Reality can *offload the referential process* and reduce collaborators’ efforts; and an empirical understanding of how this impacts the

²<https://optitrack.com>

remote collaborator’s needs for spatial information.

3. A discussion of the implications for designers and developers, including opportunities for scaling up Mixed Reality collaborative systems for remote guidance.

Since this chapter presents a controlled experiment that does not involve a domain task where the remote collaborator has pre-existing expertise, we refer to the remote collaborator who plays the role of the expert as the *helper*, and the local collaborator who plays the role of the novice as the *worker*.

6.2 Research Approach

This chapter explores the two opportunities to enhance referencing in MR-mediated physical tasks: explicitly providing the spatial information necessary for helpers to direct workers to task objects, and offloading this responsibility from the helper to the system itself using automatic visual guidance to direct workers to referenced objects. However, the helper’s need for spatial information is primarily required to direct workers to the referenced object - typically using strategies like deixis that reduce the collaborative effort required for the worker to understand the reference and identify the task object [51, 54]. Given this, *do collaborative MR systems that provide system-generated guidance for workers still need to explicitly provide helpers with spatial information?*

Our goal in this chapter is to evaluate this question in addition to exploring the opportunities to enhance referencing in collaborative MR systems. To do so, we built a prototype *2D map interface* that explicitly represents the spatial relationships of the task objects and the worker in the task environment. We also implemented *MR system guidance* to dynamically direct the worker to referenced objects as a way of offloading part of the referential process. To understand the effects of both features independently, as well as study the interplay between them, we conducted a 2x2 mixed factorial experiment. Figure 7.5 outlines our experimental approach and answers the following three research questions:

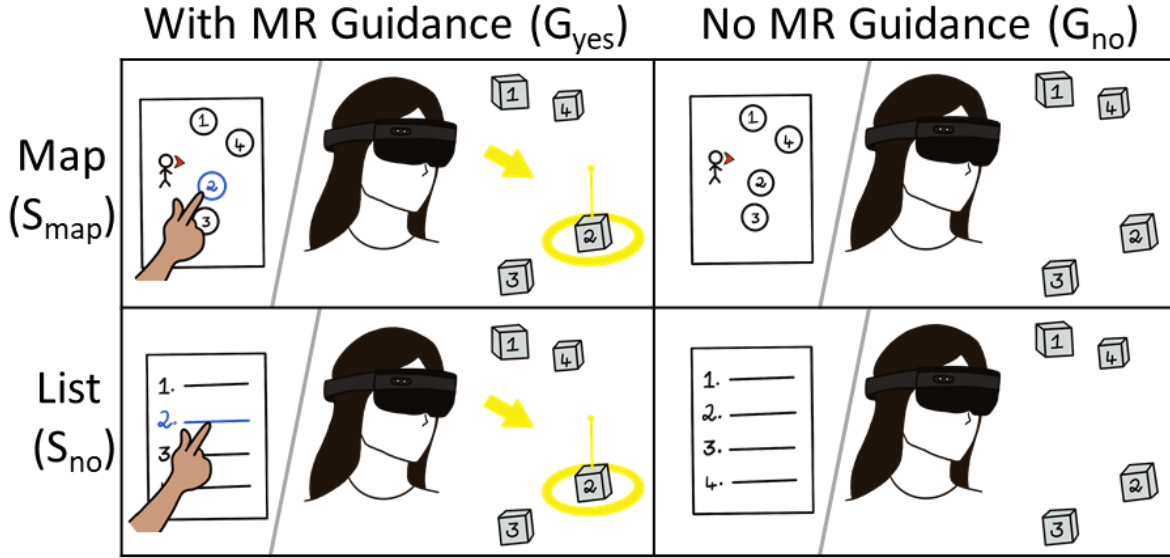


Figure 6.1. 2x2 mixed factorial study design. Our experiment studies the effect of Mixed Reality guidance for workers (G_{yes} vs. G_{no}) and the use of two different helper aids: a Map (S_{map}) or a List (S_{list}).

- **RQ1:** In collaborative mixed reality systems designed for remote guidance, does a 2D representation of the spatial relationships within the task environment provide the helper with enough information to support referencing?
- **RQ2:** Does a collaborative mixed reality system that directly implements visual guidance to objects referenced during a physical task enhance task performance and reduce the overall effort required for remote guidance?
- **RQ3:** How does the presence of system generated visual cues to referenced objects impact the helper's need for spatial knowledge when providing remote guidance for a physical task?

The overall goal is to better understand how to streamline support for referencing during physical tasks in MR environments. This knowledge is essential to design future collaborative systems with more effective representations, especially when deploying complex immersive MR systems for both helpers and workers is not an option.

6.3 Prototype Design and Implementation

To answer our questions, we implemented a prototype collaborative MR system that accommodates the different features required to run the experiment presented. When using our prototype system, the workers (see Fig. 6.2b) wear a mixed reality headset (the Microsoft HoloLens v1)³ running a custom-built application that allows them to receive *system guidance* if enabled. The helpers (see Fig. 6.2a) use a stationary workstation with a Windows desktop that shows a live feed of the *worker's first-person view*; helpers can also view a *map* or *list* interface on a tablet device (map and list are different conditions in the study) that provides additional information on the objects in the worker's environment. The *map* and *list* interface also allow the helper to initiate *system guidance* if enabled.

6.3.1 Worker's First-Person View

The prototype system provides the helper with a view of the task environment through the worker's First Person View (FPV) shown on the desktop screen. In doing so, it passively supports referencing and provides the helper with general task awareness [73]. This was implemented using a low-latency live feed of the view from the HoloLens' front camera⁴ and Microsoft's Windows' support for Miracast⁵ on both Windows Desktop computers and the HoloLens. The video feed was transmitted over the local area network and presented almost unnoticeable latency (180-210ms).

6.3.2 A Map of Task Objects

In addition to the passive support provided by the FPV, we needed to explicitly support the helpers' need for task object awareness and spatial information. We do so with a *map interface* that provides the helper with a 2D representation of the spatial relations of task objects

³<https://docs.microsoft.com/en-us/hololens/hololens1-hardware>

⁴<https://docs.microsoft.com/en-us/windows/mixed-reality/locatable-camera>

⁵<https://support.microsoft.com/en-us/help/15053/windows-8-project-wireless-screen-miracast>

in the worker's environment. As shown in Fig. 6.4a, the map interface resembles a top-view of the workspace and indicates the live positions of the task objects. It also represents the live position of the worker with respect to these objects. The task objects are represented by icons and allow the helper to know *what* objects are in the task environment, as well as *where* they are with respect to each other and the worker.

The map interface is implemented as a simple web application that can be viewed through a browser on any device. In our experiment, the helper viewed the map on an Apple iPad (See tablet in Fig. 6.2a). To show the location of both the HoloLens and the objects in the worker's space, the web application connects through a web server to a custom Windows application running on the HoloLens that relies on HoloLens's inside-out tracking system to locate objects⁶.

While the inside-out tracking updates the HoloLens's position in space over time, it does not update the location of the task objects. To get around this constraint, we set up the initial locations for the task objects through our HoloLens app and created a *Wizard-of-Oz*

⁶<https://docs.microsoft.com/en-us/windows/mixed-reality/enthusiast-guide/tracking-system>

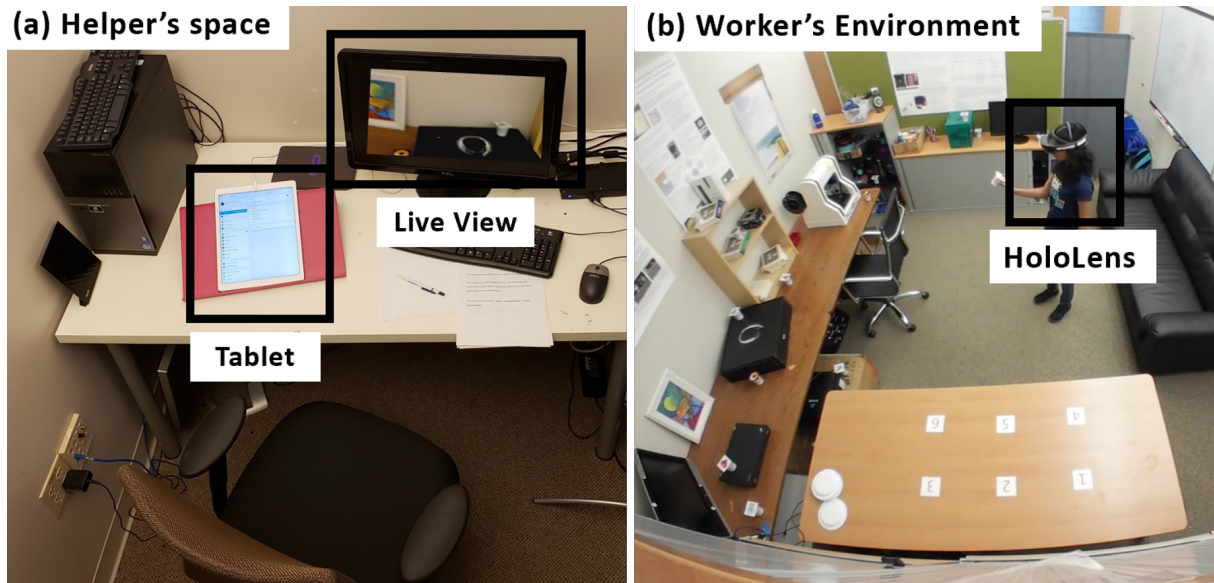


Figure 6.2. Helper's space and worker's environment. (a) The helper has a live first-person view of the worker throughout the experiment, and a specific *aid* (map or list) on a tablet. (b) The worker wears a HoloLens throughout the experiment, regardless of whether guidance is present or not.



Figure 6.3. First-person view of the worker as they are guided to an object by the MR guidance interface [59, 74] that could remotely update the locations of these objects on the map. The interface resembled the map interface. During the experiment, our wizard updated the position of an object when it was moved by clicking on its icon and then clicking on the map at the new position in the worker’s environment.

6.3.3 A List of Task Objects

The map interface explicitly provides the helper with information about (1) *what* objects are present within the task environment, and (2) *where* these objects are. However, to study the effects of the spatial information provided by the map, we needed an alternative (control) interface that only provided information about what objects were available in the task environment but did not show their positions. With this in mind, we developed a simple *list* interface (Fig. 6.4b) that shows the objects in the worker’s environment as a randomly ordered list. The helper was able to scroll through the list if it did not fit on a single page.

6.3.4 MR System Guidance through the HoloLens

To study the effect of partially offloading the referential process to the MR device, we implemented the *MR system guidance* feature. When enabled, the helper could click on the icons representing a task object on either the map or the list interface to guide the worker towards it (during the experiment, helpers only had access to one of the two interfaces at a time). Once clicked, the system (HoloLens app) dynamically directed the worker’s attention to the referenced



Figure 6.4. Helper’s aids. **(a)** *Map* showing the task objects and the worker’s position. **(b)** *List* showing all the task objects.

object.

While there are many possible visual cues to guide users to a particular location in space [27, 148], our paper does not focus on the efficacy of these methods. However, given that the effects of MR guidance cannot be separated from the influence of the visual cue itself, we needed to implement one that would not add additional confounding factors to the study. We chose to adapt the navigation technique presented in HoloCPR [148] since was specifically designed for the limited field of view of the HoloLens and proved to be intuitive. The cue included an arrow to provide guidance on which direction to look, and a circle with a vertical element to help the worker focus on the exact object (see Fig. 6.3).

6.4 Experiment

Our experiment followed a 2x2 mixed factorial design, with a between-subjects factor focused on the *presence of MR system guidance* on two levels: G_{yes} = MR guidance present, and G_{no} = no MR guidance present. The within-subjects factor was the *presence of spatial*

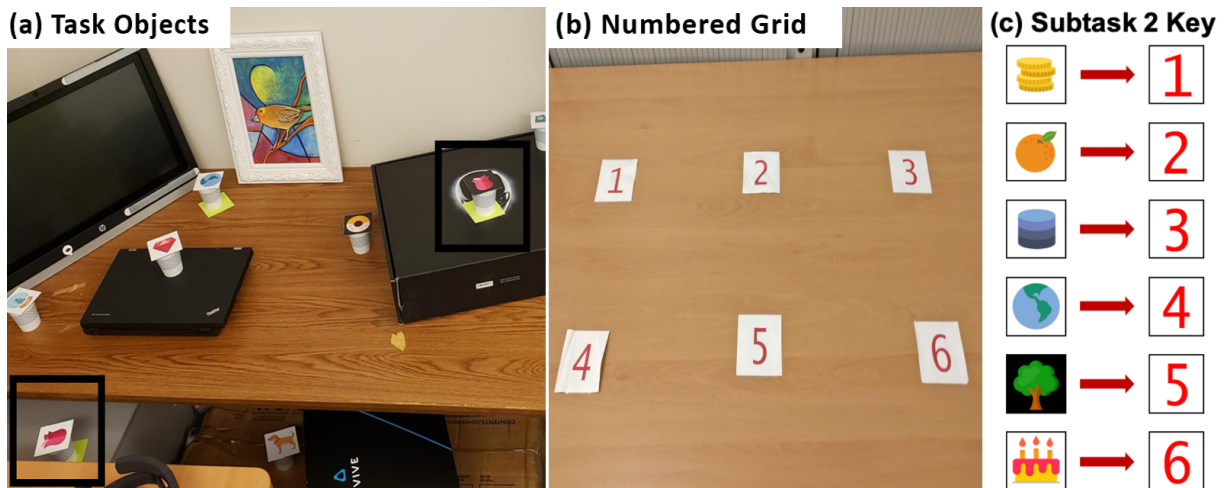


Figure 6.5. Snapshots of the worker’s environment. (a) Some of the 19 task objects. The two highlighted objects are examples of objects used in subtask 1 (pair matching). (b) Numbered grid and (c) solution key used in subtask 2.

information on two levels: S_{map} = map interface (spatial information present), and S_{list} = list interface (no spatial information present). We chose the presence of MR guidance to be a between-subjects factor since its carryover effects were unclear. However, the change in task objects and their positions between tasks ensured minimal carryover effects for the presence of spatial information when counterbalanced, allowing us to make it a within-subjects factor. The view of the task environment (provided by the FPV) and the awareness of the task objects available (provided by either the map or the list interface) were kept constant and represented our independent variables. Fig. 7.5 depicts our study design.

6.4.1 Participants and Procedure

Forty participants (17 females, 22 males, 1 preferred not to say) aged 18 to 43 ($M = 23.93$, $SD = 5.17$), were recruited to form 20 worker-helper dyads. 8 of these dyads knew each other beforehand. Most participants were students, with the majority majoring in Computer Science and Engineering (23) and others majoring in Electrical Engineering, Cognitive Science, Linguistics, Speculative design, Economics, Chemistry, and Bio-engineering. Despite a majority of computer science students, only 30% (12 people, 6 selected to be workers) had previous

experience with Microsoft HoloLens.

Each dyad was required to complete two tasks and were randomly assigned to either receive MR guidance (G_{yes}) or not (G_{no}). All dyads were given the map (S_{map}) and the list (S_{list}), one at a time, over the two tasks (the order was counterbalanced). Helper and worker roles remained constant between the two tasks.

Once the participants were welcomed and introduced to the study, they were asked to fill a simple demographics questionnaire. They then underwent a short 5-minute training session so the worker could get accustomed to the HoloLens, and the helper could get accustomed to the video feed with the first person view (FPV). They were not introduced to either the map or the list interface in this session.

After the training session, the helper was taken to the helper space to be introduced to the task at hand and was provided an instruction sheet. The helper also received an *aid* which was either the map or the list interface on the tablet. The worker, who at this point, was unaware of the task was taken to the worker space and asked to wear the HoloLens. At the end of the first task, both the helper and the worker were asked to fill out a short post-task questionnaire and the NASA-TLX questionnaire [125]. This procedure was repeated for the second task with the respective conditions. The entire experiment was also video recorded. Participants consented to be in the study as per our university's human research protection protocol and received a \$5 gift card.

6.4.2 Experiment Task and Environment

The helper and worker spaces were physically separated (see Fig. 6.2), but they were adjacent so that the helper and the worker could talk to each other. In the remainder of this section, we will refer to the worker's local space as the *task environment*. This task environment was a room that had a table with a numbered grid (see Fig. 6.5b) and 19 task objects (see Fig. 6.5a) that were spatially distributed. Each object had a picture on it with either a white or black background.

To complete each subtask, workers needed to be guided by the helper to accomplish a certain goal. The helper could use the *aid* (map or list) provided to them on the tablet as well as a live first-person view of the worker (see Fig. 6.2a). The worker wore a HoloLens at all times regardless of the study condition so that the helper could see their view (see Fig. 6.2b).

The entire task was designed to take less than 5 minutes and each dyad performed the task twice. However, the task objects and their positions in the task space were changed before the second run. There was no overlap in the set of task objects between the two runs of the task for the experiment. The task consisted of two subtasks that needed to be completed in order (sequentially):

1. **Subtask 1 - Pair Matching:** In the first subtask, collaborators had to find pairs (objects with the same picture on them) among the objects in the task space and put them together. Each pair consisted of task objects with the same picture on them but one had a black background and one had a white background. They were instructed to move the object with the white background and place it adjacent to the object with the black background. Of all the objects in the task space, only 4 pairs could be made.
2. **Subtask 2 - Object Gathering** The second task made use of the grid placed in the task space. Helpers were given a solution key (see Fig. 6.5c) that they used to help the worker find certain objects in the task space and place them on their respective positions on the grid.

While both subtasks required workers to identify particular objects in the environment and take actions on them, the actions were kept simple to allow for our measures to primarily reflect the effects of the prototype features on the referential process. The subtasks were also designed to mimic referencing in real-world tasks – Subtask 1 is similar to tasks like engine room maintenance where referents (tools and area of interest) are spread out over a wide area, while subtask 2 is similar to car repair or surgery where the tools are laid out in the task environment but the area of interest (engine part, surgical site) is in a small area like the grid.

6.4.3 Experiment Measures and Analysis

We recorded the worker’s FPV, and a view of the task environment using an external camera. The conversation between the participants during the experiment was also transcribed. We measured task performance (completion time) and analyzed communication behaviors (efficiency of communication and helper’s referencing behaviors) by coding the recorded videos and conversation transcripts using BORIS [90]. Two researchers independently coded the data (double coded 20% of it, Cohen’s Kappa = 0.886), and discussed and resolved any inconsistencies as soon as they arose. We also measured the *task load* on both the helper and the worker for each subtask they performed using the NASA-TLX questionnaire [125]. The post-task questionnaire measured the participants’ perceived success of the partnership and the importance of the specific aid (map or list) through Likert-scale questions. Participants also provided free-form feedback on the technology they used during each task in the post-task questionnaire.

6.4.4 Limitations of our study

A few aspects of our setup are important to describe to understand the possible limitations of our results. First, being a 2D interface representing a 3D space, the map showed overlapping object icons in areas where the density of task objects was high. While this did not limit our participants (they could zoom in and click), this representation of task objects is fundamentally limited by the screen’s real estate and 2D factor. Second, the lack of 3D spatial information meant that helpers sometimes described objects as next to each other when in reality they were at different heights (like on multiple levels of a shelf). However, workers adapted to it by looking up and down when searching for objects. Third, information about the task objects (map or list interface) and the FPV were on separate screens – this could have made it harder to pay attention to both at once. In general, these limitations could make the insights in this paper less generalizable for 3D representations of spatial information within more immersive environments.

Moreover, we did not observe any differences in communication behaviors because of

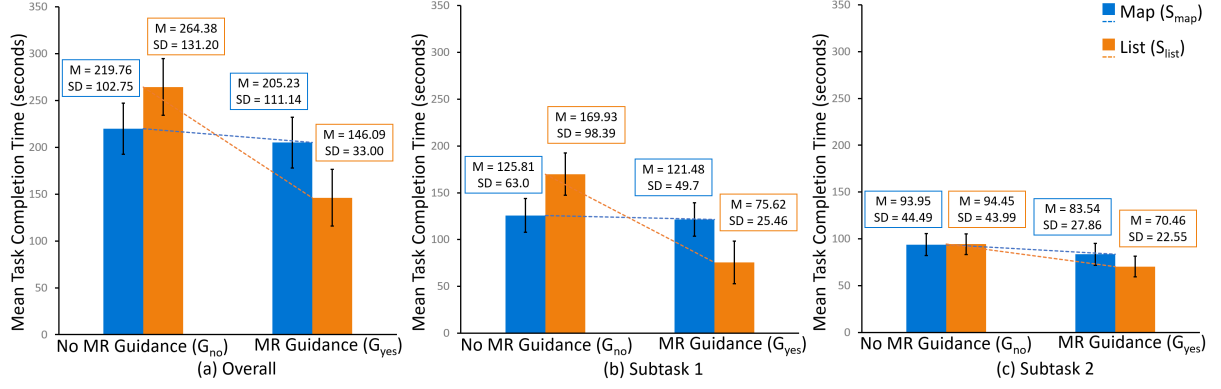


Figure 6.6. Average completion time across participants and conditions. (a) Total completion time across both subtasks. (b) Completion time for Subtask1. (c) Completion time for Subtask2.

prior familiarity among dyads - we believe this could be because of the simplicity of our task and might not reflect the effect of familiarity among collaborators for all tasks.

6.5 Results

The quantitative data collected in our experiment was analyzed using two-way mixed-model ANOVA. All our measures were normally distributed across each cell of the design except for the data from the post-task questionnaire where the nonparametric Aligned Rank Transform [343] was applied which enables the use of ANOVA after alignment and ranking. Our analysis uses an alpha of 0.05, and presents both the main effects and the interaction effects. A main effect is the statistical relationship (effect) between one factor and a measure, averaging across the levels of the other factor. In our experiment, the main effects would tell us how spatial knowledge and guidance independently affected the collaboration. An interaction effect is present when the effect of one factor on a measure depends on the level of the other factor. For our purposes, we see an interaction effect if the presence of guidance changes the effect of spatial knowledge and vice-versa.

In the remainder of this section, F represents the F-value, p the p-value, η_p^2 the effect size, M the mean, and SD the standard deviation. The error bars in Fig. 6.6, 6.7, 7.7, and 6.9 depict standard errors and we also include trend lines to make it easier to read interaction effects.

6.5.1 Task Performance

We calculated the total time it took participants to complete the full task as well as the two subtasks. We consider a lower task completion time to indicate higher performance. Fig. 6.6 shows results of task performance measurements.

Total Task Time – Participants were significantly faster in finding referenced objects with MR guidance ($F(1,18) = 5.162$, $p = 0.036$, $\eta_p^2 = 0.223$). The presence of spatial information with the map did not substantially alter task performance and there was no main effect observed ($F(1,18) = 0.064$, $p = 0.803$, $\eta_p^2 = 0.004$). Fig. 6.6a shows a crossover interaction where the effects of spatial information seem to be different depending on whether participants had MR guidance or not, but this difference was not statistically significant ($F(1,18) = 3.353$, $p = 0.084$, $\eta_p^2 = 0.157$).

Subtask 1 – In the first subtask, participants completed the task significantly faster in the presence of MR guidance ($F(1,18) = 4.914$, $p = 0.04$, $\eta_p^2 = 0.214$). No main effect was observed for spatial knowledge ($F(1,18) = 0.002$, $p = 0.963$, $\eta_p^2 = 0$). Participants with the list interface were faster in the presence of guidance, while those with the map interface were faster without guidance (see Fig. 6.6b). Overall, they were fastest when using the list with MR system guidance and the observed crossover interaction was statistically significant ($F(1,18) = 5.896$, $p = 0.026$, $\eta_p^2 = 0.247$).

Subtask 2 – There was no significant main effect observed for spatial knowledge ($F(1,18) = 0.291$, $p = 0.596$, $\eta_p^2 = 0.016$) or for system guidance ($F(1,18) = 2.381$, $p = 0.140$, $\eta_p^2 = 0.117$) in the second subtask. We also did not observe any interaction effect ($F(1,18) = 0.340$, $p = 0.567$, $\eta_p^2 = 0.019$).

6.5.2 Efficiency of Communication

We measured communication efficiency by the number of words used to communicate during a task [93]. The fewer words spoken by the speaker(s), the more efficient the collaboration is considered.

Without MR guidance, participants used an average of 636.7 words (SD = 257.34) per task with a map and 555.4 words (SD = 242.73) with a list. The presence of MR guidance significantly reduced this to 405.4 words (SD = 210.55) with the map interface and 319.1 words (SD = 154.74) with the list ($F(1,18) = 9.950$, $p = 0.005$, $\eta_p^2 = 0.356$). The presence of spatial information did not significantly change the communication efficiency ($F(1,18) = 1.681$, $p = 0.211$, $\eta_p^2 = 0.085$) and there was no observed interaction effect.

Analyzing just the words spoken by the helper shows similar results (Fig. 6.7, Top), with the presence of guidance significantly reducing the number of words spoken, indicating a higher communication efficiency ($F(1,18) = 8.623$, $p = 0.009$, $\eta_p^2 = 0.324$). Efficiency decreased when helpers used the map, but this difference was not significant ($F(1,18) = 3.371$, $p = 0.083$, $\eta_p^2 = 0.003$). When it came to workers (Fig. 6.7, Bottom), there was no observable main effect for both guidance and spatial information. No interaction effects were observed across any speaker combinations.

6.5.3 Helper's Referencing Behaviors

We also analyzed how helpers made references to the task objects or locations in the workers' task space. We first discuss their referencing patterns as a whole, and then look into references to task object and spatial references separately for further analysis. Table 6.1 outlines results of our analysis of referencing behavior.

In general, helpers made significantly more references when they used the map interface ($F(1,18) = 17.674$, $p = 0.001$, $\eta_p^2 = 0.495$). They also made significantly fewer references in the presence of system guidance ($F(1,18) = 5.646$, $p = 0.029$, $\eta_p^2 = 0.239$). There was no interaction effect.

Referring to Task Objects – When helpers verbally referred to the objects in the task environment, they either used verbal descriptions of the object like “the pink flower with the white background”, or phrases that used deictic pronouns like “*that one there*”. Deictic pronouns not used to refer to objects (like “*that's cool*”) were not considered in this analysis, nor were

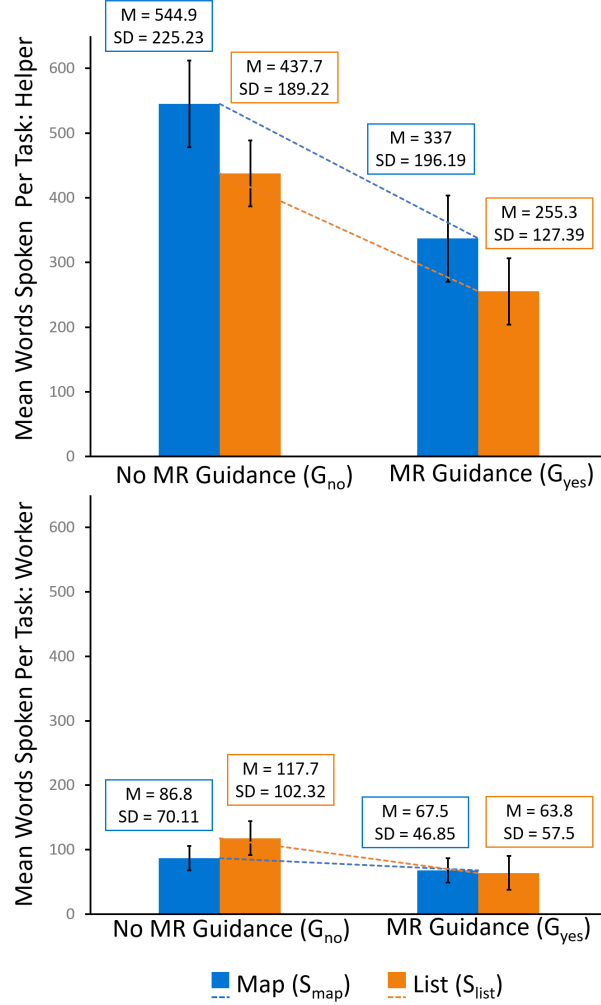


Figure 6.7. Top: Mean number of words per task by helper; Bottom: Mean number of words per task by worker.

anaphoric uses of these pronouns (like “a globe or something like *that*”).

Overall, helpers made significantly more references to objects in the worker’s environment when they used the map than when they used the list interface ($F(1,18) = 6.001$, $p = 0.025$, $\eta_p^2 = 0.25$). The interface they used did not vary the percentage of object references that were deictic pronouns ($F(1,18) = 0.79$, $p = 0.782$, $\eta_p^2 = 0.004$). Helpers made significantly fewer verbal references to objects in the presence of system guidance ($F(1,18) = 4.946$, $p = 0.039$, $\eta_p^2 = 0.216$). However, the percentage of these references that used deictic pronouns were significantly greater with MR guidance ($F(1,18) = 7.249$, $p = 0.015$, $\eta_p^2 = 0.287$). No interaction effect was

Table 6.1. Observed referencing behavior across the presence or absence of MR guidance (G_{yes} vs. G_{no}) and the different aids (S_{map} vs. S_{list}) with significant effects in green. The rightmost column shows interaction effects.

	G_{yes}		G_{no}	S_{map}		S_{list}	$G \times S$
Total Helper References	28.75	<	40.2	42	>	26.95	p=0.078
	p = 0.029			p = 0.001			
Total Task Object References	24.5	<	31.4	30.85	>	25.05	p=0.168
	p = 0.039			p = 0.025			
% of Deictic Pronouns	25.16%	>	7.1%	16.75%	>	15.5%	p=0.54
	p = 0.015			p = 0.782			
Total Spatial References	4.25	<	8.8	11.15	>	1.9	p=0.099
	p = 0.089			p < 0.001			
% of Deictic Spatial Expressions	45.5%	<	61.56%	78.78%	>	28.28%	p=0.64
	p = 0.095			p = 0.001			

observed.

Spatial References – When providing remote guidance to workers, helpers often verbally directed them towards objects of interest. When they did so, they either directly referenced areas or locations in the task environment (like “it should be on the *shelf*”), or used phrases with deictic spatial expressions like (“it’s to *your right*”).

Helpers made more spatial references when they used the map interface, and this difference was highly statistically significant ($F(1,18) = 23.137$, $p < 0.001$, $\eta_p^2 = 0.562$). The presence of spatial information from the map also caused a significant portion of these references to be made using different deictic spatial expressions ($F(1,18) = 16.264$, $p = 0.001$, $\eta_p^2 = 0.475$). The presence of system guidance did not significantly affect both how many spatial references helpers made ($F(1,18) = 3.231$, $p = 0.089$, $\eta_p^2 = 0.152$), and what percentage of these references used deixis ($F(1,18) = 3.105$, $p = 0.095$, $\eta_p^2 = 0.147$). There were also no interaction effects.

6.5.4 Collaborator Task Load

We measured the task load on both the helper and the worker for each subtask they performed using the NASA TLX questionnaire [125] (see Fig. 7.7). The task load on the helper during the task significantly reduced both in the presence of system guidance ($F(1,18) = 6.024$, $p = 0.025$, $\eta_p^2 = 0.251$) as well as when they used the map interface ($F(1,18) = 4.230$, $p = 0.05$, $\eta_p^2 = 0.19$). The worker's task load was slightly higher when the helper used the map interface ($F(1,18) = 0.129$, $p = 0.724$, $\eta_p^2 = 0.007$), and lower in the presence of system guidance ($F(1,18) = 1.296$, $p = 0.270$, $\eta_p^2 = 0.067$) but neither difference was significant. There was no significant interaction effect observed for either the helper or the worker.

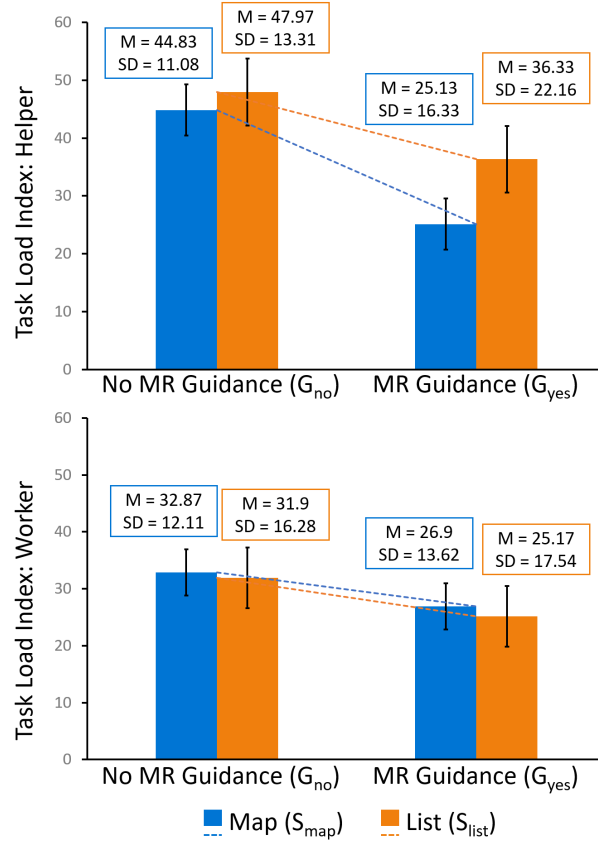


Figure 6.8. Results of the NASA TLX questionnaire. Top: Task load on helpers; Bottom: Task load on workers.

6.5.5 Participant Perceptions

After each task we measured participants' perceived success of the partnership and the role of the helper's aids (Map or List) using targeted 5-point Likert scale questions. Reliability of our questionnaire was measured using Cronbach's alpha [57], and scored high with $\alpha=0.913$.

Working Together – Helper's ratings of the perceived success of their partnership with the worker was evaluated using the Likert-scale question “My partner and I worked well together on this task” (Q1). Responses were rated on a 5-point scale from 1 (= Strongly disagree) to 5 (= Strongly agree). Ratings were consistently high ($M = 4.75$, $SD = 0.543$) across all conditions and there were no significant main or interaction effects.

It is, however, worth noting that helpers perceived the partnership to be more successful when they used a map (S_{map} : $M = 4.8$, $SD = 0.41$; S_{list} : $M = 4.7$, $SD = 0.66$) ($F(1,36) = 0.003$, $p = 0.953$, $\eta_p^2 = 0.00$), and workers perceived it to be better when the helper used the list interface (S_{map} : $M = 4.65$, $SD = 0.93$; S_{list} : $M = 4.85$, $SD = 0.37$) ($F(1,36) = 0.006$, $p = 0.938$, $\eta_p^2 = 0$).

Perceived Importance of Spatial Information Aids – The role and usefulness of the specific Map or List aid for the helper was evaluated using the following two Likert-scale questions: “How important was the aid to help you **collaborate** effectively?” (Q2), and “How important was the aid in helping you **provide assistance** to your partner?” (Q3). Responses were rated on a 5-point scale from 1 (= Not at all important) to 5 (= Extremely important).

Overall, helpers perceived the map interface to be significantly more important for general collaboration (Q2) than the list interface ($F(1,36) = 6.629$, $p = 0.014$, $\eta_p^2 = 0.156$). When asked specifically about how important the aid was for their ability to provide remote guidance (Q3), helpers still perceived the map interface as significantly more useful than the list interface ($F(1,36) = 11.471$, $p = 0.02$, $\eta_p^2 = 0.242$).

The perceived importance of both the map and the list interface increased in the presence of MR system guidance for both Q2 and Q3. This is reflected in the statistically significant main effects of guidance on both the perceived importance of the aids for general collaboration

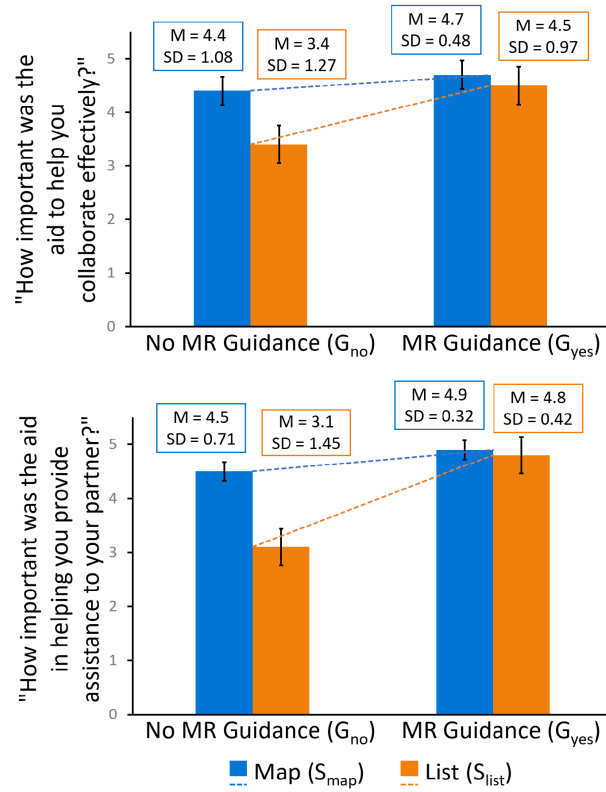


Figure 6.9. Perceived importance of the map and list aids. Top: Perceived importance on general collaboration; Bottom: Perceived importance on remote guidance.

($F(1,36) = 7.609$, $p = .009$, $\eta_p^2 = 0.174$), as well as for remote guidance ($F(1,36) = 21.204$, $p = 0.001$, $\eta_p^2 = 0.371$).

Interestingly, helpers rated the list interface to be substantially more important when they had MR system guidance than when they did not. In the presence of MR guidance, they perceived the list and the map interface as almost equally important (see Fig. 6.9). This shift in perception when MR guidance is used is reflected in the significant interaction effect for both general collaboration ($F(1,36) = 5.192$, $p = 0.029$, $\eta_p^2 = 0.126$) as well as remote guidance ($F(1,36) = 8.169$, $p = 0.007$, $\eta_p^2 = 0.185$).

6.6 Understanding Affordances

In this section, we enrich and explain the quantitative results presented in the previous section with participant feedback from our questionnaires and researcher observations of their behavior during our experiment – observations emerged from a thematic analysis of the video codes. In doing so, we also cover the major observed effects and affordances of spatial information and MR system guidance. The quotes below refer to either the helpers [H] or the workers [W].

6.6.1 Effects of 2D Spatial Information (R1)

2D spatial information allows for effective guidance even when communication is less efficient

– While the seemingly higher number of words spoken when helpers had a map could indicate a lower communication efficiency, participants’ feedback confirmed that the effect was simply because helpers were able to provide the workers with more information and thus spoke more words. Without the map (and in the absence of MR guidance), workers mentioned that once the helpers told them what object to look for, they were “*not able to help so much with where the objects were, [and the workers] mostly had to look around [themselves]*” [W19]. Helpers echoed this sentiment saying that they felt they were of “*diminished help*” [H19] with the list and sometimes “*spent more time looking [at the FPV] of the HoloLens*” [H19] in an attempt to get a sense of the layout of the task environment. Being able to use the map to understand the orientation of objects relative to the worker meant that they could use that information to direct workers and “*provide more assistance than just looking around*” [H5].

Explicit spatial information reduces effort needed for spatial deixis – The helpers’ referencing behaviors showed that the map consistently let them create more references, primarily with deictic spatial expressions. Being able to use spatial deixis is considered a cognitively expensive process as it requires extensive spatial awareness [174]. However, in our experiment, helpers had no trouble and commented that the “*easy part was [knowing] the layout of the room*” [H5]. Explicitly providing spatial information meant that they did not have to rely on a view of the task

environment to create a mental map of the space. This reduced the effort required to use spatial deixis – something that is also supported by the significant drop in task load when helpers used the map.

Additional spatial information splits helpers’ attention – Helpers used the map as an alternative source of task status; they followed the movement of objects and stated how *“following [the worker’s] progress [on the map] was less distracting than on the video stream which had a lag and jitter too”* [H15]. They were also able to correct worker actions early as they could *“easily know whether [the worker] is going to the right/wrong direction”* [H18]. However, the extreme focus on the map lead helpers to not pay attention to the FPV, and they often continued providing directions even when it was obvious that the helper had found the object. This required more back and forth verbal interaction between the participants, and reduced performance and efficiency. The following interaction (H12-W12; G_{no} , S_{map}) showcases this behavior:

H: *“So go in the corner and then right next to a snow globe...grab the white Christmas tree”*

W: *“here here, I found it...See that?” [shows object to camera]*

H: *[looking at map] “Okay, right...Somewhere around there, there’s a white Christmas tree.”*

W: *“I already [have] it...here” [shows object to camera again]*

6.6.2 Effects of MR System Guidance (R2)

MR guidance supports effective referencing through deixis even in the absence of spatial information – Our participants found MR guidance most helpful because it allowed the helper to guide the worker without knowing their spatial position in comparison to the objects. The worker was able *“to locate the object and place it wherever [the helper] wanted [them], without communicating verbally as much”* [H2].

The efficiency of the conversation primarily stemmed from the helper’s referential behaviors. First, the significantly large proportion of deictic pronouns used in the presence of MR guidance meant that most of the conversations revolving around referencing were often

just a few words (e.g. [initiates MR guidance] + “*okay now go get **that***” [H2]; [initiates MR guidance] + “*Place it next to **this** one*” [H8]). Second, we noticed that as the task progressed, some helpers shifted to not using any form of verbal referencing. They simply clicked on the icon of the object that they wanted to refer to, and let the system guide the helper. The following non-verbal exchange (H11-W11; G_{yes} , S_{list}) is an example:

H: [initiates MR guidance]

W: [puts down object at target location] + “*yeah*”

H: [initiates MR guidance]

W: [picks up object]

H: [initiates MR guidance]

W: [puts down object at target location]

Confidence in MR guidance results in less acknowledgements and increases parallelizations –

The above exchange also highlights the lack of explicit acknowledgements – a common process that is typically used to confirm the understanding of the reference [44]. Helpers often checked for acknowledgements at the beginning of the task, but they soon realized that “*after [they] tapped the desired object...the arrows took care of the rest*” [H7]. They then shifted to simply checking the FPV for confirmation instead of waiting for a verbal one, and initiated MR guidance to the next object when the worker was done. Workers also quickly adapted, and even began to expect this interaction pattern: [W13] mentioned that when the helper “*took some time to move the arrow after the [step] was done [they were] confused if [they] had successfully completed the prior [step]*”. This confidence in the MR system also allowed for parallelization as some helpers trusted that “*the arrows were able to guide [the worker] perfectly, [and they] could focus on finding the next object*” [H8].

6.6.3 Combining Spatial Information and MR Guidance (R3)

MR guidance makes spatial information superfluous – Helpers with MR guidance found it easier to choose objects in a list: H7 mentioned: “*with the map, [I] had to look over a larger area to find matches and things. With the list, [I] could easily scroll down and everything was more organized*”. Having MR guidance meant that they did not use the position of the objects very much, and found the spatial information provided by the map to be a “*an annoyance since [they] had to find the object first in the location before pointing [the workers] towards it*” [H15]. Helpers felt that using the map was confusing as they often had a lot of information, and they tended to direct the workers too much. In their feedback, workers also mentioned similar reasons, saying they received too much information when helpers had the map, and they were forced to pause or divert their attention away from the MR guidance.

Spatial information enables priming for MR Guidance – Another referential behavior that we observed when helpers used both the map and the MR system guidance was that they often *primed* the worker for the information they would receive from the MR system. This could just be information about what they would find (e.g. [initiates MR guidance] + “*Now we are picking up a coin*” [H17]), or the helper would position the worker so that when they would initiate guidance, the object was roughly in the worker’s FoV (e.g. “*Okay. pick it up and then do a 180 turn*” + [initiates MR guidance] + “*Okay, do you see that?*” [H7]). From our experiment, it is hard to say if it was indeed beneficial for the worker, but we did not see any negative impact.

MR guidance affects the perceived importance of the aids – Another place where we see an interaction effect across MR guidance and spatial information is in the perceived importance of the map or the list interface. While in the presence of MR guidance helpers reported that the list was almost as useful as the map interface, they still perceived the map as very important to be able to provide guidance throughout: “*using the map was extremely easy*” [H16].

The preference towards the map interface, even if it had limitations and diminished task

performance, could be attributed to perceiving it as more sophisticated, and therefore more useful. Some helpers acknowledged it reporting how “*the map requires more effort than the list; the list doesn’t make it fun, but is more convenient*” [H1]. The increased verbal conversation when helpers used the map could have also added to the reasons that helpers enjoyed using it: by providing more information, the map allowed the collaborators to accrue more common ground and therefore made them feel more engaged than just clicking on the list buttons.

Both MR guidance and spatial information help in unfamiliar environments – The crossover effect observed for performance in subtask 1, seemed to be absent in subtask 2. We believe that the primary reason behind it is that the collaborators were familiar with the task environment and the objects in it by the time they started the second subtask. Moving around the space in the first subtask made participants feel “*the second subtask was easier*”[W6]. Workers (and sometimes helpers) often “*remembered seeing [the referenced object] earlier, [and] was able to go straight to it*”[W19]. This suggests that the familiarity of the environment plays a role in the effectiveness of both spatial information and MR guidance.

6.7 Implications for Designers and Developers

This paper explored providing explicit information about task objects and partially offloading the referential process using MR guidance to help workers identify referenced objects. Here we discuss some of the implications that these insights might have for building collaborative MR systems optimized for referencing.

6.7.1 Offloading Referencing to MR to Increase Efficacy

Our study showed how offloading the referential process significantly increases both task performance and communication efficiency while reducing the task load on collaborators. In addition, providing explicit awareness of task objects to the helper and accurate visual cues to the worker through offloading also makes it easier to refer to occluded or hidden objects,

combats referential ambiguity, and reduces the need for explicit acknowledgements. While these characteristics are beneficial for remote guidance in general, offloading could be especially impactful in scenarios that simultaneously value efficiency and accuracy like emergency repairs or surgery.

Offloading is also well-suited for remote guidance when there is a large skill-gap between the helper and the worker. By shifting the responsibility of helping the worker identify the referenced object to the system, collaborators no longer need to build extensive common ground to reference objects, and the conversation can primarily focus on the actions that need to be taken on those objects. This can be particularly useful in scenarios like customer support where experts are helping someone outside their domain with a specific task, such as home repairs, fixing a car or a bike, etc.

Decoupling the process of making a reference, and helping the worker understand the reference through offloading, also helps when the environment is new to the helper and/or the worker. Explicit representations and MR system guidance no longer require helpers to fully understand the space to refer to objects, making it an effective technique for remote guidance in unfamiliar environments. Maintenance scenarios where parts and tasks might be standard, but the spatial layout of each environment can be different – such as a factory or engine room maintenance – are examples of where this would be useful.

Finally, while passive support for referencing through features that help view the worker's environment and provide general awareness (like the FPV) worked well in tandem with MR guidance, explicitly representing spatial information of the task objects (like the map) seemed to cause an information overload for both helpers and workers. We expect that MR guidance would not affect the positive impacts of passive awareness support. However, explicitly providing spatial information in the presence of MR guidance is not necessary and can be detrimental. An exception to this recommendation might be when the helper's tacit knowledge is tied to the position of the task objects; for example, an expert mechanic might find locating a part of an engine easier if the representation of the parts resembled the spatial layout of an engine model.

6.7.2 Making Support for Referencing More Accessible

Creating immersive MR environments requires complex set up in the helper's and the worker's environment. While the benefits of the worker being in an immersive MR environment and receiving guidance in-situ are already well known [26], helpers are typically only in immersive settings to allow for features like 3D reconstructions that increase their sense of presence and emulate face-to-face communication.

In contrast, we showed that using remote MR guidance and offloading allows helpers to successfully provide guidance even in non-immersive environments. While providing MR guidance still requires additional technology – tracking systems that provide the positions of objects in an environment – these are typically simpler to implement than systems that reconstruct the worker's environment accurately. Tracking systems also have varying complexities and fidelity as they can be sensor-based, vision-based, or both [261], giving designers and developers more freedom to choose the technology that best suits the physical task they need to support.

There are still cases where remote collaborators need powerful systems with stereoscopic displays or HMDs – like physical tasks where immersive reconstructions are truly needed and/or tracking objects externally or through computer vision is not feasible (like surgery [46]). However, these requirements limit access to remote collaborators in very specific environments. We show that using simple representations (like a web-based list or map interface) successfully reduces the effort required to reference objects. This extends rich support for referencing to everyday devices – like phones, tablets, and desktops – making guidance more scalable and increasing the network of remote experts that workers can rely on.

6.8 Conclusion

In this paper, we presented a 2x2 factorial experiment aimed at understanding how providing explicit spatial information and partially offloading the referential process through MR system guidance enhances support for referencing in collaborative Mixed Reality. Our results

show that 2D representations of spatial information provide ample support for remote helpers to provide assistance (RQ1). We also show that offloading the referential process reduces the effort required for communication (RQ2) while diminishing the role spatial information plays in the helper’s ability to provide remote guidance for physical tasks (RQ3).

Overall, we show that it is possible to design for collaborative MR systems that do not rely on expensive headsets for experts – making remote guidance and expertise more accessible, especially until immersive experiences that rely on HMDs become easier to implement and more ubiquitous. But perhaps more importantly, in reducing experts’ need for spatial information, offloaded MR guidance to task objects decouples the referential process from the task environment. Given that the need to refer to task objects is the aspect that is most tied to the task environment, the work presented in this chapter brings us one step closer to creating interfaces for remote experts that could potentially overcome the physicality of the task. The next chapter leverages this to design and evaluate an alternate interface for experts that does not fully recreate the task environment.

Acknowledgements

This chapter, in part, includes portions of material as it appears in *Do You Really Need to Know Where “That” Is? Enhancing Support for Referencing in Collaborative Mixed Reality Environments* by Janet G. Johnson, Danilo Gasques, Tommy Sharkey, Evan Schmitz, and Nadir Weibel in the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. The dissertation author was the primary investigator and author of this material.

Chapter 7

Leveraging Experts' Situated Experiences for Remote Collaboration

***Thesis Question:** Is there an alternate interface for remote experts that can tap into the benefits of high spatial presence without the need for extensive coverage of the task environment?*

Collaborative Mixed Reality (MR) systems that help extend expertise for physical tasks to remote environments often situate experts in an immersive view of the task environment to bring the collaboration closer to collocated settings. In this chapter, we design UnMapped, an alternative interface for remote experts that combines a live 3D view of the active space within the novice's environment with a static 3D recreation of the expert's own workspace to leverage their existing spatial memories within it. We evaluate the impact of this approach on single and repeated use of collaborative MR systems for remote guidance through a comparative study. Our results indicate that despite having a limited understanding of the novice's environment, using an UnMapped interface increased performance and communication efficiency while reducing experts' task load. We also outline the various affordances of providing remote experts with a familiar and spatially-stable environment to assist novices.

7.1 Introduction and Motivation

Remote experts use the skills and tacit knowledge they've developed through practicing their craft to make completing physical tasks easier, less error-prone, and more efficient for novices. However, this takes extensive coordination and requires experts to constantly understand and make inferences about the task environment and the objects within it [93]. Immersive

interfaces that provide a three-dimensional (3D) view of the task environment can make it easier for experts to understand remote environments as it lets them engage with it more naturally [278] and allows for a spatial experience that can be especially beneficial for complex tasks [136]. However, the technological difficulties of creating a fully immersive interface for remote experts that offers full coverage of the task environment (outlined in 5.3) have prompted researchers to explore a plethora of alternatives. However, none of these explorations provide the level of detail or information experts need for object-focused communication during remote collaborative physical tasks.

Even if a high-fidelity environment that mimics the real world was technologically possible, isn't perceptually the same as being physically present in it [274]. In addition, the seams and discontinuities introduced by both the technology and the distance between collaborators, alter the nature of the communication and fundamentally changes how experts experience the remote environment [140, 69]. Finally, the sustained mental and cognitive exertion required to understand remote locations can become extremely taxing on experts [245], especially when they routinely use these collaborative systems to provide guidance to multiple novices located in different distant and unfamiliar environments [152, 71].

Moreover, while our own research in the previous chapter showcases interaction techniques that provide extensive support for referencing without immersion, a fully immersive interface in and of itself can be beneficial. Room-scale immersive experiences create a compelling sense of "being there" and this enhanced perceptual richness increases users' comprehension of the virtual environment [277]. Users are, therefore, more likely to engage in more natural behaviors [278] and show higher levels of presence [34] and engagement [149] which can be especially beneficial for complex tasks [136]. Therefore, in this chapter, we set out to explore whether there is an interface for remote experts that can offer remote experts the experiential benefits that come with room-scale immersive interfaces without the need for full task environment coverage.

A key characteristic of physical tasks is that it is an extremely situated and embodied

activity that is strongly tied to the environment it occurs in [202, 36]. Consequently, in addition to being functional entities required to complete a physical task, the environment and objects within it are an integral part of how we think and behave [341]. Experts, in particular, build spatial memories and structure the environments they routinely perform physical tasks in [167]. Tapping into these situated experiences could provide new research approaches that have major implications for practitioners building MR-mediated remote guidance systems [168] and have the potential to be a key component of creating effective and enjoyable technology-mediated experiences that go beyond just being replacements for collocated communication [132].

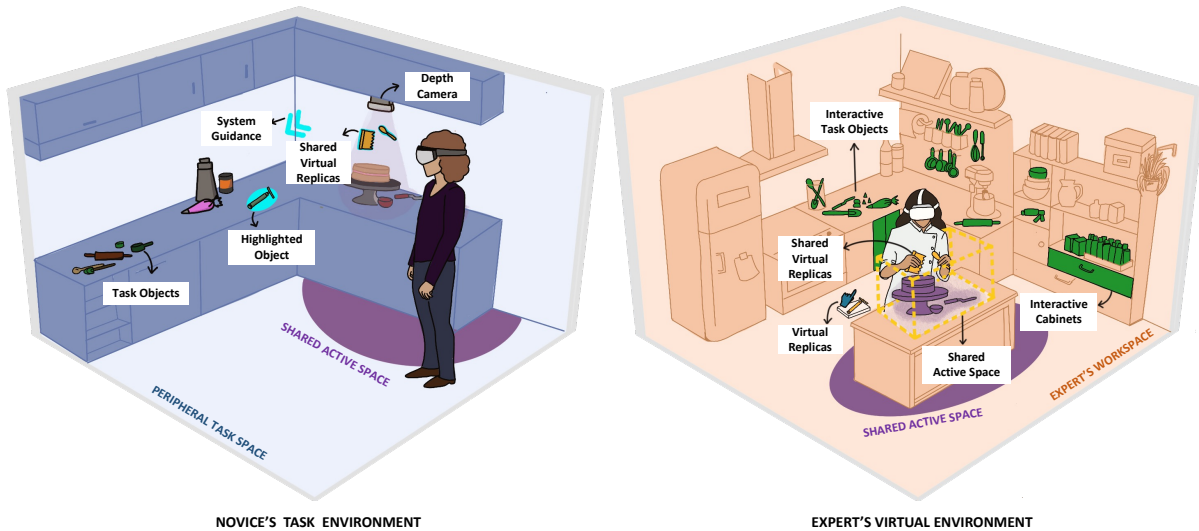


Figure 7.1. An artistic rendering of the UnMapped prototype that leverages remote experts’ situated experiences by providing an immersive virtual environment (right) that reflects their own workspace (in orange). Experts have a shared view of the active task space (in purple) streamed from a depth camera in the novice’s task environment (left). The prototype also integrates interactive virtual objects and cabinets for experts to create shared virtual replicas (in yellow) and visual guidance cues to task objects (in cyan) for novices.

In this paper, we compare experts working in a fully immersive interface that leverages their spatial memories in a familiar workspace (i.e. their own space), with an alternative interface that places experts in a full reconstruction of the novice’s environment. Results show how having experts working in a familiar environment reduces the overall effort required to provide effective guidance and eases repeated use of the system across different novice environments.

More specifically, our work contributes the following key aspects to the current research on collaborative MR systems for physical tasks:

1. The introduction of a design approach dividing the task environment into two subspaces that combines a live 3D view for the *active task space* and a static 3D view for the *peripheral task space*, within a context-aware collaborative MR system that provides system-generated visual guidance to task objects.
2. The implementation of two collaborative MR prototypes that use this approach: One that immerses the remote experts in a static 3D recreation of a familiar environment where they have established spatial memories (*UnMapped*), and one that uses a static 3D recreation of the novice’s environment instead (*Mapped*).
3. A comparative evaluation that studies the impact of these prototypes on remote guidance for physical tasks over two sessions and outlines the implications that immersing remote experts in a familiar environment has on (a) the remote guidance experience, and (b) on repeated use of collaborative MR systems.

7.2 Our Design Strategy

The task environment for a physical task can typically be divided into two broad subspaces: the *active task space*—where a bulk of the action or object manipulations that drive the task forward occurs, and the *peripheral task space*—or the rest of the environment that provides additional context and holds objects of interest required to complete the task (like tools). For instance, the surgical field and the car engine are the active task spaces for surgery and car repair respectively, and the tools or instruments needed are typically spread out in the novice’s operating space or workshop.

Existing collaborative systems tend to make design choices that focus on the active space, primarily because of the technological constraints involved with offering coverage of

both the active and the peripheral space(5.3). Now, the active space requires support for most communication processes since it fundamentally drives the task forward with support for referencing and communicating technique being key. The peripheral space, on the other hand, primarily requires support for referencing. This is especially crucial for remote guidance for physical tasks where objects in the environment play a crucial communicative role [54, 215] as their physical features and spatial relationships are directly used in the conversation. Moreover, remote guidance is most impactful for complex procedures that require specialized knowledge which often use customized tools, or those that have non-descriptive eponymous names, or both. For example, surgical sub-specialties tend to use obscure instruments that can be hard to recognize from just their names or descriptions without extensive training [76, 37]. The lack of domain knowledge among novices, therefore, means that they might not know what to look for or be able to identify a tool even if they visually register it without additional support from experts.

In this paper, we build on the prior work to design a fully immersive room-scale interface for remote experts providing MR-mediated guidance during physical tasks. A key differentiating factor of our design approach (Fig. 7.1) is that it supports the inclusion of distinct features within the active and peripheral spaces of the task environment that cater to the differing expert needs in them. We provide a live 3D view through a real-time updated point cloud within the active space, where experts need dynamic updates of the novice actions and support for both referencing and communicating action; experts can also refer to artifacts within this view using a 3D "finger pointer" object and communicate any necessary object manipulations using shared virtual replicas of the task objects. For the peripheral space, where experts primarily need referencing support, we immerse them in a static 3D mesh of the task environment and augment it with the ability to initiate visual cues that guide novices to task objects within it. This allows us to have a life-sized relatively high-fidelity view of the space that preserves the experts' ability to unambiguously refer to objects, without the need for real-time updates to the mesh.

Any view of the environment—live or static—requires experts to expend effort towards

gathering and synthesizing information about the task space from it. Experts still need to perform a visual search [75] within the virtual environment to find the tools or objects they would like to refer to or guide novice’s attention towards. And while this may be trivial in unmediated collocated collaborations, technology-mediated views of spaces change the nature of how experts experience and prioritize their attention, even without the added seams introduced by the communication medium [129]. In their paper, Johnson et al. [147] found that experts performed better when the buttons corresponding to task objects were presented as a list than when it mapped to the novice’s task environment, and having spatial awareness in addition to system guidance to objects hindered the communication between experts and novices. While a list would be difficult to parse for complex tasks with a lot of objects in the space, one implication is that experts might not need the information presented to resemble or *map* to the spatial layout of the novice’s task environment.

The inclusion of system guidance potentially reduces experts’ need to acquire the spatial information needed for referencing within the peripheral space. Moreover, the disconnected views for the two subspaces in our design approach allow us to independently change one without impacting the other. There is, therefore, an added opportunity to immerse experts in a different peripheral space without impeding their access to a dynamic view of the active space. In this paper, we explore and evaluate an alternative to recreating the novice’s peripheral task space that preserves the benefits of an immersive environment while reducing the need for experts to understand unfamiliar environments by leveraging their existing spatial memories.

7.3 Our Research Approach

Our presence in the physical world and the intimate relationship between our cognition, behavior, and external artifacts has been extensively studied under an umbrella of theories including embodied, situated, embedded, and enacted cognition [341, 217, 262]. In moving beyond screen-bound experiences, immersive and spatial interfaces have inspired HCI practitioners to

integrate how we interact with the physical world to create new design paradigms [339, 168].

Collaborative MR systems have features that build on these theories to reduce the effort it takes to communicate. Virtual pointers, gaze cues, and reconstructed hands are all technology-mediated equivalents of how we communicate with our bodies in the real world. The ability for collaborators to annotate in remote environments also allows for cognitive offloading [262] that minimizes working memory demands, and virtual replicas that can be manipulated to communicate complex actions enable experts to use epistemic actions that reduce their cognitive load [169].

A key consideration under these theories is also how we manage and adapt to the space around us. Physical task experts, in particular, have their own workspaces where they practice their craft and accrue specialized and tacit knowledge over time that is tied to their situated experiences in these spaces [202, 36]. More specifically, experts form extensive spatial memories in spaces that they are familiar with [256, 126]. This memory of artifact locations and the spatial relations between them makes it easy to retrieve objects in their environments with little to no effort [79, 256]. Experts have also been known for “jigging” or structuring their environments to simplify perception and reduce the cognitive and physical effort it takes for them to complete tasks [167]. Even when “there are standard spatial layouts of tools or instruments for a particular task, experts will modify the layout of their workspaces where object locations are personally meaningful to optimize their own performance.

When it comes to the practical use of collaborative MR systems, experts will realistically use them multiple times with novices in different environments. For example, over time a specialist trauma surgeon will help medical personnel with emergency care in rural hospitals, in ambulances, or during field surgeries. However, remote collaborations in general require sustained focus and effort that inevitably introduces a form of techno-stress and fatigue. The rise in remote work during the COVID-19 pandemic caused a phenomenon researchers now term “videoconferencing fatigue” [22] that, among many other things, is caused by the integration of frequent remote collaboration sessions to one’s work practice [71]. We can be fairly certain that

the added complexities of understanding remote environments and assisting with spatial and physical tasks from afar on a regular basis will only compound this effect, especially in critical and fast-paced settings. It is therefore crucial that we design collaborative MR systems to both reduce the effort it takes to communicate in individual sessions and simplify repeated use.

Computer interfaces have long been designed to support spatial memory and enable recall and recognition of different interface elements [266]. In general, the usability of an interface increases when users are familiar with it and a spatially-stable interface allows users to build spatial memory within it, eliminating the need for visual search. 3D interfaces support this spatial memory in an extremely intuitive way allowing users to accurately recall object locations in a spatial arrangement [55, 195].

For these reasons, our work leverages remote experts' situated experiences to make both single and repeated use of collaborative MR systems easier. More specifically, we hypothesize that immersing remote experts in a familiar and spatially-stable environment in which they already have established spatial memories will ease communication and reduce the overall effort it takes to provide remote guidance for physical tasks and make repeated use of collaborative MR systems easier. To test our hypothesis we designed and built two prototype systems based on our design approach—*Mapped* and *UnMapped*. These systems are identical except for one key difference in the peripheral space: *UnMapped* immersed remote experts in a static 3D recreation of their own familiar workspace where they have preexisting spatial awareness, and *Mapped* immersed them in a static 3D recreation of the novice's environment.

7.4 Prototype Design and Implementation

In this section, we provide details of the features that make up the *Mapped* and *UnMapped* prototype systems and outline the implementation details for them.

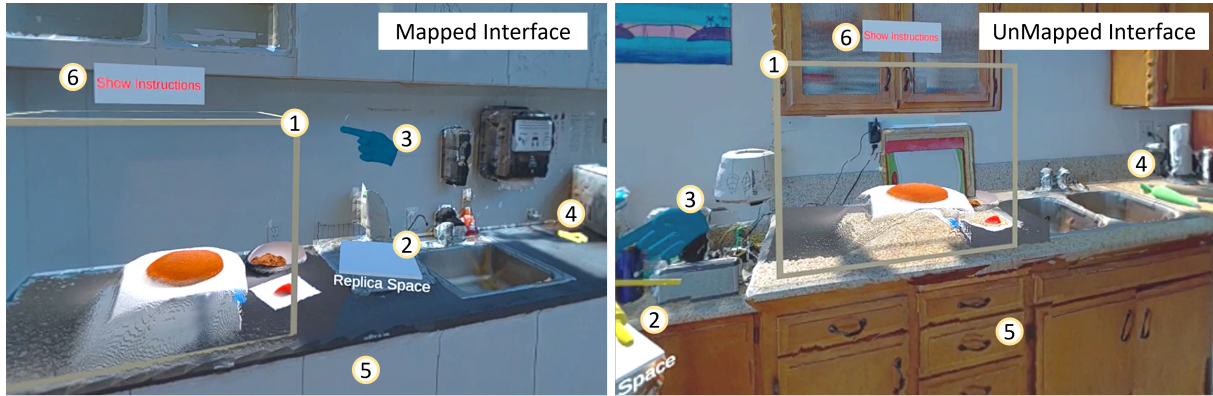


Figure 7.2. Left: A view of the Mapped VR interface that recreates one of the novice environments in our study. Right: A view of the UnMapped VR interface that recreates an expert’s personal workspace. Both interfaces include (1) A shared live point cloud of the novice’s active space, (2) the Replica Space, (3) a virtual finger pointer, (4) embedded virtual objects in the peripheral space, (5) cabinets that can be opened to reveal additional tools within them, and (6) instructions for our expert participants.

7.4.1 System Overview and User Features

Our prototype collaborative MR system provides an immersive VR interface for remote experts using an Oculus Quest 2 headset¹, and an MR interface for local novices on a Microsoft HoloLens 2 headset². The audio channel for our MR system was enabled via TeamSpeak³.

Static 3D Peripheral Space:

Once in VR, the expert is immersed in a room-scale static 3D recreation of either the novices task environment (*Mapped*, Fig. 7.2-Left) or their own workspace (*UnMapped*, Fig. 7.2-Right). The static 3D scans of the experts’ environments were made using a Matterport camera⁴. Experts could move around freely in VR within this 3D environment. The environment was also equipped with virtual objects that corresponded to task tools and a finger pointer object (Fig. 7.2-3). The virtual task tools could be present either on surfaces (Fig. 7.2-4) or within cabinets in the peripheral space (Fig. 7.2-5).

¹<https://store.facebook.com/quest/products/quest-2/>

²<https://www.microsoft.com/en-us/d/hololens-2>

³<https://teamspeak.com/en/>

⁴<https://matterport.com/cameras/pro2-3D-camera>

Shared Active Space:

The novices' task environment was equipped with an Azure Kinect⁵ (RGB+D camera) positioned over the active task space (Fig. 7.4). This camera captured and streamed a 3D color point cloud to the VR interface in real-time and allowed experts to have a view of artifacts and novice actions within the active space. The shared point cloud was embedded on top of the experts' static 3D VR environment and housed within yellow borders so users could distinguish the live regions from the static parts in VR (Fig. 7.2-1). Importantly, experts could not see live updates for any part of the novice's environment outside of this point cloud.

Using Virtual Replicas and Finger Pointer:

Our prototype system allowed experts to move the virtual finger pointer and create replicas of the virtual task objects placed in the peripheral space. To create a virtual replica, experts could "grab" the corresponding virtual tool with their VR controllers or "click" on it with a ray cast (laser) pointer. Experts could also click on cabinets to reveal virtual objects within them on a 2D menu that showed icons corresponding to the objects, allowing the expert to select any of them (cabinets could be "closed" by clicking anywhere outside the menu). Clicked objects would spawn a replica in a designated Replica Space for easy reach and subsequent manipulation (Fig. 7.2-2). The original virtual object would stay in its position but experts could now bring either the virtual finger or a replica (or both) into the active space to communicate actions or tool manipulations to the novice. When these objects were in the expert's active space, novices would see them as 3D holograms in their space as well (Fig. 7.3-1). Experts were free to move these objects outside the active space as needed, at which point the novice would no longer see the corresponding hologram in their space.

⁵<https://azure.microsoft.com/en-us/services/kinect-dk>

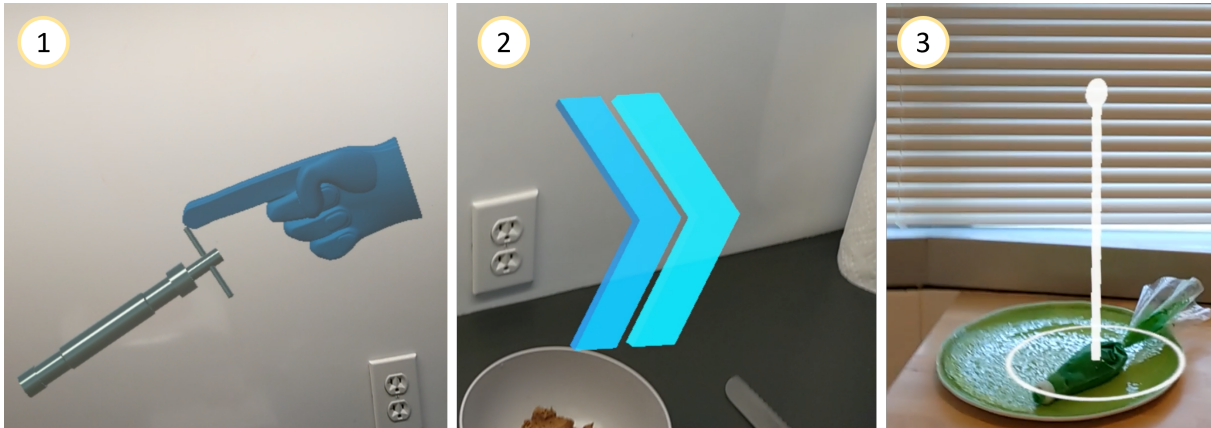


Figure 7.3. MR features for the novice in the local task environment: (1) a view of the expert’s virtual replica and virtual finger pointer used to demonstrate actions, (2) a 3D arrow, and (3) visual cues to indicate object locations in the task environment.

System Guidance cues in MR:

When experts interacted with a virtual replica of tools in the peripheral space, our prototype system automatically guided novices to the same or similar tool in their own environment if it existed. The visual cues for this guidance included 3D arrows (Fig. 7.3-2), and a circular indicator around the tool (Fig. 7.3-3) inspired by similar cues designed for the HoloLens’ small field of view [148].

7.4.2 Study-Specific Features

We added additional features to our prototype that would facilitate our study. First, the VR interface for the remote expert also included a button (Fig. 7.2-6) that would expand to show the broad goals of the task. This was done so that experts had an easy way to view their goals when in VR and ensured that our study measures were not reflective of individual differences in their retention of the task design.

Second, we did not implement real-time object detection and tracking in our prototype system and instead opted to mock this in a Wizard-of-Oz (WoZ) fashion [59, 74] that included a VR setup, MR set up, and tablet WoZ interface. Since we needed to scan multiple environments for the study, the VR setup allowed researchers to choose from a library of 3D scans to select

the static peripheral space for individual participants. It also allowed researchers to customize locations of the point cloud, Replica Space, and virtual task objects—this reflected where things were in the novice environment or expert’s workspace for Mapped and UnMapped users respectively. The MR setup allowed researchers to manually set the location of physical tools in the novice’s environment prior to a study session which the HoloLens used for the guidance feature; and the tablet WoZ interface allowed a “Wizard” to “de-highlight” objects and remove visual cues to it from a list of currently highlighted objects.

Lastly, the VR and MR setup included a calibration step to ensure that experts’ demonstrations within their active space were in the right position relative to the region captured by the point cloud. We were unable to use marker tracking since the task objects and novice actions could obscure any markers we placed. Instead, we did a simple point-to-point calibration that created a reference pose similar to one retrieved from a fiducial marker where the researcher manually selected 3 set points on the surface under the Kinect and then selected those same 3 points within the point cloud in the VR interface before a session began.

7.4.3 Implementation Details

Our prototype system was implemented in three modules: the VR system, the MR system, and the WoZ system which are connected by a web server through a series of web sockets, HTTP GET and POST requests to the server and several custom TCP sockets. Both the VR and MR systems were developed using the Unity3D game engine⁶ with the Oculus Integration⁷ and MRTK-v2⁸ packages, respectively. The web server is hosted on AWS⁹, and runs on NodeJs¹⁰. The server also hosts the WoZ interface as a Progressive Web App, which the WoZ tablet accesses via the internet.

The VR system consists of a VR headset connected to a desktop PC. The PC handles

⁶<https://unity.com>

⁷<https://assetstore.unity.com/packages/tools/integration/oculus-integration-82022>

⁸<https://github.com/microsoft/MixedRealityToolkit-Unity>

⁹<https://aws.amazon.com>

¹⁰<https://nodejs.org/en>

rendering and socket connections to both the MR system and web server, while the headset simply displays content and facilitates interaction via the built-in Oculus Link software¹¹. When experts interact with a tool, an HTTP POST message is sent to the web server and a TCP message is sent to the MR system. The MR system consists of a HoloLens application and an Intel NUC¹² computer connected to the Kinect. Visual guidance cues are initiated when the MR system receives tool selection messages from the VR system and hidden when it receives dehighlighting messages from the WoZ system.

7.5 Experiment

Our study compared the effectiveness of the *Mapped* and *UnMapped* expert interfaces with repeated use through a 2x2 mixed factorial design. The between-subjects factor—*Interface Type*—had two levels: Mapped (I_M) and UnMapped (I_U). The within-subjects factor—*Session*—had two levels: Session 1 (S_1) and Session 2 (S_2) to study the impact of repeated use.

7.5.1 Experiment Task

Experts used either Mapped or UnMapped to guide novices to collaboratively perform a cake decorating task. This task involved creating a 2-layered cake that was frosted and then decorated using icing and fondant. We chose this task primarily because, much like a variety of tasks from surgery to car repair, it included an active and peripheral task space: the decorating counter where the cake is (Fig. 7.4-Left) as the central focus for task actions, and the rest of kitchen holding the baking instruments required to complete the task as peripheral space. Cake decorating experts also spend a non-trivial amount of time perfecting their craft through formal training or in specific workspaces, and accrue concrete and tacit knowledge and decision-making skills based on what they are seeing, hearing, or feeling in the moment. Importantly, the task involves specialized tools for different operations that are typically difficult to distinguish (like

¹¹<https://store.facebook.com/help/quest/articles/headsets-and-accessories/oculus-link>

¹²<https://www.intel.com/content/www/us/en/products/details/nuc.html>

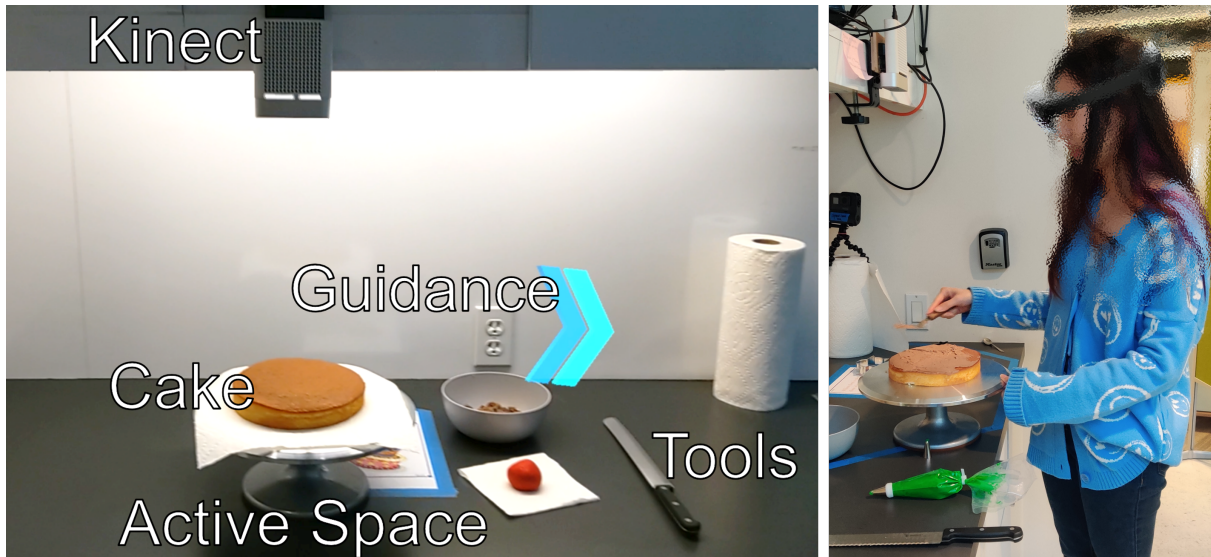


Figure 7.4. Left: The active task space in our experiment included an overhead Kinect camera, a subset of cake decorating ingredients and tools, and visual guidance cues seen through the MR headset. Right: A novice baker using MR-mediated expert guidance to decorate a cake

different piping tips) and involves specific techniques that must be used to achieve the desired result (the ability to pipe designs with icing needs more knowledge than just knowing what a piping bag is). Lastly, this task made recruitment logistically easier during the COVID-19 pandemic when access to experts and their professional spaces was extremely difficult. Many home and professional bakers use (and jig) their private kitchens to practice and experiment with their skills, which eased our access to create 3D scans of expert workspaces. Additionally, the criteria for finding a true novice was simplified since a few questions could reliably predict their comfort level with baking tools.

In our task, experts were given a specific design that they needed to collaboratively achieve with novices. This included stacking two vanilla cake layers, using chocolate buttercream as both the filling and outer frosting (achieved with a knife, a regular spatula, and an offset spatula). They also used fondant, a rolling pin, and a cookie cutter to create decorations to be placed on the cake, after which a piping bag and different piping tips were used to achieve icing decorations with two piping techniques. This cake design was chosen based on a series of pilot rounds and was considered to be complex enough that a novice would find it hard to complete on

their own, but simple enough that it could be finished within 20-30 minutes with guidance. The tools used were chosen to include objects that would be well known (knife, spatula, cookie cutter) and objects that novices either wouldn't know how to use (piping bag, coupler) or wouldn't know what they looked like (fondant extruder, specific piping tips).

7.5.2 Recruitment

We recruited a total of 60 participants of which there were 20 expert bakers and 40 novice bakers. In our experiment, each expert baker used either the Mapped or UnMapped interface (I_M or I_U) and worked with two different novices to complete the same task across two sessions (S_1 or S_2). This resulted in a total of 40 sessions.

A baker was considered an expert if they baked on a regular basis and were comfortable frosting and decorating a 2-layered cake, and a novice if they don't bake at all and/or had never baked a cake or frosted anything before. The average age of the experts in our study was 27.45 ($SD=3.28$), 8 of the 20 were male, and 12 were female. 14 experts had used a VR headset at least once before the study. 2 of the 20 experts were professional bakers and 18 were skilled home bakers. The average age of the novices in our study was 25.53 ($SD=5.16$), 26 of the 40 were male, and 14 were female. 11 novices had used a HoloLens at least once before the study.

Novice bakers were assigned to expert bakers based on their available schedules and randomized when possible. Two of the experts were familiar with both their novices from before the study, and three experts were familiar with one novice before the study (counterbalanced across conditions). Two expert participants communicated that they were not comfortable with us scanning their kitchen and were directly assigned the Mapped condition. The rest were randomly assigned to Mapped or UnMapped before the study began to give us time to scan UnMapped experts' kitchens.

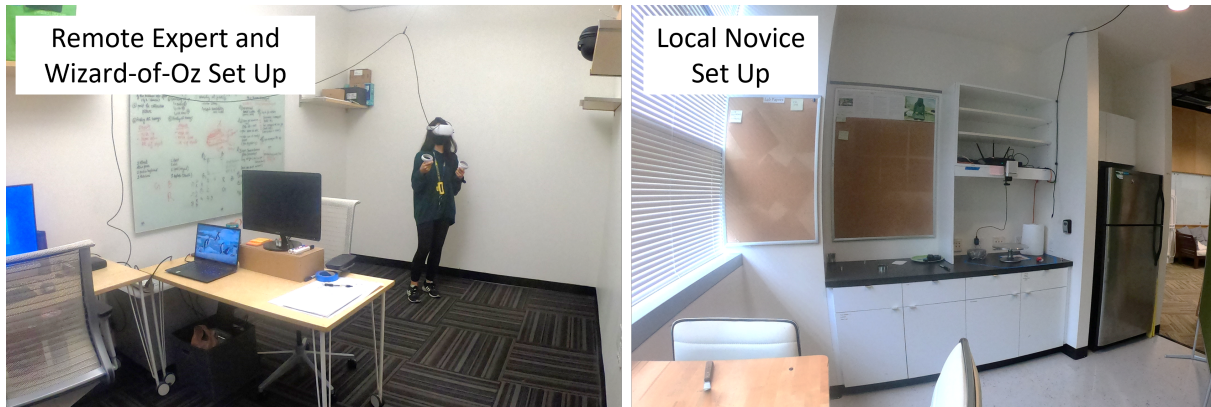


Figure 7.5. Left: Experiment space for the remote expert in our study that also included space for the Wizard; Right: One of the two novice environments in our study.

7.5.3 Experiment Setup

Scanning Kitchens:

Prior to the start of the study, one researcher scanned the personal kitchens of the 10 UnMapped experts with the Matterport camera to create their peripheral environment in VR. The researcher also made note of where experts kept their baking tools (including those that wouldn't be used in the task) and the counter they typically used for decorating—this information was used when setting up the interface prior to their session to ensure the tools would be where they are in their actual kitchen.

Expert Training:

Experts were trained on the VR interface before their first session to bring them to similar comfort levels with VR and to familiarize them with features common to Mapped and UnMapped—how to use VR controllers, initiate guidance to objects, use virtual replicas, and view the task design goals should they need it. This training lasted around 7-10 minutes, was standardized across all 20 participants, and was conducted by the same researcher regardless of the study condition. The virtual environment of the training session was a scan of a random kitchen. Experts were not shown the novice interface or the exact guidance cues but they were aware that initiating guidance meant novices would see visual cues to the same or similar object

in their space. This training also included going over the cake design of the experiment task.

Novice Environment Set Up:

To study the impact of changing novice environments, novices used one of two kitchens (seen in Fig. 7.2 and 7.5) and the order of which kitchen was used for Session 1 and 2 was counterbalanced. Each kitchen had a different layout but contained the same tools (which also included objects regularly found in kitchens that wouldn't be used in the task). However, the location of the tools was changed between spaces. The ease or difficulty of finding objects was kept similar (similar number of items within reach, on counters further away, and within cabinets). Researchers also coded these object positions within the environment for the system using the MR setup interface prior to a session.

Expert Environment Set Up:

Experts used the VR interface in a separate room on the same floor as the two kitchens (Fig. 7.5). These spaces were far enough from each other that participants in one couldn't look into or hear the other. Before the start of a session, researchers chose the appropriate 3D scan for the peripheral space and used either the locations of objects in the novice kitchens or the expert's kitchen to place objects on surfaces and within cabinets in VR through the VR setup interface. The expert's space was also where the Wizard researcher for the experiment operated the WoZ system from.

7.5.4 Study Procedure

When the participants arrived, two researchers introduced the study and explained the consent forms to the novice and expert separately outside of their respective spaces. While experts were asked to use the appropriate tools for each step if they could, their main goal was to complete the cake design. Once they reached that goal, they were free to continue decorating the cake more if they felt like it. Participants were then taken into their spaces where the researchers helped put their headsets on and started the session.

Novices were given no information about the cake design or environment or tools prior to the session. This ensured that their skill gap with the expert and familiarity with the task environment mimicked real-world scenarios (like emergency repairs and trauma surgeries) where remote guidance is used in the absence of local expertise—novices tend to be untrained in the task procedure, don't know or easily recognize specialized tools, and are often in unfamiliar environments.

While there was no formal training required for the novice, once a session started, experts were asked to guide novices to a spoon that was placed in the novice environment, and use the replica of the spoon as a system test and to help them get acquainted with each other and the interface. This was also the first time the novice saw the exact visual cues for guidance. Once this was complete, participants were free to complete the task however they chose. During the task, the Wizard researcher used the WoZ interface and a live feed of the novice's FPV to de-highlight objects when the novice picked them up. If it was obvious that guidance was initiated by mistake, the researcher removed those cues too.

After the session, experts and novices were given individual post-task questionnaires to complete. Once these were completed, they met each other and talked about the task. Researchers used this time to conduct a post-task interview. The decorated cake was divided into two and given to both participants to take home.

7.5.5 Experiment Measures and Data Analysis

We recorded the novice's FPV from the HoloLens and actions in their environment using two external GoPro cameras¹³. We also recorded the expert's VR view and the audio channel between participants. These were synced and used for video coding using BORIS [90]. Two researchers independently coded interaction and communication behaviors for 40 sessions (double-coded 4 sessions and resolved any inconsistencies to create a final coding scheme). Qualitative observations from the recordings and participant feedback from open-ended questions

¹³<https://gopro.com>

on the survey and post-task interviews were also analyzed using an open coding approach and then clustered using axial codes that resulted in the various themes of our qualitative insights.

Communication behaviors were coded using a conversation analysis approach that included coding utterances (considered as the smallest unit of a conversation—this could be a full sentence or a phrase that’s part of a longer sentence) and their intent, as well as references to objects and their type. These codes were created using a deductive approach influenced by prior work outlining collaborative communication [53, 108, 175] and refined for our study based on the double coded sessions. We only coded communication behaviors that corresponded to the peripheral space, because both Mapped and UnMapped experts were provided the same features within the active space and any communication within it would not reflect behavior changes due to the familiar or unfamiliar peripheral space. Additionally, this allowed us to reduce the impact of individual teaching styles in our measures (which primarily happened in the active space).

Our measures also included interface interaction counts, total task completion time, search time (time interval from when the expert first called for a tool to when the novice picked up the same tool), NASA TLX Task Load Index [125], Co-Presence and Perceived Message Understanding scores for collaborators’ Social Presence [124], Self-Location and Possible Actions scores for experts’ Spatial Presence [316], and a measure of experts’ cognitive representation of the novice environment using the Spatial Situation Model subscale [316]. We also included two Likert-scale questions in the post-task questionnaires (“I could tell when my partner needed assistance” and “I was able to assist my partner”) to measure experts’ perceived ability to provide guidance and modified the same questions to measure the perceived quality of guidance from the novices point of view as well.

7.6 Results

The data collected in our experiment were analyzed using a two-way mixed-model ANOVA. Survey responses were tested for internal reliability (Cronbach’s Alpha > 0.7) and

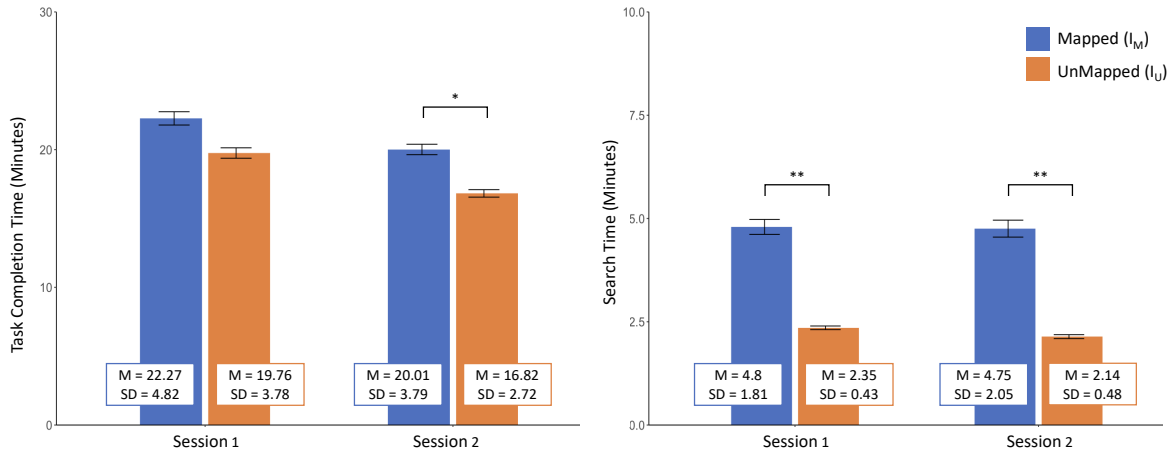


Figure 7.6. Average task completion (left) and search time (right) across study conditions with significant differences between Interface Types depicted for each Session.

then analyzed by applying the nonparametric aligned rank transform [343] which enabled the use of ANOVA after alignment and ranking. We also conducted pairwise comparisons on our data when appropriate to understand the differences between Mapped and UnMapped for each Session individually. We use an alpha of 0.05 in the results below to present the main effects of *Interface Type* and *Session* as well as the results of any pairwise comparisons. We saw no interaction effects in any of our measures and they are thus not mentioned below. Additionally, we only report the main effect of repeated use if it was significant in the interest of brevity. In the remainder of this section, F represents the F-value, p the p-value, η_p^2 the effect size, M the mean, and SD the standard deviation. The error bars in the figures depict standard errors.

7.6.1 Task Completion and Search Time

Participants completed the task significantly faster in Session 2 when compared to Session 1 ($F(1,18)=4.631$, $p=0.045$, $\eta_p^2=0.205$) and when the expert used UnMapped compared to when they used Mapped ($F(1,18)=5.4$, $p=0.032$, $\eta_p^2=0.231$). Pairwise comparison showed that this difference was not significant in Session 1, but using UnMapped led to significantly faster completion times in Session 2 ($p=0.044$) (Fig. 7.6-Left).

Isolating the time it took for participants to collaboratively search for and find the tools

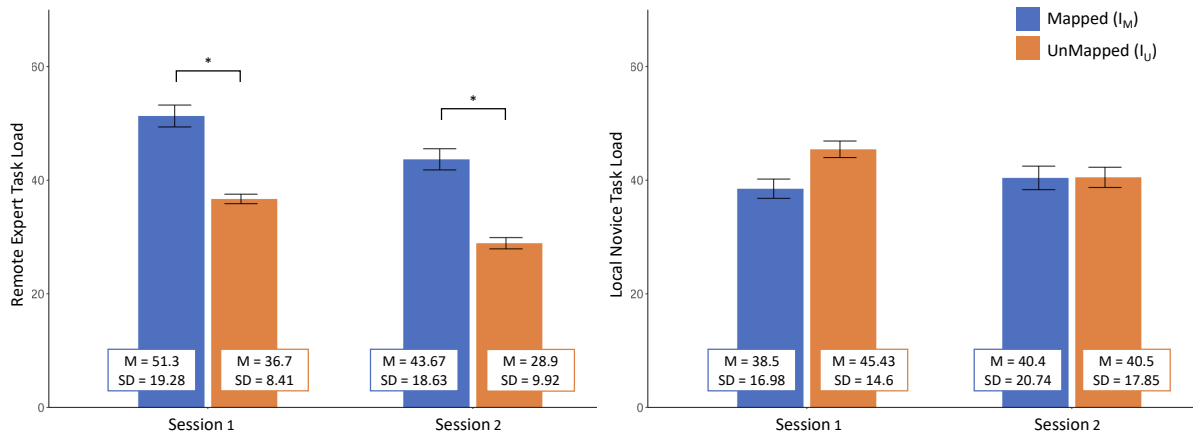


Figure 7.7. Average task load for experts (left) and novices (right) across study conditions with significant differences between Interface Types depicted for each Session.

required for the task revealed that UnMapped led to significantly faster search times overall ($F(1,18)=24.763$, $p<.001$, $\eta_p^2=0.579$) as well as in Session 1 and 2 independently ($p<.001$ for both) (Fig. 7.6-Right).

7.6.2 Collaborator Task Load

Repeated use significantly decreased experts' task load ($F(1,18)=10.472$, $p=0.005$, $\eta_p^2=0.368$). Experts also reported a significantly lower task load when they used UnMapped compared to when they used Mapped ($F(1,18)=5.569$, $p=0.03$, $\eta_p^2=0.236$). Pairwise comparisons showed that this difference was significant in both Session 1 ($p=0.042$) and Session 2 ($p=0.04$). No significant differences were found in the novices' task load regardless of the Session or experts' interface (Fig. 7.7).

7.6.3 VR Interface Interactions

The total number of interactions that experts made in VR during the task was significantly lower when experts used UnMapped ($F(1,18)=8.502$, $p=0.009$, $\eta_p^2=0.321$) overall and in Session 1 ($p=0.005$). Experts tended to open the virtual cabinets when searching for task tools in their environment and Mapped experts opened a significantly higher number of cabinets overall

Table 7.1. Expert interactions within the VR interface across study conditions. Significant differences are highlighted in green.

VR Expert Interface Interactions	Session 1		Session 2	
	Mapped (I_M)	UnMapped (I_U)	Mapped (I_M)	UnMapped (I_U)
Total number of clicks I_M (M = 33.6) > I_U (M = 19.15)	M = 36.1, SD = 15.12	M = 20.1, SD = 5.34	M = 31.1, SD = 25.04	M = 18.2, SD = 5.35
	p = 0.005		p = 0.129	
Number of cabinet clicks I_M (M = 20.75) > I_U (M = 5.4)	M = 22.8, SD = 13.36	M = 5.9, SD = 2.92	M = 18.7, SD = 21.15	M = 4.9, SD = 2.77
	p = 0.001		p = 0.056	
Number of guidance initiations I_M (M = 6.5) < I_U (M = 8.95)	M = 6.7, SD = 4.27	M = 8.8, SD = 2.15	M = 6.3, SD = 3.13	M = 9.1, SD = 1.73
	p = 0.049		p = 0.023	

($F(1,18)=16.783$, $p<.001$, $\eta_p^2=0.483$) and in both Session 1 ($p=0.001$) and 2 ($p=0.05$). Mapped experts also made significantly fewer initiations for guidance cues ($F(1,18)=4.438$, $p=0.049$, $\eta_p^2=0.198$) and this difference was significant in Session 2 ($p=0.023$) (Table. 7.1).

7.6.4 Communication Efficiency and Patterns

Both experts and novices used fewer utterances regarding the peripheral space when the expert used UnMapped compared to when they used Mapped ($F(1,18)=9.327$, $p=0.007$, $\eta_p^2=0.341$ for experts, and $F(1,18)=15.683$, $p<.001$, $\eta_p^2=0.466$ for novices). Pairwise comparison showed that this significant effect held true for both Session 1 (Expert utterance, $p=0.017$; Novice utterance, $p=0.003$) and Session 2 (Expert utterance, $p=0.022$; Novice utterance, $p=0.009$).

We also analyzed the intent for each expert utterance and the resulting communication pattern variations between Mapped and UnMapped are visualized in Fig. 7.8-Top. An utterance could have one or more of the following intents, and the measures below are the percentage of the total number of utterances of a given intent:

1. *Referring to Objects*: This was when experts made a direct reference (whether a noun or deixis) to any object (relevant to the task or not) in the passive space (Example: “Now we’re going to grab the knife”). UnMapped experts made significantly more direct object references overall ($F(1,18)=21.957$, $p<.001$, $\eta_p^2=0.55$), and in both Session 1 ($p=0.005$) & 2 ($p=0.003$).

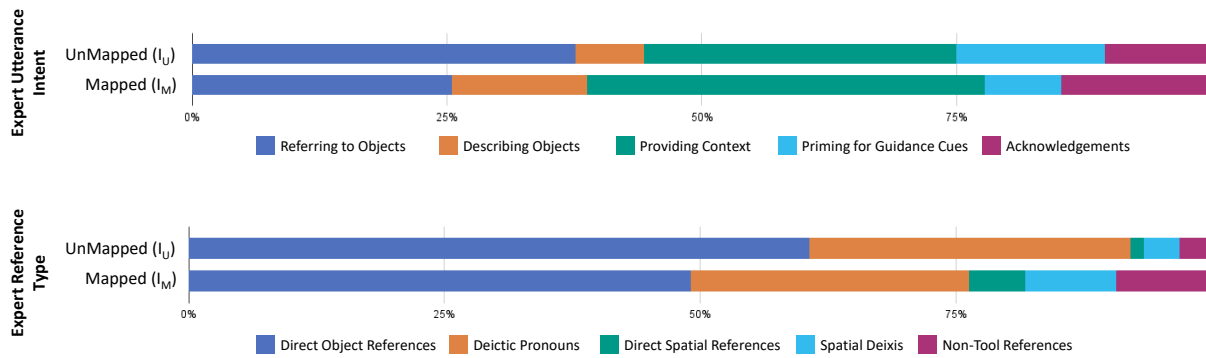


Figure 7.8. Top: Expert’s overall utterance patterns with the Mapped and UnMapped interface; Bottom: Expert’s overall reference patterns with the Mapped and UnMapped interface

2. *Describing Objects*: These utterances served to provide novices with additional details and descriptions of the object the expert was trying to guide their attention towards (Example: “It should look like a metal tube with a T on top”). UnMapped utterances to describe objects were significantly lower when compared to Mapped overall ($F(1,18)=4.822$, $p=0.041$, $\eta_p^2=0.211$) and in Session 2 ($p=0.033$).
3. *Providing Context*: These included utterances where the intent was to provide additional explanations or context around the task and why they were doing what they were doing (Example: “Having that coupler there makes it really easy to change piping tips”). There were no significant differences in how much context was provided between UnMapped and Mapped sessions.
4. *Priming for guidance cues*: Experts also primed novices, essentially giving them a heads up when they initiated guidance to objects (Example: “you should see an arrow to it now”). UnMapped experts primed for guidance significantly more than Mapped overall ($F(1,18)=4.832$, $p=0.041$, $\eta_p^2=0.212$).
5. *Acknowledgements*: These utterances primarily served as a confirmation that the novice understood the right thing or identified the correct object (Example: “okay” or “yeah, that one”). UnMapped expert utterances had significantly lower elements of acknowledgments

in Session 2 ($p < .001$).

We also dove deeper into the type of references experts used when they guided novices' attention to objects outside the active space. There were no significant differences in the number of references made by experts across either session or between those who use Mapped or UnMapped. However, there were differences in the patterns of what type of references tended to be used (Fig. 7.8-Bottom). The measures below show the percentage of the total number of references of a given type:

1. *Tool Noun or Description*: UnMapped experts made significantly more references that use a tool name or a description of it overall ($F(1,18)=4.584$, $p=0.046$, $\eta_p^2=0.203$) and during Session 1 ($p=0.031$).
2. *Deictic Pronouns*: These references included words like “that” which needed additional context to be understood. The difference in the use of deictic pronouns was not significant across interface types.
3. *Direct Spatial References*: Mapped experts made significantly more references to areas in the peripheral space (like “in the corner”) overall ($F(1,18)=8.475$, $p=0.009$, $\eta_p^2=0.32$) and in Session 1 ($p=0.002$).
4. *Spatial Deixis*: Mapped experts also made significantly more references to relative positions within the peripheral space that was deictic in nature (like “to your right”) overall ($F(1,18)= 8.876$, $p=0.008$, $\eta_p^2=0.33$) and for both Session 1 ($p=0.049$) and 2 ($p=0.005$).
5. *Non-Tool References*: UnMapped experts made significantly fewer references to other objects in the task space that were not tools required for the task overall ($F(1,18)=10.985$, $p=0.004$, $\eta_p^2=0.379$) and for both Session 1 ($p=0.01$) and 2 ($p=0.047$).

Table 7.2. Collaborators' social presence and perceived guidance quality scores across study conditions.

	Session 1		Session 2	
	Mapped (I_M)	UnMapped (I_U)	Mapped (I_M)	UnMapped (I_U)
Expert's Social Presence				
Co-presence I_M (M = 4.39), I_U (M = 4.35)	M = 4.37, SD = 0.46	M = 4.53, SD = 0.49	M = 4.41, SD = 0.59	M = 4.17, SD = 0.56
Perceived Message Understanding I_M (M = 4.06), I_U (M = 4.1)	M = 3.97, SD = 0.68	M = 4.25, SD = 0.42	M = 4.15, SD = 0.67	M = 3.95, SD = 0.64
Novice's Social Presence				
Co-presence I_M (M = 4.23), I_U (M = 4.35)	M = 4.03, SD = 0.81	M = 4.67, SD = 0.47	M = 4.43, SD = 0.44	M = 4.23, SD = 0.69
Perceived Message Understanding I_M (M = 3.99), I_U (M = 4.05)	M = 3.62, SD = 0.94	M = 4.0, SD = 0.86	M = 4.37, SD = 0.52	M = 4.1, SD = 0.44
Perceived Guidance Quality				
Remote Expert I_M (M = 4.35), I_U (M = 4.36)	M = 4.27, SD = 0.49	M = 4.6, SD = 0.44	M = 4.33, SD = 0.42	M = 4.1, SD = 0.68
Local Novice I_M (M = 4.4), I_U (M = 4.37)	M = 4.15, SD = 0.9	M = 4.37, SD = 0.82	M = 4.65, SD = 0.44	M = 4.38, SD = 0.49

Table 7.3. Experts' spatial presence scores and cognitive representation of the novice's environment across study conditions.

	Session 1		Session 2	
	Mapped (I_M)	UnMapped (I_U)	Mapped (I_M)	UnMapped (I_U)
Expert's Spatial Presence				
Self-Location I_M (M = 3.91), I_U (M = 4.2)	M = 3.9, SD = 0.98	M = 4.15, SD = 0.49	M = 3.93, SD = 1.07	M = 4.23, SD = 0.68
Possible Actions I_M (M = 3.75), I_U (M = 3.74)	M = 3.8, SD = 0.79	M = 3.86, SD = 0.94	M = 3.7, SD = 1.05	M = 3.6, SD = 1.09
Expert's Cognitive Representation of Novice Environment				
Spatial Situation Model I_M (M = 3.82) > I_U (M = 2.67)	M = 3.82, SD = 0.8	M = 2.73, SD = 1.16	M = 3.82, SD = 0.99	M = 2.62, SD = 1.12

7.6.5 Self-Reported Questionnaire Scores

Both experts and novices reported similar social presence scores regardless of which interface the expert used (Fig. 7.2). Experts also reported no differences in their perceived ability to provide guidance, while novices reported no perceivable change in how well they were being assisted by the experts in any of our conditions (Fig. 7.2). Additionally, experts reported similar levels of spatial presence in both Mapped and UnMapped interfaces (Fig. 7.3). There were no significant effects observed for any of these measures.

A measure of experts' mental model or cognitive representation of the novice's envi-

ronment revealed significantly lower scores for UnMapped experts when compared to Mapped experts ($F(1,18)=9.525$, $p=0.006$, $\eta_p^2=0.346$) (Fig. 7.3). There were no other significant effects.

7.7 Observational Results and Participant Feedback

In this section, we report on common feedback and observations across all 40 sessions and discuss our qualitative results specific to the use of Mapped and UnMapped overall and with repeated use.

Experts in our study used the live point cloud of the active space to successfully sustain task status awareness. However, the noise and distortion inherent with point clouds hindered their ability to view fine details on the cake. The camera being positioned overhead also meant that the sides of the cake weren't always visible and novices often blocked the view when working on the cake. As such experts found *"the task view was a bit irritating"*[E16]. However, experts learned to work with the imperfections by having novices rotate the cake to accommodate parallax errors and occasionally step back for a full unblocked view. Experts also used their own experiences to *"fill in the gaps"*[E8] and relied on verbal communication to keep track of subtle changes within the active space. Additionally, tracking system drifts introduced calibration errors that made pointing at specific parts of the cake hard, especially for *"fine motor movement [like] precise frosting movements"*[E13], so experts resorted to gesturing at broad regions and providing more specific information verbally.

Experts didn't seem to be affected by the lack of dynamic updates about any "prep" actions unless novices took longer than expected outside the active space without verbal updates. When that happened, experts would ask for an update after a period of silence. Having a static view of the peripheral space meant that experts couldn't confirm whether the novice had the correct tool. Novices, therefore, *"held tools in front of the [point cloud] for [the expert]"*[N7] but experts still found it challenging to provide confirmation because their point cloud was *"grainy"*[E1] and metallic instruments wouldn't show up well.

Experts and novices both found system guidance *“tremendously helpful [because] it avoided having to explain the features of tools”*[E18] and mentioned that it *“was the easiest part”*[N4], especially for cluttered areas or when tool names didn’t give novices a sense of what to look for. E16 articulated this sentiment saying *“having the objects pointed out helped immensely, it made not having a great view tolerable”*. In supporting the referencing that’s crucial for physical tasks, system guidance allowed experts to easily direct novices’ attention to objects without extensive verbal descriptions. Even when experts knew tool locations, they preferred *“not having to give [their] partner exact locations”*[E20] and novices found the visual cues *“much more useful than understanding [the expert’s] speech”*[N3].

Experts and novices also emphasized that using virtual replicas to demonstrate action was key to seamless communication when it included specialized baking tools and techniques that were alien to the novice. N7 reiterated this saying *“I liked the piping demonstrations with the holograms...I would have no idea if you would have audio described it to me”*.

7.7.1 Behavioral and Experience Differences Between Mapped and UnMapped

Being in a familiar environment increased experts’ confidence and decreased stress:

In general, UnMapped experts had a more jovial tone and voiced feeling extremely comfortable during the task. They mentioned that *“it genuinely kind of felt like I was working in my kitchen”*[E8], and that this familiarity of the environment made the task less daunting. For instance, E19 mentioned *“it actually put me more at ease just because of my environment and I felt more calm”*. E6 also mentioned that this came as a surprise, saying *“I thought it wouldn’t matter at first but then it was actually really helpful because I can just be like ‘oh where do I keep the offset spatula’...I feel it reduced a little bit of my mental load [which was] unexpected and I actually felt like [the novice] was in my space making a cake”*.

UnMapped experts also emphasized that it wasn’t just their memory of object locations that made them feel more confident—having familiar objects in their peripheral vision that

they were used to while working in their own kitchens also played a part. E18 recalled this phenomenon saying *“I found it very helpful [that] I could even see my plants a little bit, that actually helps a lot...obviously it was more distorted than reality but it was still helpful”*.

Mapped allowed for collaborative searches, UnMapped removed the need for searching:

When the position of a tool wasn't obvious from a quick scan of the peripheral space, Mapped experts described the tool they were looking for and engaged in a collaborative search where both participants moved around and opened cabinets in their respective environments. Some experts searched their virtual environment as they were describing the tool and only asked novices to join in if they hadn't spotted it when they were done with the description. Mapped experts found searching was especially hard when objects were occluded or in unexpected locations and mentioned the imperfections in the 3D scan made visual search harder. There were two Mapped sessions where neither participant found the relevant tool but experts were able to assist novices with using other tools as a workaround.

Experts who used UnMapped, on the other hand, found directing novices to baking instruments trivial and intuitive since they were *“already familiarized [with the environment]...this is very easy on my end...I know the rolling pin is going to be where it is in my actual home, my baking tools are in these drawers...it was really nice way to navigate through the kitchen”*[E11]. They were also more forgiving towards the imperfections of the 3D scan and said that *“it just felt like the apartment was covered in paper”*[E7] and that they stopped noticing the imperfections once they focused on the cake and didn't need a detailed view to find objects in it.

Mapped views required experts to attend to the full task environment, UnMapped views allowed them to focus on the active space:

Having spatial awareness of the novices peripheral space allowed Mapped experts to use spatial references and other objects to ground their references even when they used guidance. However, the post-task interviews revealed that experts felt the need to constantly scan the full environment to ensure the task was moving forward smoothly. They would often visually scan

the peripheral space for the next tool while novices completed the current step and primarily used verbal updates to keep track of actions in the active space. E3 mentioned that splitting their attention between the two spaces so novices had less searching to do when they reached the next step was the “*hardest part*”.

In contrast, UnMapped experts didn’t need to expend energy towards finding objects and also experienced some downtime when novices were completing relatively long actions like frosting the cake. None of the experts realized the novice in the second session was in a different environment. Many experts mentioned noticing something different but “*just assumed that was a slightly different positioning of the camera*”[E11]. When asked whether they would have preferred to see the novices’ full environment, UnMapped experts said “*if I was in the same space as [the novice] I could be like ‘oh its on your left’...[but] honestly it has very little impact because once we are in the task, I’m just looking at the task space, I don’t care about the surroundings*”[E16]. E18 and E13 echoed this feedback saying “*all I needed to see was the cake and ingredients, it didn’t really matter where it would have been*” because “*none of my [instructions] were about [the] space, it was about the actual action*”.

A Mapped or UnMapped view impacted the use of virtual replicas for acknowledgment:

We observed that virtual replicas were often used as a form of acknowledgment and visual confirmation for novices to check if they got the right tool. However, Mapped experts tended to do so as a way to repair other forms of acknowledgment—like when verbal descriptions were not enough or when the point cloud view was not clear. Once UnMapped experts were confident that the guidance cues were reliable, they no longer needed verbal acknowledgments and preemptively put virtual replicas in the active space such that novices saw it on their way back from picking up the tool. For example, E8 told their novice “*I’m gonna have you grab another type of spatula, an offset spatula <clicks to initiate guidance and moves replica into active space> you should see arrows to it now. I also put it in your space*”. As the task progressed,

experts stopped explicitly stating this and the acknowledgments became non-verbal, making the task flow more seamless.

Mapped experts compared their experience to being collocated, UnMapped experts felt like it was a fundamentally different experience:

While all experts mentioned feeling strongly immersed in VR, Mapped experts tended to compare their experience with instructing someone when collocated and recalled feeling restricted because *“figuring out how to describe objects without being physically present to gesture or show things myself was strange and tricky”*. They felt that although *“the environment looks real, [they] can’t really do anything...even though it feels like we are in the same space”*[E1]. UnMapped experts, on the other hand, voiced that this experience was fundamentally different. Some experts felt like they were working with the novice in their own kitchen, and others felt like they were in a whole new hybrid space. When reflecting on this, E17 mentioned *“I didn’t feel like it was really my kitchen [since] there’s some distortion which cuts the reality [and] I was not feeling like I was in [the novice space]...but I was feeling like we had the connection”*.

The Mapped system can be used in the absence of system guidance, the UnMapped system is dependent on it working:

While UnMapped experts found their experience enjoyable overall, they also reflected on the fact that they were *“depending completely on the technology to point [the novice] to the right direction”*[E18]. In our study, the HoloLens briefly lost tracking in the novice environment in two Mapped sessions and one UnMapped session which resulted in the visual cues pointing novices in a different direction from where the object of interest was. Mapped experts were still able to use the static view to search for objects and redirect novices to them while the UnMapped expert could not do more than verbally describe the object and let the novice search for it independently.

7.7.2 Learning from Repeated Use

All experts felt more confident in their abilities to provide effective guidance in the second session because they got used to the interface and knew what to expect from it. They also modified their instructions because they thought *“about the things [where the] first novice [got] confused and [were] trying to proactively address them”*[E17]. Experts also prepped novices about the technological limitations at the start of the task and often reminded them that their additional questions were because of it. For example, E6 constantly reassured their novice when asking for more information saying things like *“I just wanted to double check, it’s just the limitations of the VR”*.

In general, experts trusted the guidance feature more and used fewer verbal acknowledgments when they used it the second time. Reflecting on it, E18 said *“I had a better sense of like...they can see [an arrow] right now, this is fine...so I had less of a feeling of [saying] ‘look around for an arrow’...I trusted [the feature] more”*. However, more trust in the feature didn’t impact Mapped experts’ ability to use it more and the primary difference we saw in the second session compared to their first was that they looked around the peripheral space more intently during the start to preemptively find tools, but would let novices do more independent searching for tools they didn’t catch in this initial scan when the time for the steps that involved it came.

7.8 Discussion

Our results indicate that when paired with a context-aware system that can direct novices to objects of interest, a static view of the peripheral space and a live view of the active space in a task environment allows remote experts to experience the benefits of a fully immersive interface that sufficiently supports their ability to provide effective guidance during collaborative physical tasks.

7.8.1 The Impact of a Familiar Environment on Remote Guidance

Having remote experts immersed in a familiar environment where they have previously established spatial memories led to better performance and decreased task load. Additionally, the disconnect with the novice's task environment did not impact the experts' spatial presence within the virtual environment. The UnMapped approach also allowed for a similar sense of social presence among collaborators, despite remote experts having a limited understanding of the novice's environment outside of the shared active space. Novices' task load and perceived effectiveness of guidance were also not impacted by the experts being in a different environment from their own.

Being immersed in a recreation of their own workspaces where they had prior knowledge of object locations also meant that experts spent little to no time searching for task-relevant objects, making initiating system guidance for novices trivial. In contrast, when immersed in a recreation of the novice environment that was unfamiliar to them, they used this feature less as they relied on a collaborative search for objects with the novices if it wasn't readily visible to them. In fundamentally changing how effectively and frequently system guidance can be used, the UnMapped approach no longer needed experts to align or adapt how they refer to objects for it to be easily understood by the novice (a strategic process that's commonplace in object-focused communication [35, 88, 54, 139]). Experts could use direct references and minimize the additional verbal information and spatial expressions they needed to disambiguate their phrases, resulting in more efficient communication.

The compounding effects of reduced visual search and communication effort when a familiar environment was paired with system guidance also led to larger experiential benefits. Experts perceived it to be fundamentally different from being collocated, and felt more at ease as it freed up their limited attentional resources. This also allowed them to focus on the actions within the active space where their expertise and tacit knowledge are most impactful. Designing for the nuances of expert needs during collaborative tasks, therefore, allowed us to leverage the

strengths of an immersive environment to simplify perception and offload cognition [167]. In facilitating an experience that didn't fully imitate physical proximity, the UnMapped approach brings us a step closer to creating collaborative MR systems that allow us to go *beyond being there* [132].

7.8.2 The Implications of a Spatially Stable Environment on Repeated Use

The theory of least collaborative effort [51] suggests that collaborators minimize the total effort required for an interaction, and we often allocate our attention to environmental artifacts based on their ability to contribute to the overall task success [78]. This means that we often preemptively reduce the cognitive resources we make available if their perceived benefits don't substantially impact our collaborative goals. The added frustration of having to search unfamiliar environments [266] is clearly visible in our experiment. Specifically, the difficulty of understanding remote environments in the first session caused experts using the Mapped interface to only spend time comprehending the novice environment in the beginning of the task and opt to revert back to verbal communication (indicated by the increased heterogeneity in utterance patterns and the drop in significant differences of spatial references as well as qualitative insights) for objects that weren't found in this quick search. Experts who used UnMapped, on the other hand, experienced a more seamless conversation with their novices that included non-verbal acknowledgments because of their ability to use system guidance and their increased trust in it the second time.

Therefore, while the increase in overall performance and decrease in task load with repeated use indicate that all experts benefit from previous experience with the interface and the task, having a familiar environment that was spatially-stable across sessions provided additional benefits and heavily increased this effect. While it is difficult to isolate the impact of the familiar space from the effects of having just any spatially-stable interface in our study, we can reasonably predict that repeated use with varying novice environments is less disruptive

with a stable peripheral space—even if it is one that doesn’t initially hold any meaning to the expert. Experts can develop familiarity and spatial memory in any spatially-stable virtual environment over time, making this approach a valuable way to reduce the cognitive burden of understanding multiple remote environments when experts use collaborative MR systems regularly. Additionally, these environments can be designed to be personalizable, allowing experts to structure and progressively optimize their immersive interfaces to work better for themselves.

7.8.3 Designing for Practical Use and Limitations

Despite the relatively straightforward benefits of our design approach on collaborator experience, it is important to acknowledge the impact of technology on the practical implications of our work and limitations of our experiment.

First, the relatively lower-fidelity live 3D view of the active space made a non-trivial impact on user experiences. This could have impacted the self-reported questionnaire scores, especially for experts who used the Mapped interface, since they compared the experience to collocated communication more than experts who used the UnMapped interface. However, since this feature was kept constant across conditions, our results still reliably indicate that the UnMapped approach doesn’t introduce additional negative effects to collaborator experiences. Moreover, our ability to provide real-time high-fidelity views will inevitably increase as technology gets better [170].

Second, the ability to provide system guidance as a feature is fundamentally dependent on technology like object tracking for tools in the novice’s environment. Having an interactive static peripheral space also needs prior setup to curate the experts’ tools and environment based on the tasks they are predicted to provide guidance for, with UnMapped requiring additional personalization for experts. While we employed a WoZ approach for this, designers and system developers can choose from a plethora of techniques like computer vision, infrared markers, and RFID tags combined with machine learning or neural network architectures [127] to detect

and classify complex objects in the real world. These techniques have already been used in MR systems to track tools during physical tasks like engine components [201] and surgical instruments [206, 233]. Similar methods can be used to detect and track domain-specific task objects in the novice’s environment, and pair them with the tools in the expert’s virtual task environment.

While this still requires computational resources and prior setup, a context-aware system with system guidance tied to a static 3D view allows us to provide a fully immersive environment while shifting the real-time complexity over from the network—where limited bandwidth and delays will likely remain a constraint in our near future—to individual local environments. Doing so also enables practitioners to choose the technology that best fits the needs of the task they are supporting as different techniques allow for varying levels of fidelity and accuracy [261]. For example, RFID tags provide a low-cost option that can cope well in a cluttered space [110] and real-time vision-based tracking can be modified to be computationally frugal [198, 15]. It is, however, important to note that the Mapped interface is more conducive to recovery if technology fails with system guidance since experts can still make spatial references within the peripheral space based on the static view if objects’ locations haven’t drastically changed.

Third, even if bandwidth wasn’t a constraint, our design approach would allow system designers to use more computational resources to enhance and optimize the view and communication experience within the crucial active region of the task environment. Moreover, designing around experts’ spatial memories made them more tolerant towards artifact distortions in the recreated environment, and having this environment be their own workspace even introduced positive effects on their level of comfort due to the familiarity of objects in their peripheral vision. Additionally, while a VR interface allows experts to provide guidance from anywhere, the UnMapped approach could potentially be implemented as an MR interface that experts could use from within their physical workspaces as well.

Another factor to consider is that our experiment studies the impact of the peripheral space being a familiar environment where experts have established spatial memories. We believe

a majority of our results primarily reflect the impact of this spatial memory where experts have preexisting knowledge of object locations in the UnMapped condition. While we did see experiential benefits of their environment being a familiar space (like lower stress and added comfort), we cannot reliably predict whether users would experience similar benefits without this prior knowledge of object locations. There is, therefore, merit in further exploring the isolated impact of familiar artifacts in the environment for remote collaborative tasks in the future.

We also acknowledge that our approach might not be suited for all types of physical tasks like room-scale tasks where the active space is the whole environment, tasks that don't use domain-specific tools, or if experts don't typically have a workspace where the spatial relationships of objects are meaningful to them. This approach would also be less impactful if novice environments don't vary, or the skill gap between collaborators is negligible as novices would already be familiar with the domain-specific tools used.

That said, remote collaborative systems are most impactful when local collaborators are true novices, especially in emergency and resource-strapped settings. Designing interfaces that ease participation for both experts and novices is thus crucial to extending our collective skills and expanding the reach of expertise. The overarching goal of our work, therefore, has been to examine the technology capabilities at our disposal to shape and explore effective design approaches that we can take when building remote collaborative MR systems. With this paper, we have added the Mapped and UnMapped approaches to the repertoire of possible design choices that practitioners can use to decide what features a specific task setting calls for.

7.9 Conclusion

In this chapter, we introduced a new design approach for remote expert interfaces in context-aware collaborative MR systems, and explored an immersive environment that leveraged experts' situated experiences in familiar workspaces as an alternative to recreating the complete task environment. We designed and built the *UnMapped* prototype which allowed remote experts

to work out of static 3D reconstructions of their personal workspace that embedded a live 3D view of the active space within the novice’s task environment. Experts could also use virtual replicas of task-relevant objects to communicate action and initiate system-generated visual guidance cues that directed novices towards tools in their peripheral space. We also built *Mapped*, an identical prototype that immersed remote experts in a static reconstruction of the novice’s environment instead of their own workspace, and compared the effectiveness of UnMapped against it on remote guidance for physical tasks overall and with repeated use of the system through a 2x2 mixed factorial experiment.

Our results reveal multiple benefits of having a familiar and spatially-stable environment where experts have preexisting spatial memories, and showcases how this makes it more conducive for experts to use MR-mediated guidance systems as a routine part of their work practice. Despite having a limited understanding of the novice’s environment, using the UnMapped interface increased performance and communication efficiency, while reducing task load for both experts and novices. UnMapped also allowed experts to focus their mental resources on the shared active space and eliminated the need for visual search within the peripheral task space. The paper also discusses the various benefits that providing remote experts with a familiar and spatially-stable environment has on collaborator experiences and sustained use. Additionally, through this work, we demonstrate the impact that considering users’ existing work practices and strengths has on creating more usable and effective collaborative MR systems.

Finally, the *UnMapped* approach is one that offers high spatial presence without full coverage of the task environment. In other words, it is, to the best of our knowledge, the first interface for remote collaborative tasks that goes beyond recreating physical proximity while still providing the rich communicative support required for it. The next chapter concludes this dissertation and discusses the approach that made this possible.

Acknowledgements

We thank UC San Diego's Design Lab and The Qualcomm Institute for providing the space and helping with the logistics required to conduct this research. We would also like to thank our participants for their valuable time, and for allowing us to access their personal spaces. Finally, thank you to Jiju Johnson for lending us the scanning equipment used in our experiment.

This chapter, in part, includes portions of material as it appears in *UnMapped: Leveraging Experts' Situated Experiences to Ease Remote Guidance in Collaborative Mixed Reality* by Janet G. Johnson, Tommy Sharkey, Iramuali Cynthia Butarbutar, Danica Xiong, Ruijie Huang, Lauren Sy, and Nadir Weibel in the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. The dissertation author was the primary investigator and author of this material.

Chapter 8

Conclusion

This dissertation conceptually and empirically explored how XR can support remote collaborative physical tasks where an expert who is geographically separated from the task environment could guide a novice present within it. In this chapter, I summarize the work it presents. I also reflect on the research approach it takes and discuss opportunities for future work that could benefit from it.

8.1 A Summary and Discussion

The recent advancements in XR brought about an opportunity to provide richer support for collaborative physical tasks than was previously possible because MR facilitated the provision of in-situ cues that enhanced novices' ability to follow procedural guidance. When it came to supporting the remote expert, immersive technologies allowed them to be immersed in a virtual recreation of the task environment where they could engage with novices in ways that mirrored how they would in the real world.

While recreating the experience of being present in the same physical space offers a multitude of benefits to the collaborative experience, my interest in XR as a collaborative medium stemmed from its ability to provide users with experiences that could go beyond the constraints of what is possible in the real world. Incorporating this into the design of systems for collaborative physical tasks in a substantial way, however, proved to be a uniquely challenging problem – primarily because of how intricately tied to the task environment the communication required for it is. Providing a view of the environment, in particular, is seen as indispensable to remote

experts' ability to provide effective guidance. Therefore, at the start of this dissertation, it was unclear how we could break past the constraints of the physicality of the task, what an interface that realized that vision would be, and how it could benefit the collaboration in a meaningful way.

Motivated by these uncertainties, I aimed to take a human-centered approach to understand how XR's unique capabilities could be leveraged to create systems that ease the effort required for remote collaborative physical tasks. My thesis statement claimed that leveraging these capabilities to design information representations and communicative mechanisms that minimize the environment awareness required for object-focused communication during remote physical tasks would lead to more flexible collaborative environments without fully replicating physical proximity. The research initiatives undertaken in this dissertation provided empirical evidence and design strategies to support this thesis. Each chapter uncovered a piece of the puzzle that expanded our understanding of the communication task and the design space of what's possible, and seemingly small steps came together to create the UnMapped interface which offered an experience that felt fundamentally different from being collocated. Here, I will summarize both our research approach and how each chapter contributed to my overall research goal.

Given the complexities of the task and the many technological capabilities we have at hand, there are always a plethora of interface possibilities and design paths that can be taken. A key component of understanding how to expand the design space in meaningful ways is to first have a humanistic understanding of the task at hand. Chapter 2 presented a conceptual framework and a native model for collaborative XR that outlined collaborator needs and described how XR could impact their situated experience. This framework gave us a sense of what technology could achieve and established the theoretical foundations for subsequent research.

Chapter 3 then studied the utility of MR for novices during time-critical tasks by designing and evaluating MR guidance cues for cardiopulmonary resuscitations against screenbound instructions. The primary goal here was to understand whether MR's spatial awareness and

technology as it stands today offered enough support for novices to comprehend and act on instructions for physical tasks. However, a secondary outcome of the work in this chapter was that we now had an MR guidance cue designed to work with the limitations of technology that we were confident could direct novices to objects in the task environment without external support.

Chapter 4, on the other hand, explored how the spatial and immersive capabilities of XR could support remote experts in their ability to provide such MR guidance to novices during emergency trauma surgery. While the interface designed here mirrored many of the features we see in existing collaborative XR systems today, a realistic evaluation of this interface being used during emergency surgery on perfused cadavers uncovered in-depth insights into how these features impacted the collaboration. This chapter highlighted both the information and communicative richness of the medium as well as the persistent challenges of sustaining object-focused communication. It also brought to light how collaborator needs vary within different regions of the task environment and offered empirical evidence for referencing being the communication process most tied to the environment. This then prompted us to dig a little deeper to understand how this critical communication process has been supported to date.

Chapter 5 presented a communication-focused overview of the design landscape of existing XR-mediated collaborative systems based on the conceptual framework introduced in this chapter. Looking at the research out there through a theoretical lens helped us break down how the various research and design explorations in the field supported the different communicative needs of physical tasks. More specifically, it helped us understand the two key research gaps concerning referencing that could reduce the joint effort required for it.

Chapter 6 then explores the two opportunities identified to enhance support by referencing. It leveraged MR's spatial or environment awareness to both provide explicit representations for the spatial information experts needed as well as explore the guidance cues evaluated in Chapter 3 to then be able to partially offload the communication of these references to the system itself. This, in turn, allowed us to study the impact of offloading guidance on remote experts' need for spatial information in a controlled setting, which told us that we could provide a high level of

support for physical task referencing without the experts needing extensive awareness.

All of our research efforts so far then came together in Chapter 7 to create an approach that was designed based on remote collaborators' varying needs within the task environment. The ability to decouple views of the active and peripheral space then gave us the freedom to swap out the peripheral space with a recreation of experts' own workspaces with the UnMapped interface – something that was only possible because it leveraged the insights from the previous chapter. Additionally, the previous chapter evaluated being able to partially offload referencing in a controlled manner that isolated the process. In contrast, the work in this chapter evaluated the technique in a more ecologically valid manner where we could see how it would play out in a more realistic task that involved all the other communication processes and complexities typically seen during physical tasks. Moreover, in providing an interface that wasn't fully tied to the task environment, the research in this chapter was also able to explore how we could use the immersive aspects of XR to work with the experts and make both single and repeated use of the system easier.

Overall, the work in this dissertation provides a blueprint for how we can systematically leverage emerging technology capabilities to reduce the effort it takes to partake in rich remote collaboration. It also demonstrates the impact of considering users' existing work practices and strengths on creating more usable and effective collaborative MR systems. It is our hope that this inspires researchers and practitioners to explore the different design choices that technology affords to create rich collaborative experiences, even if it doesn't fully recreate reality or mimic how we collaborate with others in our physical spaces.

8.2 Looking forward

Continued research toward understanding and designing for the intricate challenges of facilitating seamless collaboration at a distance is vital as distributed work environments become more ubiquitous in an increasingly connected world. The impact of having better collaborative

systems that make expertise more accessible to regions that don't have it, in particular, cannot be overstated – it holds the promise of transforming various industries and leveling the playing field across the world by bringing knowledge and skills that were previously out of reach. For example, real-time support from expert medical practitioners situated elsewhere can reduce the healthcare disparities between the urban and rural world and reduce our reaction time during natural disasters and emergencies. In general, effective remote collaborative systems allow many industries to benefit from global expertise without the need for significant investments in physical infrastructure and substantial environmental and fiscal costs towards travel. In this section, I outline key research areas for building more sophisticated systems that can amplify support for human collaboration.

Foundational Research for Physical Tasks

This dissertation focused on leveraging XR's visual and spatial capabilities to enhance support for object-focused communication. The last two chapters, in particular, revolved around foundational research towards enhancing support for referencing, as it was identified as the key communication process that could not occur without information from the environment. However, the visualization of the current design landscape in Chapter 5 (Table 5.1) reveals many opportunities to leverage XR's various capabilities to enhance support for different collaborator needs. Table 8.1 builds on this to visualize how the different capabilities of XR currently support the various collaborator needs during physical tasks. For example, if you take a cell, say, [C1,N1] that corresponds to capability C1 and need N1, the number in the cell indicates the number of features that use capability C1 to support need N1. A darker shade indicates a higher frequency, and the gaps in the table indicate possible research opportunities for the future.

This table showcases that even though MR environments have the capability to integrate information that we perceive through senses other than our vision, these tend to be grossly underused. For example, while environment awareness is primarily supported using visual techniques, the tacit knowledge of many physical task experts often rely on other senses as

Table 8.1. A visualization of the how different XR capabilities have been leveraged to support the various communication needs.

			Information and Communicative Needs																		
			Environment		Collaborator Awareness								Communicative								
			Verbal Channel	Passive Source	Explicit Representation	Passive Source	Explicit Representation								Communication Process Supported			Type of Mechanism			
				View of the task environment			Presence of task objects	Spatial information of task objects	Secondary Awareness	View of collaborator(s)	Collaborator's location	Collaborator's pose or orientation	Collaborator's view	Collaborator's gaze	Collaborator's facial expressions	Collaborator's action	Secondary awareness		Referencing	Communicating technique	Acknowledgement
Mixed Reality Capabilities	Present multi-sensory virtual content	2D visual content	4				4	1	1	2	1	1	1	1	4	3			4	1	
		3D visual content	3	1			4	3	3	2	1				5	5			6	1	
		Audio signals	1					1					1		1				2		
		Haptic feedback		1			1								1				1		
		Virtual smell and/or taste																			
	Modify user's experience with the real world	Augment the real world	1		1			2	3	3	2	1				5	5			7	1
		Diminish aspects of the real world																			
		Modify the presentation of reality		1				1	1	2		1				1	1			1	
		Block out the real world		3				3													
		Recreate remote locations		3	1			4								1	2			2	
	Enable spatial interfaces and understanding	Spatially distribute content	1	4	1			5	3	3	2	1				5	5			7	1
		Anchor content to the real world						2	3	3	2	1				5	3			5	
		Detect and track objects			1											1	1			2	
		Track user location and orientation						1	3	2	2										
	Enable human understanding	Understand natural language																			
		Detect non-verbal cues							1	2	1	1	1		1						
		Detect physiological state											1		1						
Provide independence of experience	Personalize content and interface																				
	Allow freedom of movement		2				2														
	Allow freedom of perspective		4				3														
	Ability to modify environments														3	2			3		

well – surgeons quite regularly rely on how a particular tissue feels during an operation, and car mechanics often listen for certain sounds to help them diagnose an engine problem. Therefore, incorporating and relaying relevant non-visual information to remote experts could optimize the support level offered for certain tasks. Additionally, even simple tweaks to existing features, like

having more spatial audio, could add to the realism and embodied nature of the collaborative experience. In general, exploring the multi-modal nature of both XR and communication itself is key to uncovering more of the promised potential for rich communication with the technology.

We also see a complete lack of support for acknowledgment and outlining strategy in today's collaborative XR systems. While being able to share external resources or use annotations to provide context (like in Chapter 4) supports outlining strategy, we don't offer any dedicated support for it today. When it comes to acknowledgments, providing confirmations or even subtle indications of confusion regarding each others' actions is an aspect of communication often taken for granted in collocated settings. Subtle behavioral cues like verbal backchannels, a nod, or a smile help the conversation move forward smoothly. A possible future direction for research could be to understand if MR's understanding of user speech, head movements, and additional body language can provide a channel for smoother acknowledgments in remote collaboration.

On that note, our research in chapters 6 and 7 showcased how some features can allow for new forms of acknowledgment as well. Being able to trust that the system would direct novices to the right task object reduced the need for expert acknowledgement and that communication process became more implicit during the collaboration. UnMapped allowed for virtual replicas to be used as a form of non-verbal acknowledgement. However, we were only able to uncover these unexpected advantages of interaction techniques through in-depth evaluations of the communication patterns and collaborator behaviors. Therefore, while it is crucial to engage in foundational research that explores how XR's unique capabilities can shape remote collaboration, researchers should also ensure that thorough human-centered evaluations accompany it to understand how the different features in this new medium play out in practice – helping us understand the extent of support these features provide, when they can be leveraged, and the true gaps that remain.

Building for Sustainable Interactions

Collaborating over physical tasks from a distance, as we've seen, can be an extremely taxing and physically demanding process. Therefore, if we are asking, say, an expert surgeon, to engage in remote collaborative sessions with different novices as a regular part of their work practice, we need to make sure we are designing systems and interactions that can be sustained over time without inducing burnout or extensive fatigue. Designing systems tailored or personalized for different users or user groups is a major research area that can reduce the learning curve and make collaborative systems less frustrating to use for remote experts and novices alike.

Another key research area that is severely under-explored today is the design of systems that make repeated use easier. To the best of our knowledge, the UnMapped approach is the only research effort where a primary design goal was to ease repeated use of the system. Future research should, therefore, look towards building collaborative systems that are designed to be easy to use over time and exploring how user behavior and work practices evolve when collaborative XR systems are used for more than a few sessions. This will be crucial to ensuring that we are moving towards more sustainable collaboration practices that eventually encourage continued adoption, more productivity, and fewer errors as expertise and teamwork become increasingly distributed.

While a majority of our work and the field as a whole focuses on exploring the immersive and 3D nature of these technologies, they remain taxing to use today and aren't always the most practical option in a variety of scenarios. The cost of these technologies today also restricts it from being one that can be ubiquitously distributed across the world. Chapter 6 in this dissertation highlights how simple 2D representations like a map or list of task objects can offer substantial support for complex communication processes like referencing without requiring immersion. It is, therefore, just as important for future researchers to continue exploring 2D representations and communication cues that enhance support for non-immersive remote users to ensure collaborative

systems remain accessible and scalable in a variety of domains and scenarios.

Understanding Domain-Specific Collaborations

An aspect of this dissertation I'd like to highlight is that it would have been incredibly difficult for us to isolate the difficulties of referencing and engage in the subsequent foundational research for it without the domain-specific research that came first (especially in Chapter 4 with trauma surgery). Given the nascent nature of the field, it is understandable how a majority of our research efforts have gone towards building interaction techniques and mechanisms that would work in a variety of scenarios. However, collaborative systems operate within specific social and organizational contexts, and I believe the technology is at a turning point where we can observe it in use for real tasks and generate insights that can advance collaborative XR systems further.

While more general research provides us with a good foundation to stand on, domain-specific research allows us to understand the nuances, complexities, and challenges of different tasks and foster an environment where specialized solutions can be developed. This means that we can tailor approaches that optimize support for physical tasks in a variety of domains and specific contextual factors – which in turn enables the design of systems that align with users' existing work practices and expectations and ensures that we are maximizing the overall impact these systems can have in their respective contexts. In addition to aligning to the intricacies of each domain, domain-specific research also fosters meaningful innovation that would not have otherwise been possible – acting as a catalyst for broader innovation in general.

8.3 Closing Remarks

In conclusion, the research in this dissertation lays the foundations for rich collaborative experiences that go beyond the constraints of face-to-face interactions while making it easier and more sustainable to scale real-world expertise. It offers a conceptual lens and specific design ideas that empower practitioners to create collaborative XR systems that facilitate effective communication and guidance during physical tasks. The work in this dissertation also expands

the design space for collaborative XR and demonstrates how remote physical tasks can be effectively supported without fully replicating physical proximity. By minimizing the experiential overhead for collaborators and leveraging XR's unique capabilities, these novel design choices enhance task performance and communication efficiency and hold the promise of revolutionizing how expertise is shared and accessed across different geographic locations. It is my hope that the takeaways from this dissertation go beyond the specific features explored and speak to the importance of taking a holistic, structured, and human-centered approach to creating more thoughtfully designed collaborative systems that are adaptable and capable of supporting evolving collaborative practices in the real world.

Appendix A

Papers Reviewed

The following papers were reviewed and analyzed based on the conceptual framework for the work presented in chapter 5.

[3, 4, 6, 5, 7, 8, 9, 10, 13, 12, 11, 14, 17, 18, 21, 24, 25, 33, 32, 38, 39, 41, 42, 47, 48, 49, 61, 62, 64, 65, 70, 80, 81, 82, 85, 86, 92, 99, 98, 97, 101, 103, 104, 105, 116, 117, 119, 118, 131, 133, 134, 136, 135, 137, 138, 145, 145, 143, 142, 144, 150, 151, 153, 155, 164, 163, 159, 158, 162, 161, 160, 157, 165, 171, 172, 178, 181, 186, 185, 184, 194, 188, 189, 190, 187, 193, 192, 196, 199, 200, 205, 203, 207, 213, 225, 224, 235, 236, 241, 239, 238, 240, 246, 249, 250, 248, 251, 255, 258, 259, 305, 265, 267, 269, 271, 279, 281, 100, 291, 294, 292, 293, 295, 298, 302, 299, 303, 301, 304, 300, 306, 307, 308, 310, 312, 313, 317, 318, 325, 327, 329, 319, 320, 322, 332, 331, 324, 335, 338, 347, 348, 352, 353, 354, 355]

Appendix B

Helper Task Instructions

The following pages are the instructions helpers received in the task for the experiment outlined in chapter 6.

TASK 1

You will be working together with your partner to complete a task.

Your partner has certain objects in their room (that they don't know of) that can be seen on the aid provided to you. You will be guiding them to complete the task.

This task has two subtasks - **please complete them in order (sequentially).**

SUBTASK 1 - PAIR MATCHING

Your task is to find pairs (**objects with the same picture on them**) among the objects in your partner's room and put them together.

- Each pair has one object with a black background and another object with a white background.
- You must guide your partner to **pick up the white object and place it next to a black one.**



There are 4 pairs.

If there are duplicate objects, you only need to move one.

SUBTASK 2 - OBJECT GATHERING

In your partner's room there is a grid is made up of 6 numbers in the following order:

1	2	3
4	5	6

Your task is to help your partner **find the following objects in the room and place them in their respective positions on the grid.**



TASK 2

You will be working together with your partner to complete a task.

Your partner has certain objects in their room (that they don't know of) that can be seen on the aid provided to you. You will be guiding them to complete the task.

This task has two subtasks - **please complete them in order (sequentially).**

SUBTASK 1 - PAIR MATCHING

Your task is to find pairs (**objects with the same picture on them**) among the objects in your partner's room and put them together.

- Each pair has one object with a black background and another object with a white background.
- You must guide your partner to **pick up the white object and place it next to a black one.**



There are 4 pairs.

If there are duplicate objects, you only need to move one.

SUBTASK 2 - OBJECT GATHERING

In your partner's room there is a grid is made up of 6 numbers in the following order:

1	2	3
4	5	6

Your task is to help your partner **find the following objects in the room and place them in their respective positions on the grid.**



Appendix C

Cake Decorating Task

The following pages show the different steps involved in the cake decorating task for the experiment outlined in chapter 7. The task object used is highlighted in blue.



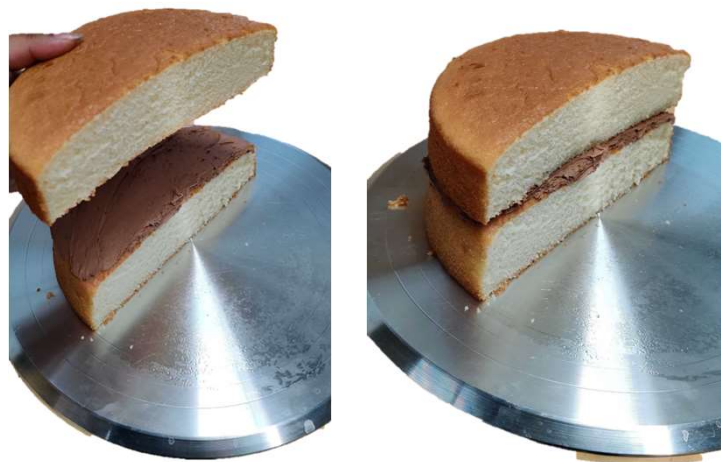
There is a cake layer on a turntable or rotating cake stand.



Cut the cake in the middle using a **serrated knife**.



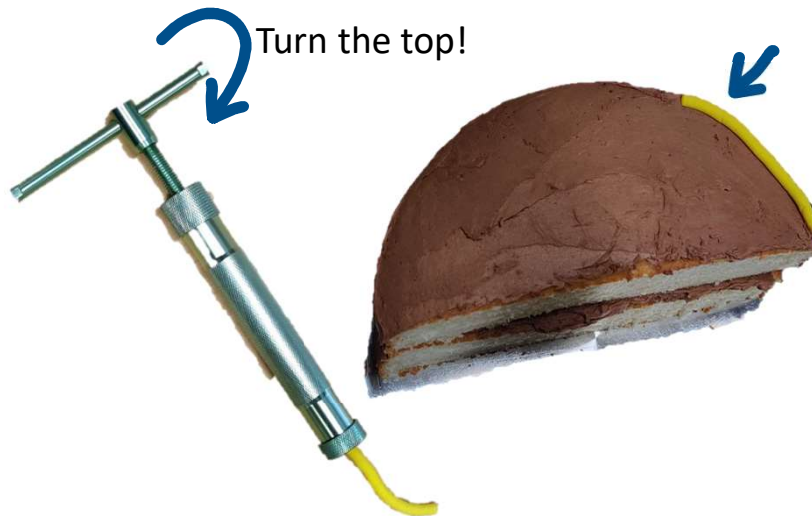
Add a layer of chocolate buttercream on top of one half of the cake with a **rubber spatula**.



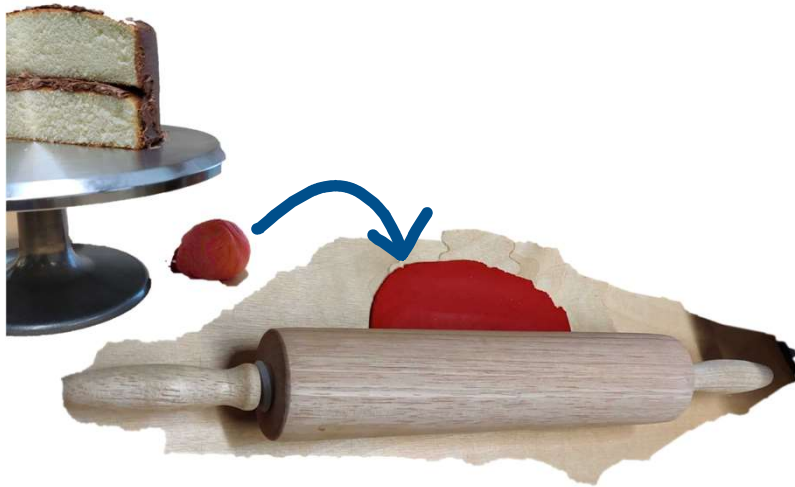
Stack the other half of the cake on top of the one with buttercream.



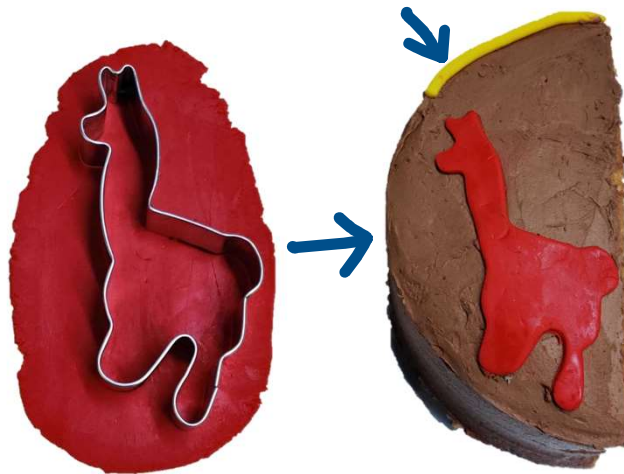
Frost the cake with an **offset spatula**.
You can leave the straight side unfrosted.



Use a **fondant extruder** to get a string of fondant
(a few inches long is fine). Place it on the right
side of the curved edge of the cake.



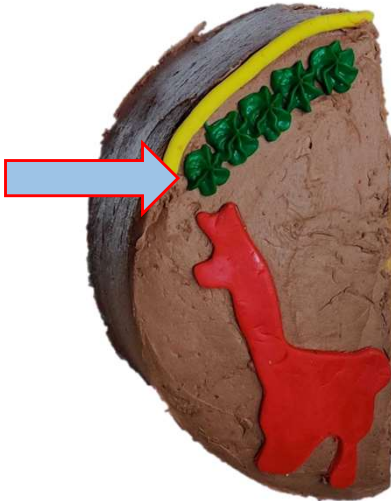
Use a **rolling pin** to flatten the fondant ball into a sheet.



Use a **llama cookie cutter** to add a fondant llama to the cake. Make sure the fondant string is above its head.



Put the **star piping tip** on the piping bag
(unscrew the outer part of the coupler to do so)



Pipe 5 stars under the fondant string



Change the piping tip to the **flower piping tip**.
(unscrew the outer part of the coupler to do so)



Pipe grass under the llama's feet



THE TASK IS DONE

Feel free to add more detail if you want.

Bibliography

- [1] Mark S Ackerman, Christine A Halverson, Thomas Erickson, and Wendy A Kellogg. *Resources, co-evolution and artifacts: Theory in CSCW*. Springer Science & Business Media, 2007.
- [2] Marilyn Jager Adams, Yvette J Tenney, and Richard W Pew. Situation awareness and the cognitive management of complex systems. *Human factors*, 37(1):85–104, 1995.
- [3] Matt Adcock, Stuart Anderson, and Bruce Thomas. Remotefusion: real time depth camera fusion for remote collaboration on physical tasks. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, pages 235–242. ACM, 2013.
- [4] Matt Adcock, David Feng, and Bruce Thomas. Visualization of off-surface 3d viewpoint locations in spatial augmented reality. In *Proceedings of the 1st symposium on Spatial user interaction*, pages 1–8, 2013.
- [5] Matt Adcock and Chris Gunn. Using projected light for mobile remote guidance. *Computer Supported Cooperative Work (CSCW)*, 24:591–611, 2015.
- [6] Matt Adcock, Dulitha Ranatunga, Ross Smith, and Bruce H Thomas. Object-based touch manipulation for remote guidance of physical tasks. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*, pages 113–122, 2014.
- [7] Deepak Akkil, Jobin Mathew James, Poika Isokoski, and Jari Kangas. Gazetorch: Enabling gaze awareness in collaborative physical tasks. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1151–1158, 2016.
- [8] Leila Alem, Franco Tecchia, and Weidong Huang. Handsonvideo: Towards a gesture based mobile ar system for remote collaboration. In *Recent trends of mobile collaborative augmented reality systems*, pages 135–148. Springer, 2011.
- [9] Leila Alem, Franco Tecchia, and Weidong Huang. Remote tele-assistance system for maintenance operators in mines. 2011.
- [10] Judith Amores, Xavier Benavides, and Pattie Maes. Showme: A remote collaboration system that supports immersive gestural communication. In *Proceedings of the 33rd*

Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pages 1343–1348, 2015.

- [11] Dan Andersen, Voicu Popescu, Maria Eugenia Cabrera, Aditya Shanghavi, Brian Mullis, Sherri Marley, Gerardo Gomez, and Juan P Wachs. An augmented reality-based approach for surgical telementoring in austere environments. *Military medicine*, 182(suppl_1):310–315, 2017.
- [12] Daniel Andersen, Voicu Popescu, Maria Eugenia Cabrera, Aditya Shanghavi, Gerardo Gomez, Sherri Marley, Brian Mullis, and Juan Wachs. Virtual annotations of the surgical field through an augmented reality transparent display. *The Visual Computer*, 32(11):1481–1498, 2016.
- [13] Daniel Andersen, Voicu Popescu, Maria Eugenia Cabrera, Aditya Shanghavi, Gerardo Gomez, Sherri Marley, Brian Mullis, and Juan P Wachs. Medical telementoring using an augmented reality transparent display. *Surgery*, 159(6):1646–1653, 2016.
- [14] David Anton, Gregorij Kurillo, and Ruzena Bajcsy. User experience and interaction performance in 2d/3d telecollaboration. *Future Generation Computer Systems*, 82:77–88, 2018.
- [15] Kittipat Apicharttrisorn, Xukan Ran, Jiasi Chen, Srikanth V. Krishnamurthy, and Amit K. Roy-Chowdhury. Frugal following: Power thrifty object detection and tracking for mobile augmented reality. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, SenSys ’19, page 96–109, New York, NY, USA, 2019. Association for Computing Machinery.
- [16] Ferran Argelaguet, Alexander Kulik, André Kunert, Carlos Andujar, and Bernd Froehlich. See-through techniques for referential awareness in collaborative virtual reality. *International Journal of Human-Computer Studies*, 69(6):387–400, 2011.
- [17] Doris Aschenbrenner, Michael Rojkov, Florian Leutert, Jouke Verlinden, Stephan Lukosch, Marc Erich Latoschik, and Klaus Schilling. Comparing different augmented reality support applications for cooperative repair of an industrial robot. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 69–74. IEEE, 2018.
- [18] Huidong Bai, Prasanth Sasikumar, Jing Yang, and Mark Billinghurst. A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13, 2020.
- [19] Martin Bauer, Gerd Kortuem, and Zary Segall. “where are you pointing at?” a study of remote collaboration in a wearable videoconference system. In *Digest of Papers. Third International Symposium on Wearable Computers*, pages 151–158. IEEE, 1999.
- [20] Michel Beaudouin-Lafon, Susanne Bødker, and Wendy E Mackay. Generative theories of interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 28(6):1–54, 2021.

- [21] Xavier Benavides, Judith Amores, and Pattie Maes. Remot-io: a system for reaching into the environment of a remote collaborator. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 99–100, 2015.
- [22] Andrew A Bennett, Emily D Champion, Kathleen R Keeler, and Sheila K Keener. Video-conference fatigue? exploring changes in fatigue after videoconference meetings during covid-19. *Journal of Applied Psychology*, 106(3):330, 2021.
- [23] DK Berlo. An introduction to theory and practice. process of communication. 1960.
- [24] Reinhard Bernsteiner, Christian Ploder, Thomas Dilger, Johannes Nigg, Teresa Spieß, and Rebecca Weichelt. An analysis of an augmented reality application to support service staff in industrial maintenance. In *Knowledge Management in Organizations: 15th International Conference, KMO 2021, Kaohsiung, Taiwan, July 20-22, 2021, Proceedings 15*, pages 457–467. Springer, 2021.
- [25] Mark Billingham, Kunal Gupta, Masai Katsutoshi, Youngho Lee, Gun Lee, Kai Kunze, and Maki Sugimoto. Is it in your eyes? explorations in using gaze cues for remote collaboration. *Collaboration Meets Interactive Spaces*, pages 177–199, 2016.
- [26] Mark Billingham and Hirokazu Kato. Collaborative augmented reality. *Communications of the ACM*, 45(7):64–70, 2002.
- [27] Frank Biocca, Arthur Tang, Charles Owen, and Fan Xiao. Attention funnel: omnidirectional 3d cursor for mobile augmented reality platforms. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1115–1122, 2006.
- [28] Lennart Björneborn. Adjacent possible. In *The Palgrave encyclopedia of the possible*, pages 16–28. Springer, 2023.
- [29] Jonas Blattgerste, Benjamin Streng, Patrick Renner, Thies Pfeiffer, and Kai Essig. Comparing conventional and augmented reality instructions for manual assembly tasks. In *Proc. of the 10th International Conference on Pervasive Technologies Related to Assistive Environments*, pages 75–82, 2017.
- [30] Susanne Bødker. When second wave hci meets third wave challenges. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, pages 1–8, 2006.
- [31] Etai M Bogen, Knut M Augestad, Hiten RH Patel, and Rolv-Ole Lindsetmo. Telementoring in education of laparoscopic surgeons: An emerging technology. *World journal of gastrointestinal endoscopy*, 6(5):148, 2014.
- [32] Sébastien Bottecchia, Jean-Marc Cieutat, and Jean-Pierre Jessel. Tac: Augmented reality system for collaborative tele-assistance in the field of maintenance through internet. In *Proceedings of the 1st Augmented Human International Conference*, pages 1–7, 2010.

- [33] Sébastien Bottecchia, Jean-Marc Cieutat, Christophe Merlo, and Jean-Pierre Jessel. A new AR interaction paradigm for collaborative teleassistance system: the POA. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 3(1):35–40, 2009.
- [34] M. Boukhris, A. Paljic, and D. Lafon-Pham. 360° versus 3d environments in vr headsets for an exploration task. In *Proceedings of the 27th International Conference on Artificial Reality and Telexistence and 22nd Eurographics Symposium on Virtual Environments, ICAT-EGVE '17*, page 71–78, Goslar, DEU, 2017. Eurographics Association.
- [35] Susan E Brennan and Herbert H Clark. Conceptual pacts and lexical choice in conversation. *Journal of experimental psychology: Learning, memory, and cognition*, 22(6):1482, 1996.
- [36] Svend Brinkmann and Lene Tanggaard. Toward an epistemology of the hand. *Studies in philosophy and education*, 29(3):243–257, 2010.
- [37] M Ayodele Buraimoh, Jane Z Liu, Stephen B Sundberg, and Michael P Mott. Eponymous instruments in orthopaedic surgery. *The Iowa orthopaedic journal*, 37:211, 2017.
- [38] Davide Calandra, Alberto Cannavò, and Fabrizio Lamberti. Improving AR-powered remote assistance: a new approach aimed to foster operator’s autonomy and optimize the use of skilled resources. *The International Journal of Advanced Manufacturing Technology*, 114(9):3147–3164, 2021.
- [39] Marina Carbone, Cinzia Freschi, Stefano Mascioli, Vincenzo Ferrari, and Mauro Ferrari. A wearable augmented reality platform for telemedicine. In *Augmented Reality, Virtual Reality, and Computer Graphics: Third International Conference, AVR 2016, Lecce, Italy, June 15-18, 2016. Proceedings, Part II 3*, pages 92–100. Springer, 2016.
- [40] T P Caudell and D W Mizell. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *Proc. HICSS '92*, pages 659–669 vol.2, January 1992.
- [41] Yuan-Chia Chang, Hao-Chuan Wang, Hung-kuo Chu, Shung-Ying Lin, and Shuo-Ping Wang. Alpharead: Support unambiguous referencing in remote collaboration with readable object annotation. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 2246–2259, 2017.
- [42] Georgios Chantziaras, Andreas Triantafyllidis, Aristotelis Papaprodromou, Ioannis Chatzikonstantinou, Dimitrios Giakoumis, Athanasios Tsakiris, Konstantinos Votis, and Dimitrios Tzovaras. An augmented reality-based remote collaboration platform for worker assistance. In *International Conference on Pattern Recognition*, pages 404–416. Springer, 2021.
- [43] Jeffrey W Chastine, Kristine Nagel, Ying Zhu, and Luca Yearsoovich. Understanding the design space of referencing in collaborative augmented reality environments. In *Proceedings of graphics interface 2007*, pages 207–214. ACM, 2007.

- [44] Jeffrey W Chastine, Ying Zhu, and Jon A Preston. A framework for inter-referential awareness in collaborative environments. In *2006 International Conference on Collaborative Computing: Networking, Applications and Worksharing*, pages 1–5. IEEE, 2006.
- [45] Henry Chen, Austin S Lee, Mark Swift, and John C Tang. 3d collaboration method over hololens™ and skype™ end points. In *Proceedings of the 3rd International Workshop on Immersive Media Experiences*, pages 27–30, 2015.
- [46] Long Chen, Thomas W Day, Wen Tang, and Nigel W John. Recent developments and future challenges in medical mixed reality. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 123–135. IEEE, 2017.
- [47] Sicheng Chen, Miao Chen, Andreas Kunz, Asim Evren Yantaç, Mathias Bergmark, Anders Sundin, and Morten Fjeld. Semarbeta: mobile sketch-gesture-video remote support for car drivers. In *Proceedings of the 4th Augmented Human International Conference*, pages 69–76, 2013.
- [48] Sung Ho Choi, Minseok Kim, and Jae Yeol Lee. Situation-dependent remote AR collaborations: Image-based collaboration using a 3d perspective map and live video-based collaboration with a synchronized vr mode. *Computers in Industry*, 101:51–66, 2018.
- [49] Marina Cidota, Stephan Lukosch, Dragos Datcu, and Heide Lukosch. Comparing the effect of audio and visual notifications on workspace awareness using head-mounted displays for remote collaboration in augmented reality. *Augmented Human Research*, 1:1–15, 2016.
- [50] Herbert H Clark. Using language. 1996.
- [51] Herbert H Clark and Susan E Brennan. Grounding in communication. 1991.
- [52] Herbert H Clark and Catherine R Marshall. Definite knowledge and mutual knowledge. 1981.
- [53] Herbert H Clark and Edward F Schaefer. Contributing to discourse. *Cognitive science*, 13(2):259–294, 1989.
- [54] Herbert H Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22(1):1–39, 1986.
- [55] Andy Cockburn and Bruce McKenzie. Evaluating spatial memory in two and three dimensions. *International Journal of Human-Computer Studies*, 61(3):359–373, 2004.
- [56] Benjamin F Crabtree and William L Miller. *Doing qualitative research*. Sage Publications, 1999.
- [57] Lee J Cronbach. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334, 1951.

- [58] Richard L Daft and Robert H Lengel. Organizational information requirements, media richness and structural design. *Management science*, 32(5):554–571, 1986.
- [59] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. Wizard of oz studies: why and how. In *Proceedings of the 1st international conference on Intelligent user interfaces*, pages 193–200, 1993.
- [60] Archi Dasgupta, Mark Manuel, Rifat Sabbir Mansur, Nabil Nowak, and Denis Gračanin. Towards real time object recognition for context awareness in mixed reality: A machine learning approach. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 262–268. IEEE, 2020.
- [61] Dragoş Datcu, Marina Cidota, Stephan Lukosch, David Martinez Oliveira, and Mikael Wolff. Virtual co-location to support remote assistance for inflight maintenance in ground training for space missions. In *Proceedings of the 15th International Conference on Computer Systems and Technologies*, pages 134–141, 2014.
- [62] Dragoş Datcu, Stephan G Lukosch, and Heide K Lukosch. Handheld augmented reality for distributed collaborative crime scene investigation. In *Proceedings of the 19th International Conference on Supporting Group Work*, pages 267–276, 2016.
- [63] Ryan Anthony J de Belen, Huyen Nguyen, Daniel Filonik, Dennis Del Favero, and Tomasz Bednarz. A systematic review of the current state of collaborative mixed reality technologies: 2013–2018. 2019.
- [64] Francesco De Pace, Federico Manuri, Andrea Sanna, and Davide Zappia. A comparison between two different approaches for a collaborative mixed-virtual environment in industrial maintenance. *Frontiers in Robotics and AI*, page 18, 2019.
- [65] Iñigo Fernandez del Amo, John Erkoyuncu, Romain Frayssinet, Cristina Vazquez Reynel, and Rajkumar Roy. Structured authoring for ar-based communication to enhance efficiency in remote diagnosis for complex equipment. *Advanced Engineering Informatics*, 45:101096, 2020.
- [66] Arindam Dey, Mark Billinghurst, Robert W Lindeman, and J Edward Swan. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI*, 5:37, 2018.
- [67] Ersin Dincelli and Alper Yayla. Immersive virtual reality in the age of the metaverse: A hybrid-narrative review based on the technology affordance perspective. *The Journal of Strategic Information Systems*, 31(2):101717, 2022.
- [68] Alan Dix. Computer supported cooperative work: A framework. In *Design issues in CSCW*, pages 9–26. Springer, 1994.
- [69] Gwyneth Doherty-Sneddon, Anne Anderson, Claire O’malley, Steve Langton, Simon Garrod, and Vicki Bruce. Face-to-face and video-mediated communication: A comparison

- of dialogue structure and task performance. *Journal of Experimental Psychology: Applied*, 3(2):105, 1997.
- [70] Veronika Domova, Elina Vartiainen, and Marcus Englund. Designing a remote video collaboration system for industrial settings. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, pages 229–238, 2014.
 - [71] Nicola Döring, Katrien De Moor, Markus Fiedler, Katrin Schoenenberg, and Alexander Raake. Videoconference fatigue: A conceptual analysis. *International journal of environmental research and public health*, 19(4):2061, 2022.
 - [72] Mingsong Dou and Henry Fuchs. Temporally enhanced 3d capture of room-sized dynamic scenes with commodity depth cameras. In *2014 IEEE Virtual Reality (VR)*, pages 39–44. IEEE, IEEE, 2014.
 - [73] Paul Dourish and Victoria Bellotti. Awareness and coordination in shared workspaces. In *CSCW*, volume 92, pages 107–114, 1992.
 - [74] Steven Dow, Jaemin Lee, Christopher Oezbek, Blair MacIntyre, Jay David Bolter, and Maribeth Gandy. Wizard of oz interfaces for mixed reality applications. In *CHI’05 extended abstracts on human factors in computing systems*, pages 1339–1342, 2005.
 - [75] Miguel P Eckstein. Visual search: A retrospective. *Journal of vision*, 11(5):14–14, 2011.
 - [76] Abraham El-Sedfy and Ronald S Chamberlain. Surgeons and their tools: A history of surgical instruments and their innovators—part I: Place the scissors on the mayo stand. *The American Surgeon*, 80(11):1089–1092, 2014.
 - [77] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billingham. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, 131:81–98, 2019.
 - [78] MR Ensley. Toward a theory of situation awareness in dynamic systems. *Human factors*, 37:85–104, 1995.
 - [79] Russell A Epstein, J Stephen Higgins, Karen Jablonski, and Alana M Feiler. Visual scene processing in familiar and unfamiliar environments. *Journal of neurophysiology*, 97(5):3670–3683, 2007.
 - [80] Omid Fakourfar, Kevin Ta, Richard Tang, Scott Bateman, and Anthony Tang. Stabilized annotations for mobile remote assistance. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 1548–1560, 2016.
 - [81] Dikai Fang, Huahu Xu, Xiaoxian Yang, and Minjie Bian. An augmented reality-based method for remote collaborative real-time assistance: from a system perspective. *Mobile Networks and Applications*, 25:412–425, 2020.

- [82] Martin Feick, Anthony Tang, and Scott Bateman. Mixed-reality for object-focused remote collaboration. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*, pages 63–65, 2018.
- [83] Steven Feiner, Blair Macintyre, and Dorée Seligmann. Knowledge-based augmented reality. *Commun. ACM*, 36(7):53–62, July 1993.
- [84] J Finlay and A Dix. Human computer interaction prentice hall. *Hemel Hempstead*, 13(2):218–234, 1993.
- [85] Roland Fischer, Andre Mühlenbrock, Farin Kulapichitr, Verena Nicole Usler, Dirk Weyhe, and Gabriel Zachmann. Evaluation of point cloud streaming and rendering for vr-based telepresence in the or. In *Virtual Reality and Mixed Reality: 19th EuroXR International Conference, EuroXR 2022, Stuttgart, Germany, September 14–16, 2022, Proceedings*, pages 89–110. Springer, 2022.
- [86] Philipp Fleck, Fernando Reyes-Aviles, Christian Pirchheim, Clemens Arth, and Dieter Schmalstieg. Exploring tele-assistance for cyber-physical systems with maui. In *Computer Vision, Imaging and Computer Graphics Theory and Applications: 15th International Joint Conference, VISIGRAPP 2020 Valletta, Malta, February 27–29, 2020, Revised Selected Papers*, pages 431–452. Springer, 2022.
- [87] Nick V Flor. Side-by-side collaboration: A case study. *International Journal of Human-Computer Studies*, 49(3):201–222, 1998.
- [88] Anouschka Foltz, Judith Gaspers, Kristina Thiele, Prisca Stenneken, and Philipp Cimiano. Lexical alignment in triadic communication. *Frontiers in psychology*, 6:127, 2015.
- [89] Adam Fouse, Nadir Weibel, Edwin Hutchins, and James D Hollan. Chronoviz: a system for supporting navigation of time-coded data. In *Proc. CHI’11*, pages 299–304, 2011.
- [90] Olivier Friard and Marco Gamba. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in ecology and evolution*, 7(11):1325–1330, 2016.
- [91] Markus Funk, Andreas Bächler, Liane Bächler, Thomas Kosch, Thomas Heidenreich, and Albrecht Schmidt. Working with augmented reality?: A long-term analysis of in-situ instructions at the assembly workplace. In *Proc. of the 10th International Conference on Pervasive Technologies Related to Assistive Environments*, pages 222–229. ACM, 2017.
- [92] Susan R Fussell, Robert E Kraut, and Jane Siegel. Coordination of communication: Effects of shared visual context on collaborative work. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 21–30. ACM, 2000.
- [93] Susan R Fussell, Leslie D Setlock, and Robert E Kraut. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 513–520. ACM, 2003.

- [94] Susan R Fussell, Leslie D Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam DI Kramer. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*, 19(3):273–309, 2004.
- [95] Joaquín M Fuster. Physiology of executive functions: The perception-action cycle. 2002.
- [96] Brian R Gaines. Modeling and forecasting the information sciences. *Information Sciences*, 57:3–22, 1991.
- [97] Lei Gao, Huidong Bai, Mark Billinghurst, and Robert W Lindeman. User behaviour analysis of mixed reality remote collaboration with a hybrid view interface. In *Proceedings of the 32nd Australian Conference on Human-Computer Interaction*, pages 629–638, 2020.
- [98] Lei Gao, Huidong Bai, Weiping He, Mark Billinghurst, and Robert W Lindeman. Real-time visual representations for mobile mixed reality remote collaboration. In *SIGGRAPH Asia 2018 Virtual & Augmented Reality*, pages 1–2. 2018.
- [99] Lei Gao, Huidong Bai, Gun Lee, and Mark Billinghurst. An oriented point-cloud view for mr remote collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, page 8. ACM, 2016.
- [100] Lei Gao, Huidong Bai, Rob Lindeman, and Mark Billinghurst. Static local environment capturing and sharing for mr remote collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, pages 1–6. 2017.
- [101] Danilo Gasques, Janet G. Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuo-qun Robin Xu, Enrique Zavala, Yifei Zhang, Wanze Xie, Xinming Zhang, Konrad Davis, Michael Yip, and Nadir Weibel. ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI ’21, pages 1–14, New York, NY, USA, May 2021. Association for Computing Machinery.
- [102] Danilo Gasques, Janet G Johnson, Tommy Sharkey, and Nadir Weibel. What you sketch is what you get: Quick and easy augmented reality prototyping with pintar. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2019.
- [103] Steffen Gauglitz, Cha Lee, Matthew Turk, and Tobias Höllerer. Integrating the physical environment into mobile remote collaboration. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 241–250, 2012.
- [104] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, pages 197–205, 2014.

- [105] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 449–459, 2014.
- [106] William W Gaver. Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 79–84, 1991.
- [107] William W Gaver, Abigail Sellen, Christian Heath, and Paul Luff. One is not enough: Multiple views in a media space. In *Proceedings of the INTERACT’93 and CHI’93 Conference on Human Factors in Computing Systems*, pages 335–341. ACM, 1993.
- [108] Darren Gergle, Robert E Kraut, and Susan R Fussell. Action as language in a shared visual space. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 487–496, 2004.
- [109] James J Gibson. The theory of affordances. *Hilldale, USA*, 1(2):67–82, 1977.
- [110] Eduardo Luis Gomes, Mauro Fonseca, André Eugenio Lazzaretti, Anelise Munaretto, and Carlos Guerber. Clustering and hierarchical classification for high-precision rfid indoor location systems. *IEEE Sensors Journal*, 22(6):5141–5149, 2021.
- [111] George O Goodman and Mark J Abel. Communication and collaboration: Facilitating cooperative work through communication. *Office Technology and People*, 3(2):129–145, 1987.
- [112] Robert Graham, Margaret A McCoy, and Andrea M Schultz. The public experience with cardiac arrest. *National Academies Press (US)*, 2015.
- [113] Jens Emil Grønbæk. Designing extended reality experiences for the future of hybrid meetings. *XRDS: Crossroads, The ACM Magazine for Students*, 29(1):20–25, 2022.
- [114] Barbara J Grosz and Candace L Sidner. Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3):175–204, 1986.
- [115] Jonathan Grudin and Steven E Poltrock. Computer-supported cooperative work and groupware. *Advances in computers*, 45:269–320, 1997.
- [116] Sebastian Günther, Sven Kratz, Daniel Avrahami, and Max Mühlhäuser. Exploring audio, visual, and tactile cues for synchronous remote assistance. In *Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference*, pages 339–344, 2018.
- [117] Kunal Gupta, Gun A Lee, and Mark Billinghurst. Do you see what i see? the effect of gaze tracking on task space remote collaboration. *IEEE transactions on visualization and computer graphics*, 22(11):2413–2422, 2016.

- [118] Pavel Gurevich, Joel Lanir, and Benjamin Cohen. Design and implementation of teleadvisor: a projection-based augmented reality system for remote collaboration. *Computer Supported Cooperative Work (CSCW)*, 24(6):527–562, 2015.
- [119] Pavel Gurevich, Joel Lanir, Benjamin Cohen, and Ran Stone. Teleadvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 619–622, 2012.
- [120] Carl Gutwin. The effects of network delays on group work in real-time groupware. In *ECSCW 2001*, pages 299–318. Springer, 2001.
- [121] Carl Gutwin and Saul Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11(3-4):411–446, 2002.
- [122] Sebastian Günther, Sven Kratz, Daniel Avrahami, and Max Mühlhäuser. Exploring Audio, Visual, and Tactile Cues for Synchronous Remote Assistance. In *Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference*, pages 339–344, Corfu Greece, June 2018. ACM.
- [123] Christine A Halverson. Activity theory and distributed cognition: Or what does cscw need to do with theories? *Computer Supported Cooperative Work (CSCW)*, 11(1):243–267, 2002.
- [124] Chad Harms and Frank Biocca. Internal consistency and reliability of the networked minds measure of social presence. In *Seventh annual international workshop: Presence*, volume 2004. Universidad Politecnica de Valencia Valencia, Spain, 2004.
- [125] Sandra G Hart and Lowell E Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.
- [126] Adamantini Hatzipanayioti, Alexia Galati, Marianna Pagkratidou, and Marios N Avraamides. Are spatial memories for familiar environments orientation dependent? *Journal of Cognition*, 4(1), 2021.
- [127] SJ Hayward, Kate van Lopik, Christopher Hinde, and AA West. A survey of indoor location technologies, techniques and applications in industry. *Internet of Things*, page 100608, 2022.
- [128] Eric B Hekler, Predrag Klasnja, Jon E Froehlich, and Matthew P Buman. Mind the theoretical gap: interpreting, using, and developing behavioral theory in hci research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3307–3316, 2013.
- [129] Jason Helbing, Dejan Draschkow, and Melissa L-H Vo. Semantic and syntactic anchor object information interact to make visual search in immersive scenes efficient. *Journal of Vision*, 20(11):573–573, 2020.

- [130] Steven J Henderson and Steven K Feiner. Augmented reality in the psychomotor phase of a procedural task. In *2011 10th IEEE international symposium on mixed and augmented reality*, pages 191–200. IEEE, 2011.
- [131] Keita Higuch, Ryo Yonetani, and Yoichi Sato. Can eye help you? effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5180–5190, 2016.
- [132] Jim Hollan and Scott Stornetta. Beyond being there. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 119–125, 1992.
- [133] Weidong Huang and Leila Alem. Supporting hand gestures in mobile remote collaboration: a usability evaluation. In *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, pages 211–216. British Computer Society, 2011.
- [134] Weidong Huang and Leila Alem. Gesturing in the air: supporting full mobility in remote collaboration on physical tasks. *Journal of Universal Computer Science*, 19(8):1158–1174, 2013.
- [135] Weidong Huang and Leila Alem. Handsinair: a wearable system for remote collaboration on physical tasks. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*, pages 153–156, 2013.
- [136] Weidong Huang, Leila Alem, and Franco Tecchia. Handsin3d: supporting remote guidance with immersive virtual environments. In *IFIP Conference on Human-Computer Interaction*, pages 70–77. Springer, 2013.
- [137] Weidong Huang, Leila Alem, Franco Tecchia, and Henry Been-Lirn Duh. Augmented 3d hands: a gesture-based mixed reality system for distributed collaboration. *Journal on Multimodal User Interfaces*, 12(2):77–89, 2018.
- [138] Weidong Huang, Mark Billingham, Leila Alem, and Seungwon Kim. Handsintouch: Sharing gestures in remote collaboration. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, pages 396–400, 2018.
- [139] Ellen A Isaacs and Herbert H Clark. References in conversation between experts and novices. *Journal of experimental psychology: general*, 116(1):26, 1987.
- [140] Hiroshi Ishii, Minoru Kobayashi, and Kazuho Arita. Iterative design of seamless collaboration media. *Communications of the ACM*, 37(8):83–97, 1994.
- [141] Kumi Ishii, Mary Madison Lyons, and Sabrina A Carr. Revisiting media richness theory for today and future. *Human Behavior and Emerging Technologies*, 1(2):124–131, 2019.
- [142] Allison Jing, Kunal Gupta, Jeremy McDade, Gun A Lee, and Mark Billingham. Comparing gaze-supported modalities with empathic mixed reality interfaces in remote collaboration. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 837–846. IEEE, 2022.

- [143] Allison Jing, Kieran May, Gun Lee, and Mark Billinghurst. Eye see what you see: Exploring how bi-directional augmented reality gaze visualisation influences co-located symmetric collaboration. *Frontiers in Virtual Reality*, 2:697367, 2021.
- [144] Allison Jing, Kieran May, Brandon Matthews, Gun Lee, and Mark Billinghurst. The impact of sharing gaze behaviours in collaborative mixed reality. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–27, 2022.
- [145] Allison Jing, Kieran William May, Mahnoor Naeem, Gun Lee, and Mark Billinghurst. Eyemr-vis: a mixed reality system to visualise bi-directional gaze behavioural cues between remote collaborators. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–4, 2021.
- [146] Michal Joachimczak, Juan Liu, and Hiroshi Ando. Real-time mixed-reality telepresence via 3d reconstruction with hololens and commodity depth sensors. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 514–515. ACM, 2017.
- [147] Janet G Johnson, Danilo Gasques, Tommy Sharkey, Evan Schmitz, and Nadir Weibel. Do you really need to know where “that” is? enhancing support for referencing in collaborative mixed reality environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI ’21, New York, NY, USA, 2021. Association for Computing Machinery.
- [148] Janet G. Johnson, Danilo Gasques Rodrigues, Madhuri Gubbala, and Nadir Weibel. HoloCPR: Designing and evaluating a mixed reality interface for time-critical emergencies. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth ’18, page 67–76, New York, NY, USA, 2018. Association for Computing Machinery.
- [149] Steven Johnson, Madeleine Gibson, and Bilge Mutlu. Handheld or handsfree? remote collaboration via lightweight head-mounted displays and handheld devices. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 1825–1836, 2015.
- [150] Yannik Kalbas, Hoijoon Jung, John Ricklin, Ge Jin, Mingjian Li, Thomas Rauer, Shervin Dehghani, Nassir Navab, Jinman Kim, and Hans-Christoph Pape. Remote interactive surgery platform (RISP): Proof of concept for an augmented-reality-based platform for surgical telementoring. *Journal of Imaging*, 9(3):56, 2023.
- [151] Jari Kangas, Antti Sand, Tero Jokela, Petri Piippo, Peter Eskolin, Marja Salmimaa, and Roope Raisamo. Remote expert for assistance in a physical operational task. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2018.
- [152] Stephen Kaplan. The restorative benefits of nature: Toward an integrative framework. *Journal of environmental psychology*, 15(3):169–182, 1995.

- [153] Shunichi Kasahara and Jun Rekimoto. Jackin: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th augmented human international conference*, pages 1–8, 2014.
- [154] Stuart A Kauffman. *Investigations*. Oxford University Press, 2000.
- [155] Cedric Kervegant, Julien Castet, Juliette Vauchez, and Charles Bailly. Distant assist cursor (DAC): Designing an augmented reality system to facilitate remote collaboration for novice users. In *Companion Proceedings of the 2021 Conference on Interactive Surfaces and Spaces*, pages 8–11, 2021.
- [156] Kourosh Khoshelham and Sander Oude Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *sensors*, 12(2):1437–1454, 2012.
- [157] S Kim, W Huang, CM Oh, G Lee, M Billinghamurst, and SJ Lee. View types and visual communication cues for remote collaboration. *Computers, Materials and Continua*, pages 4363–4379, 2023.
- [158] Seungwon Kim, Mark Billinghamurst, Chilwoo Lee, and Gun Lee. Using freeze frame and visual notifications in an annotation drawing interface for remote collaboration. *KSII Transactions on Internet and Information Systems (TIIS)*, 12(12):6034–6056, 2018.
- [159] Seungwon Kim, Mark Billinghamurst, and Gun Lee. The effect of collaboration styles and view independence on video-mediated remote collaboration. *Computer Supported Cooperative Work (CSCW)*, 27:569–607, 2018.
- [160] Seungwon Kim, Allison Jing, Hanhoon Park, Soo-hyung Kim, Gun Lee, and Mark Billinghamurst. Use of gaze and hand pointers in mixed reality remote collaboration. In *The 9th International Conference on Smart Media and Applications. SMA, Jeju, Republic of Korea*, pages 1–6, 2020.
- [161] Seungwon Kim, Allison Jing, Hanhoon Park, Gun A Lee, Weidong Huang, and Mark Billinghamurst. Hand-in-air (hia) and hand-on-target (hot) style gesture cues for mixed reality collaboration. *IEEE Access*, 8:224145–224161, 2020.
- [162] Seungwon Kim, Gun Lee, Weidong Huang, Hayun Kim, Woontack Woo, and Mark Billinghamurst. Evaluating the combination of visual communication cues for hmd-based mixed reality remote collaboration. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–13, 2019.
- [163] Seungwon Kim, Gun Lee, Nobuchika Sakata, and Mark Billinghamurst. Improving co-presence with augmented visual communication cues for sharing experience through video conference. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 83–92. IEEE, 2014.
- [164] Seungwon Kim, Gun A Lee, Nobuchika Sakata, Andreas Dünser, Elina Vartiainen, and Mark Billinghamurst. Study of augmented gesture communication cues and view sharing in

- remote collaboration. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 261–262. IEEE, 2013.
- [165] David Kirk, Tom Rodden, and Danaë Stanton Fraser. Turn it this way: grounding collaborative action with remote gestures. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 1039–1048, 2007.
 - [166] David Kirk and Danae Stanton Fraser. Comparing remote gesture technologies for supporting collaborative physical tasks. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1191–1200. ACM, 2006.
 - [167] David Kirsh. The intelligent use of space. *Artificial intelligence*, 73(1-2):31–68, 1995.
 - [168] David Kirsh. Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(1):1–30, 2013.
 - [169] David Kirsh and Paul Maglio. On distinguishing epistemic from pragmatic action. *Cognitive science*, 18(4):513–549, 1994.
 - [170] Petrus EJ Kivi, Markku J Mäkitalo, Jakub Žádník, Julius Ikkala, Vinod Kumar Malamal Vadakital, and Pekka O Jääskeläinen. Real-time rendering of point clouds with photorealistic effects: A survey. *IEEE Access*, 10:13151–13173, 2022.
 - [171] Michael Kleiber, Thomas Alexander, Carsten Winkelholz, and Christopher M Schlick. User-centered design and evaluation of an integrated ar-vr system for tele-maintenance. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1443–1448. IEEE, 2012.
 - [172] Jan Kolkmeier, Emiel Harmsen, Sander Giesselink, Dennis Reidsma, Mariët Theune, and Dirk Heylen. With a little help from a holographic friend: The openimpress mixed reality telepresence toolkit for remote collaboration systems. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pages 1–11, 2018.
 - [173] Ivor Kovic and Ileana Lulic. Mobile phone in the chain of survival. *Resuscitation*, 82(6):776–779, 2011.
 - [174] Robert E Kraut, Susan R Fussell, and Jane Siegel. Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*, 18(1-2):13–49, 2003.
 - [175] Robert E Kraut, Darren Gergle, and Susan R Fussell. The use of visual information in shared visual spaces: Informing the development of virtual co-presence. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 31–40, 2002.
 - [176] Robert E Kraut, Mark D Miller, and Jane Siegel. Collaboration in performance of physical tasks: Effects on outcomes and communication. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pages 57–66. ACM, 1996.

- [177] Gregorij Kurillo, Evan Hemingway, Mu-Lin Cheng, and Louis Cheng. Evaluating the accuracy of the azure kinect and kinect v2. *Sensors*, 22(7):2469, 2022.
- [178] Gregorij Kurillo, Allen Y Yang, Victor Shia, Aaron Bair, and Ruzena Bajcsy. New emergency medicine paradigm via augmented telemedicine. In *Virtual, Augmented and Mixed Reality: 8th International Conference, VAMR 2016, Held as Part of HCI International 2016, Toronto, Canada, July 17-22, 2016. Proceedings* 8, pages 502–511. Springer, 2016.
- [179] Hideaki Kuzuoka. Spatial workspace collaboration: a sharedview video support system for remote collaboration capability. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 533–540, 1992.
- [180] Yeram Kwon, Sungwon Lee, Jihoon Jeong, and Wonjoon Kim. Heartisense: a novel approach to enable effective basic life support training without an instructor. In *Proc. CHI '14*, pages 1699–1704, 2014.
- [181] Joel Lanir, Ran Stone, Benjamin Cohen, and Pavel Gurevich. Ownership and control of point of view in remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2243–2252, 2013.
- [182] Mary P Larsen, Mickey S Eisenberg, Richard O Cummins, and Alfred P Hallstrom. Predicting survival from out-of-hospital cardiac arrest: a graphic model. *Annals of emergency medicine*, 22(11):1652–1658, 1993.
- [183] Rifat Latifi, RS Weinstein, JM Porter, M Ziemba, D Judkins, D Ridings, R Nassi, T Valenzuela, M Holcomb, and F Leyva. Telemedicine and telepresence for trauma and emergency care management. *Scandinavian journal of Surgery*, 96(4):281–289, 2007.
- [184] Khanh-Duy Le, Kening Zhu, and Morten Fjeld. Mirrortablet: exploring a low-cost mobile system for capturing unmediated hand gestures in remote collaboration. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, pages 79–89, 2017.
- [185] Morgan Le Chénéchal, Thierry Duval, Valérie Gouranton, Jérôme Royan, and Bruno Arnaldi. Vishnu: virtual immersive support for helping users an interaction paradigm for collaborative remote guiding in mixed reality. In *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*, pages 9–12. IEEE, 2016.
- [186] Morgan Le Chnchal, Thierry Duval, Valrie Gouranton, Jrôme Royan, and Bruno Arnaldi. The stretchable arms for collaborative remote guiding. In *Proceedings of International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments*, 2015.
- [187] Geonsun Lee, HyeongYeop Kang, JongMin Lee, and JungHyun Han. A user study on view-sharing techniques for one-to-many mixed reality collaborations. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 343–352. IEEE, 2020.

- [188] Gun Lee, Seungwon Kim, Youngho Lee, Arindam Dey, Thammathip Piumsomboon, Mitchell Norman, and Mark Billinghurst. Improving collaboration in augmented video conference using mutually shared gaze. In *Proceedings of the 27th International Conference on Artificial Reality and Telexistence and 22nd Eurographics Symposium on Virtual Environments*, pages 197–204, 2017.
- [189] Gun A Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. Sharedsphere: Mr collaboration through shared live panorama. In *SIGGRAPH Asia 2017 Emerging Technologies*, pages 1–2. 2017.
- [190] Gun A Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. A user study on mr remote collaboration using live 360 video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 153–164. IEEE, 2018.
- [191] Stephen G Lee, Y-S Ma, Georg L Thimm, and J Verstraeten. Product lifecycle management in aviation maintenance, repair and overhaul. *Computers in industry*, 59(2-3):296–303, 2008.
- [192] Yongjae Lee and Byounghyun Yoo. Xr collaboration beyond virtual reality: Work in the real world. *Journal of Computational Design and Engineering*, 8(2):756–772, 2021.
- [193] Yongjae Lee, Byounghyun Yoo, and Soo-Hong Lee. Sharing ambient objects using real-time point cloud streaming in web-based xr remote collaboration. In *The 26th International Conference on 3D Web Technology*, pages 1–9, 2021.
- [194] Youngho Lee, Katsutoshi Masai, Kai Kunze, Maki Sugimoto, and Mark Billinghurst. A remote collaboration system with empathy glasses. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 342–343. IEEE, 2016.
- [195] Chia-Ling Li, M Pilar Aivar, Dmitry M Kit, Matthew H Tong, and Mary M Hayhoe. Memory and visual search in naturalistic 2d and 3d environments. *Journal of vision*, 16(8):9–9, 2016.
- [196] Chengyuan Lin, Daniel Andersen, Voicu Popescu, Edgar Rojas-Munoz, Maria Eugenia Cabrera, Brian Mullis, Ben Zarzaur, Kathryn Anderson, Sherri Marley, and Juan Wachs. A first-person mentee second-person mentor ar interface for surgical telementoring. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 3–8. IEEE, 2018.
- [197] Diane J Litman and James F Allen. Discourse processing and commonsense plans. *Intentions in communication*, pages 365–388, 1990.
- [198] Luyang Liu, Hongyu Li, and Marco Gruteser. Edge assisted real-time object detection for mobile augmented reality. In *The 25th annual international conference on mobile computing and networking*, pages 1–16, 2019.

- [199] Stephan Lukosch, Heide Lukosch, Dragoş Datcu, and Marina Cidota. Providing information on the spot: Using augmented reality for situational awareness in the security domain. *Computer Supported Cooperative Work (CSCW)*, 24:613–664, 2015.
- [200] Tiago Madeira, Bernardo Marques, João Alves, Paulo Dias, and Beatriz Sousa Santos. Exploring annotations and hand tracking in augmented reality for remote collaboration. In *Human Systems Engineering and Design III: Proceedings of the 3rd International Conference on Human Systems Engineering and Design (IHSED2020): Future Trends and Applications, September 22-24, 2020, Juraj Dobrila University of Pula, Croatia 3*, pages 83–89. Springer, 2021.
- [201] Ana Malta, Mateus Mendes, and Torres Farinha. Augmented reality maintenance assistant using yolov5. *Applied Sciences*, 11(11):4758, 2021.
- [202] Trevor HJ Marchand. Embodied cognition and communication: studies with british fine woodworkers. *Journal of the Royal Anthropological Institute*, 16:S100–S120, 2010.
- [203] Bernardo Marques, Carlos Ferreira, Samuel Silva, Paulo Dias, and Beatriz Sousa Santos. Is social presence (alone) a general predictor for good remote collaboration? comparing video and augmented reality guidance in maintenance procedures. *Virtual Reality*, pages 1–14, 2023.
- [204] Bernardo Marques, Samuel Silva, Paulo Dias, and Beatriz Sousa-Santos. An ontology for evaluation of remote collaboration using augmented reality. *Computer Supported Cooperative Work*, 2021.
- [205] Bernardo Marques, Samuel Silva, Paulo Dias, and Beatriz Sousa-Santos. Does size matter? exploring how standard and large-scale displays affect off-site experts during ar-remote collaboration. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*, pages 1–3, 2022.
- [206] Alejandro Martin-Gomez, Haowei Li, Tianyu Song, Sheng Yang, Guangzhi Wang, Hui Ding, Nassir Navab, Zhe Zhao, and Mehran Armand. Sttar: Surgical tool tracking using off-the-shelf augmented reality head-mounted displays. *arXiv preprint arXiv:2208.08880*, 2022.
- [207] Riccardo Masoni, Francesco Ferrise, Monica Bordegoni, Michele Gattullo, Antonio E Uva, Michele Fiorentino, Ernesto Carrabba, and Michele Di Donato. Supporting remote maintenance in industry 4.0 through augmented reality. *Procedia manufacturing*, 11:1296–1302, 2017.
- [208] Charlotta Mellander and Richard Florida. The rise of skills: Human capital, the creative class, and regional development. *Handbook of regional science*, pages 707–719, 2021.
- [209] Helena M Mentis, Yuanyuan Feng, Azin Semsar, and Todd A Ponsky. Remotely shaping the view in surgical telementoring. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.

- [210] Microsoft. Overview of dynamics 365 remote assist on hololens and hololens 2, April 2020.
- [211] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE Trans. on Inf. Syst.*, 77(12):1321–1329, 1994.
- [212] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telem manipulator and telepresence technologies*, volume 2351, pages 282–292. International Society for Optics and Photonics, 1995.
- [213] Peter Mohr, Shohei Mori, Tobias Langlotz, Bruce H. Thomas, Dieter Schmalstieg, and Denis Kalkofen. Mixed reality light fields for interactive remote assistance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI ’20, page 1–12, New York, NY, USA, 2020. Association for Computing Machinery.
- [214] Dariush Mozaffarian and Emelia J. Benjamin. Heart disease and stroke statistics—2015 update. *Circulation*, 2016.
- [215] Jens Müller, Roman Rädle, and Harald Reiterer. Remote collaboration with mixed reality displays: How shared virtual landmarks facilitate spatial referencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 6481–6486, 2017.
- [216] Bonnie A Nardi, Heinrich Schwarz, Allan Kuchinsky, Robert Leichner, Steve Whittaker, and Robert Sciabassi. Turning away from talking heads: The use of video-as-data in neurosurgery. In *Proceedings of the INTERACT’93 and CHI’93 Conference on Human Factors in Computing Systems*, pages 327–334, 1993.
- [217] Albert Newen, Shaun Gallagher, and Leon De Bruin. 4E Cognition: Historical Roots, Key Concepts, and Central Issues. pages 3–16, 09 2018.
- [218] Anne Nielsen, Dan Isbye, Freddy Lippert, and Lars Rasmussen. Can mass education and a television campaign change the attitudes towards cardiopulmonary resuscitation in a rural community? *SJTREM*, 21(1):39, 2013.
- [219] Lars Niklasson, Maria Riveiro, Fredrik Johansson, Anders Dahlbom, Goran Falkman, Tom Ziemke, Christoffer Brax, Thomas Kronhamn, Martin Smedberg, and Hakan Warston. Extending the scope of situation analysis. In *2008 11th international conference on information fusion*, pages 1–8. IEEE, 2008.
- [220] Jerry Nolan, Jasmeet Soar, and Harald Eikeland. The chain of survival. *Resuscitation*, 71(3):270–271, 2006.
- [221] Donald A Norman. *The psychology of everyday things*. Basic books, 1988.
- [222] Donald A Norman. Design principles for cognitive artifacts. *Research in Engineering Design*, 4(1):43–50, 1992.

- [223] Donald A. Norman and Stephen W. Draper. *User Centered System Design; New Perspectives on Human-Computer Interaction*. L. Erlbaum Associates Inc., 1986.
- [224] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 405–415, 2015.
- [225] Ohan Oda, Mengu Sukan, Steven Feiner, and Barbara Tversky. Poster: 3d referencing for remote task assistance in augmented reality. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 179–180. IEEE, 2013.
- [226] Gary M Olson and Judith S Olson. Distance matters. *Human-computer interaction*, 15(2-3):139–178, 2000.
- [227] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, and Mingsong Dou. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 741–754, 2016.
- [228] Jiazhi Ou, Susan R Fussell, Xilin Chen, Leslie D Setlock, and Jie Yang. Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 242–249. ACM, 2003.
- [229] Antti Oulasvirta, Esko Kurvinen, and Tomi Kankainen. Understanding contexts by being there: case studies in bodystorming. *Personal and ubiquitous computing*, 7(2):125–134, 2003.
- [230] Matthew J Page, David Moher, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, and Sue E Brennan. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *British Medical Journal*, 372, 2021.
- [231] Cristian Pamparău and Radu-Daniel Vatavu. A research agenda is needed for designing for the user experience of augmented and mixed reality: A position paper. In *19th International Conference on Mobile and Ubiquitous Multimedia*, pages 323–325, 2020.
- [232] Nohyoung Park, Yeram Kwon, Sungwon Lee, Woontack Woo, and Jihoon Jeong. Projected AR-Based Interactive CPR Simulator. In *Proc. VAMR '13*, pages 83–89, 2013.
- [233] Olivier Pauly, Benoit Diotte, Pascal Fallavollita, Simon Weidert, Ekkehard Euler, and Nassir Navab. Machine learning-based augmented reality for improved surgical scene understanding. *Computerized Medical Imaging and Graphics*, 41:55–60, 2015.
- [234] Thomas Pechmann and Werner Deutsch. The development of verbal and nonverbal devices for reference. *Journal of experimental child psychology*, 34(2):330–341, 1982.

- [235] Jana Pejoska-Laajola, Sanna Reponen, Marjo Virnes, and Teemu Leinonen. Mobile augmented communication for remote collaboration in a physical work context. *Australasian Journal of Educational Technology*, 33(6), 2017.
- [236] Tomislav Pejisa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, pages 1716–1725, 2016.
- [237] Thammathip Piumsomboon, Arindam Day, Barrett Ens, Youngho Lee, Gun Lee, and Mark Billinghurst. Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, pages 1–5. 2017.
- [238] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI*, 6(5):02, 2019.
- [239] Thammathip Piumsomboon, Gun A Lee, Jonathon D Hart, Barrett Ens, Robert W Lindeman, Bruce H Thomas, and Mark Billinghurst. Mini-me: an adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, page 46. ACM, 2018.
- [240] Thammathip Piumsomboon, Gun A Lee, Andrew Irlitti, Barrett Ens, Bruce H Thomas, and Mark Billinghurst. On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–17, 2019.
- [241] Thammathip Piumsomboon, Youngho Lee, Gun Lee, and Mark Billinghurst. Covar: a collaborative virtual and augmented reality system for remote collaboration. In *SIGGRAPH Asia 2017 Emerging Technologies*, pages 1–2. 2017.
- [242] Ronald Poelman, Oytun Akman, Stephan Lukosch, and Pieter Jonker. As if being there: mediated reality for crime scene investigation. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 1267–1276, 2012.
- [243] Steven Poltrock and Jonathan Grudin. Cscw, groupware and workflow: experiences, state of art, and future trends. In *CHI 98 Conference Summary on Human Factors in Computing Systems*, pages 119–120, 1998.
- [244] Michael Prilla. “I simply watched where she was looking at” coordination in short-term synchronous cooperative mixed reality. *Proceedings of the ACM on Human-Computer Interaction*, 3(GROUP):1–21, 2019.
- [245] Joseph Psotka. Immersive training systems: Virtual reality and education and training. *Instructional science*, 23(5):405–431, 1995.

- [246] Dulitha Ranatunga, David Feng, Matt Adcock, and Bruce Thomas. Towards object based manipulation in remote guidance. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1–6. IEEE, 2013.
- [247] Abhishek Ranjan, Jeremy P Birnholtz, and Ravin Balakrishnan. An exploratory analysis of partner action and camera control in a video-mediated collaborative task. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 403–412. ACM, 2006.
- [248] Troels Rasmussen, Tiare Feuchtner, Weidong Huang, and Kaj Grønæk. Supporting workspace awareness in remote assistance through a flexible multi-camera system and augmented reality awareness cues. *Journal of Visual Communication and Image Representation*, 89:103655, 2022.
- [249] Troels A Rasmussen and Weidong Huang. Scenecam: Using ar to improve multi-camera remote collaboration. In *SIGGRAPH Asia 2019 XR*, pages 36–37. 2019.
- [250] Troels Ammitsbøl Rasmussen and Weidong Huang. Scenecam: Improving multi-camera remote collaboration using augmented reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 28–33. IEEE, 2019.
- [251] Manuel Rebol, Claudia Ranniger, Colton Hood, Erin Horan, Adam Rutenberg, Neal Sikka, Yasser Ajabnoor, Safinaz Alshikah, and Krzysztof Pietroszek. Mixed reality communication system for procedural tasks. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*, pages 1–3, 2022.
- [252] Charles Rich, Candace L Sidner, and Neal Lesh. Collagen: Applying collaborative discourse theory to human-computer interaction. *AI magazine*, 22(4):15–15, 2001.
- [253] J David Richardson. General surgeon shortage in the united states: fact or fiction, causes and consequences. *Social work in public health*, 26(5):513–523, 2011.
- [254] Mauro Del Rio, Vittorio Meloni, Francesca Frexia, Francesco Cabras, Roberto Tumbarello, Sabrina Montis, Andrea Marini, and Gianluigi Zanetti. Augmented reality for supporting real time telementoring: an exploratory study applied to ultrasonography. In *Proceedings of the 2nd International Conference on Medical and Health Informatics*, pages 218–222, 2018.
- [255] Kostia Robert, Dingyun Zhu, Weidong Huang, Leila Alem, and Tom Gedeon. Mobile-helper: remote guiding using smart mobile devices, hand gestures and augmented reality. In *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications*, pages 1–5, 2013.
- [256] Jessica Robin and Morris Moscovitch. The effects of spatial contextual familiarity on remembered scenes, episodic memories, and imagined future events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2):459, 2014.

- [257] Yvonne Rogers, Liam Bannon, and Graham Button. Rethinking theoretical frameworks for hci: report on an interchi'93 workshop, amsterdam, 24–25th april, 1993. *ACM SIGCHI Bulletin*, 26(1):28–30, 1994.
- [258] Edgar Rojas-Muñoz, Dan Andersen, Maria Eugenia Cabrera, Voicu Popescu, Sherri Marley, Ben Zarzaur, Brian Mullis, and Juan P Wachs. Augmented reality as a medium for improved telementoring. *Military medicine*, 184(Supplement_1):57–64, 2019.
- [259] Edgar Rojas-Muñoz, Maria E Cabrera, Chengyuan Lin, Daniel Andersen, Voicu Popescu, Kathryn Anderson, Ben L Zarzaur, Brian Mullis, and Juan P Wachs. The system for telementoring with augmented reality (star): A head-mounted display to improve surgical coaching and confidence in remote areas. *Surgery*, 2020.
- [260] Edgar Rojas-Muñoz, Maria Eugenia Cabrera, Chengyuan Lin, Natalia Sánchez-Tamayo, Dan Andersen, Voicu Popescu, Kathryn Anderson, Ben Zarzaur, Brian Mullis, and Juan P Wachs. Telementoring in leg fasciotomies via mixed-reality: Clinical evaluation of the star platform. *Military Medicine*, 185(Supplement_1):513–520, 2020.
- [261] Somaieh Rokhsaritalemi, Abolghasem Sadeghi-Niaraki, and Soo-Mi Choi. A review on mixed reality: Current trends, challenges and prospects. *Applied Sciences*, 10(2):636, 2020.
- [262] Wolff-Michael Roth and Alfredo Jornet. Situated cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(5):463–478, 2013.
- [263] Andrea Sanna and Federico Manuri. A survey on applications of augmented reality. *Advances in Computer Science: an International Journal*, 5(1):18–27, 2016.
- [264] Bhuvaneswari Sarupuri, Gun A Lee, and Mark Billingham. Using augmented reality to assist forklift operation. In *Proc. OzCHI '16*, pages 16–24, 2016.
- [265] Prasanth Sasikumar, Lei Gao, Huidong Bai, and Mark Billingham. Wearable remotefusion: A mixed reality remote collaboration system with local eye gaze and remote hand gesture sharing. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 393–394. IEEE, 2019.
- [266] Joey Scarr, Andy Cockburn, and Carl Gutwin. Supporting and exploiting spatial memory in user interfaces. *Foundations and Trends® in Human–Computer Interaction*, 6(1):1–84, 2013.
- [267] Giuseppe Scavo, Fridolin Wild, and Peter Scott. The ghosthands ux: telementoring with hands-on augmented reality instruction. In *Workshop Proceedings of the 11th International Conference on Intelligent Environments*, pages 236–243. IOS Press, 2015.
- [268] Wilbur Ed Schram. The process and effects of mass communication. 1954.

- [269] Jwawon Seo, Ignacio Avellino, Damaruka Priya Rajasagi, Anita Komlodi, and Helena M Mentis. Holomenter: Enabling remote instruction through augmented reality mobile views. *Proceedings of the ACM on Human-Computer Interaction*, 7(GROUP):1–29, 2023.
- [270] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.
- [271] Mahesh B Shenai, Marcus Dillavou, Corey Shum, Douglas Ross, Richard S Tubbs, Alan Shih, and Barton L Guthrie. Virtual interactive presence and augmented reality (vipar) for remote surgical assistance. *Operative Neurosurgery*, 68(suppl.1):ons200–ons207, 2011.
- [272] Tobias Sielhorst, Tobias Obst, Rainer Burgkart, Robert Riener, and Nassir Navab. An augmented reality delivery simulator for medical training. In *Proc. AMI-ARCS '04*, volume 141, pages 11–20, 2004.
- [273] Richard Skarbez, Frederick P Brooks, Jr, and Mary C Whitton. A survey of presence and related concepts. *ACM Computing Surveys (CSUR)*, 50(6):1–39, 2017.
- [274] Richard Skarbez, Missie Smith, and Mary C Whitton. Revisiting milgram and kishino’s reality-virtuality continuum. *Frontiers in Virtual Reality*, 2:647997, 2021.
- [275] Mel Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.
- [276] Mel Slater, Domna Banakou, Alejandro Beacco, Jaime Gallego, Francisco Macia-Varela, and Ramon Oliva. A separate reality: An update on place illusion and plausibility in virtual reality. *Frontiers in Virtual Reality*, 3:914392, 2022.
- [277] Mel Slater, Vasilis Linakis, Martin Usoh, and Rob Kooper. Immersion, presence and performance in virtual environments: An experiment with tri-dimensional chess. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST ’96, page 163–172, New York, NY, USA, 1996. Association for Computing Machinery.
- [278] Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6(6):603–616, 1997.
- [279] Rajinder S Sodhi, Brett R Jones, David Forsyth, Brian P Bailey, and Giuliano Maciocci. BeThere: 3d mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 179–188, 2013.
- [280] Maurício Sousa, Rafael Kufner dos Anjos, Daniel Mendes, Mark Billinghurst, and Joaquim Jorge. Warping deixis: Distorting gestures to enhance collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

- [281] Maximilian Speicher, Jingchen Cao, Ao Yu, Haihua Zhang, and Michael Nebeling. 360anywhere: Mobile ad-hoc collaboration in any environment using 360 video and augmented reality. *Proceedings of the ACM on Human-Computer Interaction*, 2(EICS):1–20, 2018.
- [282] Maximilian Speicher, Brian D Hall, and Michael Nebeling. What is mixed reality? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2019.
- [283] Michael B Spring and Vichita Vathanophas. Peripheral social awareness information in collaborative work. *Journal of the American Society for Information Science and Technology*, 54(11):1006–1013, 2003.
- [284] Aaron Stafford, Wayne Piekarski, and Bruce H Thomas. Implementation of god-like interaction techniques for supporting collaboration between outdoor ar and indoor tabletop users. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 165–172. IEEE, 2006.
- [285] Neville A Stanton, Peter RG Chambers, and John Piggott. Situational awareness and safety. *Safety science*, 39(3):189–204, 2001.
- [286] Ohan Oda Carmine Elvezio Mengu Sukan, Steven Feiner, and Barbara Tversky. Virtual replicas for remote assistance in virtual and augmented reality. 2015.
- [287] Dag Svanaes and Gry Seland. Putting the users center stage: role playing and low-fi prototyping enable end users to design mobile systems. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 479–486, 2004.
- [288] J Edward Swan and Joseph L Gabbard. Survey of user-based experimentation in augmented reality. In *Proc. 1st Intl. Conf. on VR*, pages 1–9, 2005.
- [289] Zsolt Szalavári, Dieter Schmalstieg, Anton Fuhrmann, and Michael Gervautz. “studierstube”: An environment for collaboration in augmented reality. *Virtual Reality*, 3(1):37–48, 1998.
- [290] Gabor Sziebig, I Rudas, M Demiralp, and N Mastorakis. Achieving total immersion: Technology trends behind augmented reality- a survey. In *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering*, number 5. WSEAS, 2009.
- [291] Matthew D Tadlock, Erik J Olson, Danilo Gasques, Roland Champagne, Michael J Krzyzaniak, Shawn A Belverud, Vijay Ravindra, Jakob Kerns, Pamela M Choi, and Jennifer Deveraux. Mixed reality surgical mentoring of combat casualty care related procedures in a perfused cadaver model: Initial results of a randomized feasibility study. *Surgery*, 172(5):1337–1345, 2022.

- [292] Matthew Tait and Mark Billinghurst. View independence in remote collaboration using ar. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 309–310. IEEE, 2014.
- [293] Matthew Tait and Mark Billinghurst. The effect of view independence in a collaborative ar system. *Computer Supported Cooperative Work (CSCW)*, 24(6):563–589, 2015.
- [294] Matthew Tait, Tony Tsai, Nobuchika Sakata, Mark Billinghurst, and Elina Vartiainen. A projected augmented reality system for remote collaboration. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1–6. IEEE, 2013.
- [295] Keisuke Tajimi, Nobuchika Sakata, Keiji Uemura, and Shogo Nishida. Remote collaboration using real-world projection interface. In *2010 IEEE International Conference on Systems, Man and Cybernetics*, pages 3008–3013. IEEE, 2010.
- [296] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. Comparative effectiveness of augmented reality in object assembly. In *Proc. CHI '03*, pages 73–80, 2003.
- [297] John C Tang. Findings from observational studies of collaborative work. *International Journal of Man-machine studies*, 34(2):143–160, 1991.
- [298] Franco Tecchia, Leila Alem, and Weidong Huang. 3d helping hands: a gesture based mr system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*, pages 323–328, 2012.
- [299] Theophilus Teo, Gun A. Lee, Mark Billinghurst, and Matt Adcock. 360drops: Mixed reality remote collaboration using 360 panoramas within the 3d scene. In *SIGGRAPH Asia 2019 Emerging Technologies*, pages 1–2. 2019.
- [300] Theophilus Teo, Ashkan F. Hayati, Gun A. Lee, Mark Billinghurst, and Matt Adcock. A technique for mixed reality remote collaboration using 360 panoramas in 3d reconstructed scenes. In *25th ACM Symposium on Virtual Reality Software and Technology*, pages 1–11, 2019.
- [301] Theophilus Teo, Louise Lawrence, Gun A Lee, Mark Billinghurst, and Matt Adcock. Mixed reality remote collaboration combining 360 video and 3d reconstruction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–14, 2019.
- [302] Theophilus Teo, Gun A Lee, Mark Billinghurst, and Matt Adcock. Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, pages 406–410, 2018.
- [303] Theophilus Teo, Gun A Lee, Mark Billinghurst, and Matt Adcock. Merging live and static 360 panoramas inside a 3d scene for mixed reality remote collaboration. In *2019 IEEE*

International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pages 22–25. IEEE, 2019.

- [304] Theophilus Teo, Gun A Lee, Mark Billinghurst, and Matt Adcock. Supporting visual annotation cues in a live 360 panorama-based mixed reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1187–1188. IEEE, IEEE, 2019.
- [305] Theophilus Teo, Mitchell Norman, Gun A Lee, Mark Billinghurst, and Matt Adcock. Exploring interaction techniques for 360 panoramas inside a 3d reconstructed scene for mixed reality remote collaboration. *Journal on Multimodal User Interfaces*, 14:373–385, 2020.
- [306] Frank ter Haar, Sylvie Dijkstra-Soudarissanane, Piotr Zuraniewski, Rick Hindriks, Karim El Assal, Simon Gunkel, Galit Rahim, and Omar Niamut. Remote expert assistance system for mixed-hmd clients over 5g infrastructure. In *Proceedings of the 14th Conference on ACM Multimedia Systems*, pages 318–322, 2023.
- [307] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 161–174, 2019.
- [308] Huayuan Tian, Gun A Lee, Huidong Bai, and Mark Billinghurst. Using virtual replicas to improve mixed reality remote collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 2023.
- [309] Jenifer Tidwell. *Designing interfaces: Patterns for effective interaction design*. ” O’Reilly Media, Inc.”, 2010.
- [310] Philipp Tiefenbacher, Tobias Gehrlich, and Gerhard Rigoll. Impact of annotation dimensionality under variable task complexity in remote guidance. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 189–190. IEEE, 2015.
- [311] Antonio E Uva, Michele Gattullo, Vito M Manghisi, Daniele Spagnulo, Giuseppe L Cascella, and Michele Fiorentino. Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations. *Intl J.Adv. Manufact Tech*, pages 1–13, 2017.
- [312] Jeremy Venerella, Tyler Franklin, Lakpa Sherpa, Hao Tang, and Zhigang Zhu. Integrating ar and vr for mobile remote collaboration. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 104–108. IEEE, 2019.
- [313] Jeremy Venerella, Lakpa Sherpa, Hao Tang, and Z Zhui. A lightweight mobile remote collaboration using mixed reality. In *Proceedings of Computer Vision and Pattern Recognition*, pages 1–4, 2019.

- [314] Alonso H Vera, Thomas Kvan, Robert L West, and Simon Lai. Expertise, collaboration and bandwidth. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 503–510, 1998.
- [315] Cordula Vesper, Laura Schmitz, Lou Safra, Natalie Sebanz, and Günther Knoblich. The role of shared visual information for joint action coordination. *Cognition*, 153:118–123, 2016.
- [316] Peter Vorderer, Werner Wirth, Feliz Ribeiro Gouveia, Frank Biocca, Timo Saari, Lutz Jäncke, Saskia Böcking, Holger Schramm, Andre Gysbers, and Tilo Hartmann. Mec spatial presence questionnaire. *Retrieved Sept, 18:2015*, 2004.
- [317] Wolfgang Vorraber, Johannes Gasser, Helena Webb, Dietmar Neubacher, and Philipp Url. Assessing augmented reality in production: Remote-assisted maintenance with hololens. *Procedia CIRP*, 88:139–144, 2020.
- [318] Junfeng Wang, Yaqing Feng, Cheng Zeng, and Shiqi Li. An augmented reality based system for remote collaborative maintenance instruction of complex products. In *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 309–314. IEEE, 2014.
- [319] Peng Wang, Xiaoliang Bai, Mark Billinghamurst, Shusheng Zhang, Dechuan Han, Mengmeng Sun, Zhuo Wang, Hao Lv, and Shu Han. Haptic feedback helps me? a vr-sar remote collaborative system with tangible interaction. *International Journal of Human–Computer Interaction*, 36(13):1242–1257, 2020.
- [320] Peng Wang, Xiaoliang Bai, Mark Billinghamurst, Shusheng Zhang, Weiping He, Dechuan Han, Yue Wang, Haitao Min, Weiqi Lan, and Shu Han. Using a head pointer or eye gaze: The effect of gaze on spatial ar remote collaboration for physical tasks. *Interacting with Computers*, 32(2):153–169, 2020.
- [321] Peng Wang, Xiaoliang Bai, Mark Billinghamurst, Shusheng Zhang, Sili Wei, Guangyao Xu, Weiping He, Xiangyu Zhang, and Jie Zhang. 3dgam: using 3d gesture and cad models for training on mixed reality remote collaboration. *Multimedia Tools and Applications*, pages 1–26, 2020.
- [322] Peng Wang, Xiaoliang Bai, Mark Billinghamurst, Shusheng Zhang, Sili Wei, Guangyao Xu, Weiping He, Xiangyu Zhang, and Jie Zhang. 3dgam: using 3d gesture and cad models for training on mixed reality remote collaboration. *Multimedia Tools and Applications*, 80(20):31059–31084, 2021.
- [323] Peng Wang, Xiaoliang Bai, Mark Billinghamurst, Shusheng Zhang, Xiangyu Zhang, Shuxia Wang, Weiping He, Yuxiang Yan, and Hongyu Ji. Ar/mr remote collaboration on physical tasks: A review. *Robotics and Computer-Integrated Manufacturing*, 72:102071, 2021.
- [324] Peng Wang, Yue Wang, Mark Billinghamurst, Huizhen Yang, Peng Xu, and Yanhong Li. Behere: a vr/sar remote collaboration system based on virtual replicas sharing gesture and avatar in a procedural task. *Virtual Reality*, pages 1–22, 2023.

- [325] Peng Wang, Shusheng Zhang, Xiaoliang Bai, Mark Billinghurst, Weiping He, Mengmeng Sun, Yongxing Chen, Hao Lv, and Hongyu Ji. 2.5 dhands: a gesture-based mr remote collaborative platform. *The International Journal of Advanced Manufacturing Technology*, 102(5-8):1339–1353, 2019.
- [326] Peng Wang, Shusheng Zhang, Xiaoliang Bai, Mark Billinghurst, Weiping He, Shuxia Wang, Xiaokun Zhang, Jiayang Du, and Yongxing Chen. Head pointer or eye gaze: Which helps more in mr remote collaboration? In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1219–1220. IEEE, 2019.
- [327] Peng Wang, Shusheng Zhang, Xiaoliang Bai, Mark Billinghurst, Li Zhang, Shuxia Wang, Dechuan Han, Hao Lv, and Yuxiang Yan. A gesture-and head-based multimodal interaction platform for mr remote collaboration. *The International Journal of Advanced Manufacturing Technology*, 105(7-8):3031–3043, 2019.
- [328] Shiyao Wang, Michael Parsons, Jordan Stone-McLean, Peter Rogers, Sarah Boyd, Kristopher Hoover, Oscar Meruvia-Pastor, Minglun Gong, and Andrew Smith. Augmented reality as a telemedicine platform for remote procedural training. *Sensors*, 17(10):2294, 2017.
- [329] Tzu-Yang Wang, Yuji Sato, Mai Otsuki, Hideaki Kuzuoka, and Yusuke Suzuki. Developing an ar remote collaboration system with semantic virtual labels and a 3d pointer. In *International Conference on Human-Computer Interaction*, pages 407–417. Springer, 2020.
- [330] X Wang, SK Ong, and AYC Nee. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing*, 4(1):1–22, 2016.
- [331] Yue Wang, Peng Wang, Zhiyong Luo, and Yuxiang Yan. A novel ar remote collaborative platform for sharing 2.5 d gestures and gaze. *The International Journal of Advanced Manufacturing Technology*, 119(9):6413–6421, 2022.
- [332] Zenglei Wang, Shusheng Zhang, and Xiaoliang Bai. A mixed reality platform for assembly assistance based on gaze interaction in industry. *The International Journal of Advanced Manufacturing Technology*, 116:3193–3205, 2021.
- [333] Nadir Weibel, Danilo Gasques, Janet Johnson, Thomas Sharkey, Zhuoqun Robin Xu, Xinming Zhang, Enrique Zavala, Michael Yip, and Konrad Davis. ARTEMIS: Mixed-reality environment for immersive surgical telementoring. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–4, 2020.
- [334] Robert B Welch, Theodore T Blackmon, Andrew Liu, Barbara A Mellers, and Lawrence W Stark. The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence. *Presence: Teleoperators & Virtual Environments*, 5(3):263–273, 1996.

- [335] Shigeru Wesugi and Yoshiyuki Miwa. “lazy susan” communication system for remote, spatial and physical collaborative works. In *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP06)*, pages 8–pp. IEEE, 2006.
- [336] Ross T Whitaker, Chris Crampton, David E Breen, Mihran Tuceryan, and Eric Rose. Object calibration for augmented reality. *Comput. Graph. Forum*, 14(3):15–27, 1 August 1995.
- [337] Steve Whittaker. Theories and methods in mediated communication. In *Handbook of discourse processes*, pages 246–289. Routledge, 2003.
- [338] Aiden Wickey and Leila Alem. Analysis of hand gestures in remote collaboration: some design recommendations. In *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*, pages 87–93, 2007.
- [339] Amanda Williams, Eric Kabisch, and Paul Dourish. From interaction to participation: Configuring space through embodied interaction. In *International Conference on Ubiquitous Computing*, pages 287–304. Springer, Springer, 2005.
- [340] Ederyn Williams. Experimental comparisons of face-to-face and mediated communication: A review. *Psychological bulletin*, 84(5):963, 1977.
- [341] Margaret Wilson. Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4):625–636, 2002.
- [342] Mads Wissenberg, Freddy K Lippert, Fredrik Folke, Peter Weeke, Carolina Malta Hansen, Erika Frischknecht Christensen, Henning Jans, Poul Anders Hansen, Torsten Lang-Jensen, and Jonas Bjerring Olesen. Association of national initiatives to improve cardiac arrest management with rates of bystander intervention and patient survival after out-of-hospital cardiac arrest. *JAMA*, 310(13):1377–1384, 2013.
- [343] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 143–146, 2011.
- [344] Nelson Wong and Carl Gutwin. Where are you pointing? the accuracy of deictic pointing in cves. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1029–1038, 2010.
- [345] Nelson Wong and Carl Gutwin. Support for deictic pointing in cves: still fragmented after all these years’. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 1377–1387. ACM, 2014.
- [346] Hsin-Kai Wu, Silvia Wen-Yu Lee, Hsin-Yi Chang, and Jyh-Chong Liang. Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62:41–49, 2013.

- [347] YuXiang Yan, Xiaoliang Bai, Weiping He, Shuxia Wang, XiangYu Zhang, Peng Wang, Liwei Liu, and Bing Zhang. Can you notice my attention? a novel information vision enhancement method in mr remote collaborative assembly. *The International Journal of Advanced Manufacturing Technology*, pages 1–23, 2023.
- [348] Jing Yang, Prasanth Sasikumar, Huidong Bai, Amit Barde, Gábor Sörös, and Mark Billinghurst. The effects of spatial auditory and visual cues on mixed reality remote collaboration. *Journal on Multimodal User Interfaces*, pages 1–16, 2020.
- [349] Michelle Yeh and Christopher D Wickens. Attention and trust biases in the design of augmented reality displays. Technical report, Illinois Univ. at Urbana-Champaign, Savoy-Aviation Research Lab, 2000.
- [350] George Yule. *The study of language*. Cambridge university press, 2016.
- [351] George Yule. *The study of language*. Cambridge university press, 2020.
- [352] Nadia Zenati-Henda, Abdelkader Bellarbi, Samir Benbelkacem, and Mahmoud Belhocine. Augmented reality system based on hand gestures for remote maintenance. In *2014 International Conference on Multimedia Computing and Systems (ICMCS)*, pages 5–8. IEEE, 2014.
- [353] Sunny Zhang, Brennan Jones, Sean Rintel, and Carman Neustaedter. Xrmas: Extended reality multi-agency spaces for a magical remote christmas. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*, pages 203–207, 2021.
- [354] Zhenning Zhang, Zhigeng Pan, Weiqing Li, and Zhiyong Su. Imitative collaboration: A mirror-neuron inspired mixed reality collaboration method with remote hands and local replicas. *Journal of Visual Communication and Image Representation*, 88:103600, 2022.
- [355] Jakob Zillner, Erick Mendez, and Daniel Wagner. Augmented reality remote collaboration with dense reconstruction. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 38–39. IEEE, 2018.