

# UC San Diego

## UC San Diego Electronic Theses and Dissertations

### Title

Designing for misalignment in Augmented Reality displays: Towards a Theory of Contextual Aids

### Permalink

<https://escholarship.org/uc/item/2qw9r0d0>

### Author

Gasques Rodrigues, Danilo

### Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

**Designing for misalignment in Augmented Reality displays: Towards a Theory of  
Contextual Aids**

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

in

Computer Science

by

Danilo Gasques Rodrigues

Committee in charge:

Nadir Weibel, Chair  
Scott Klemmer  
Sundar Murugappan  
Larry Smarr  
Erik Virree  
Michael Yip

2023

Copyright

Danilo Gasques Rodrigues, 2023

All rights reserved.

The dissertation of Danilo Gasques Rodrigues is approved,  
and it is acceptable in quality and form for publication on  
microfilm and electronically.

University of California San Diego

2023

## DEDICATION

To those who have fought hard and won, who have held on and succeeded, and those who are still bravely in their battle, never let your kindness and understanding be diminished; let it be the unwavering light that guides you.

## EPIGRAPH

*Nobody ever figures out what life is all about, and it doesn't matter. Explore the world. Nearly everything is really interesting if you go into it deeply enough.*

— Richard P. Feynman

## TABLE OF CONTENTS

Dissertation Approval Page . . . . .	iii
Dedication . . . . .	iv
Epigraph . . . . .	v
Table of Contents . . . . .	vi
List of Figures . . . . .	ix
Acknowledgements . . . . .	xi
Vita . . . . .	xv
Abstract of the Dissertation . . . . .	xvii
Chapter 1	
Introduction . . . . .	1
1.1 Making sense of physical task guidance . . . . .	1
1.2 Imagining the patient body . . . . .	2
1.3 What is Augmented Reality? . . . . .	3
1.4 Authoring Augmented Reality interfaces: key challenges . . . . .	5
1.5 Scope of Dissertation . . . . .	8
1.6 Thesis Statement and Contributions . . . . .	8
Chapter 2	
HoloCPR: Computer-assisted guidance in emergency scenarios . . . . .	10
2.1 Introduction . . . . .	10
2.2 Related Work . . . . .	12
2.3 Designing a situated checklist for Mixed Reality . . . . .	13
2.3.1 Role-Playing a heart attack . . . . .	14
2.3.2 Pilot Study . . . . .	14
2.3.3 Iterative Design Process . . . . .	15
2.3.4 Design Challenges and Proposed Solutions . . . . .	16
2.4 HoloCPR . . . . .	21
2.5 Evaluation . . . . .	24
2.5.1 Participants . . . . .	24
2.5.2 Experimental Design . . . . .	25
2.5.3 Data Analysis . . . . .	26
2.6 Results . . . . .	27
2.6.1 Response Time . . . . .	27
2.6.2 Performance and Transition Time . . . . .	28
2.6.3 Procedural Adherence . . . . .	28
2.6.4 Subjective Feedback . . . . .	30

	2.7	Discussion . . . . .	31
	2.8	Closing Remarks . . . . .	32
Chapter 3		ARTEMIS: Collaborative guidance in a trauma telementoring . . . . .	34
	3.1	Introduction . . . . .	34
	3.2	Designing a Mixed Reality Telementoring System . . . . .	36
	3.2.1	Role-playing in a mock-up operating room . . . . .	37
	3.2.2	How do mentors guide mentees? . . . . .	38
	3.2.3	How much information does the mentor need from the mentee side? . . . . .	39
	3.2.4	What do mentors want to show to mentees? . . . . .	41
	3.2.5	Summary . . . . .	45
	3.3	ARTEMIS overview . . . . .	45
	3.3.1	Novice Surgeon’s Interface . . . . .	46
	3.3.2	Expert Surgeon’s Interface . . . . .	49
	3.3.3	Calibrating 3D annotations . . . . .	52
	3.4	Implementation . . . . .	53
	3.4.1	Overcoming HoloLens Limitations . . . . .	54
	3.4.2	Rapid Technology integration . . . . .	55
	3.4.3	Rendering the Point-cloud . . . . .	56
	3.4.4	Calibrating Different Coordinate Systems . . . . .	56
	3.5	System Evaluation . . . . .	58
	3.5.1	Procedure . . . . .	58
	3.5.2	Results and Discussion . . . . .	60
	3.6	Conclusion . . . . .	63
	3.7	Closing Remarks . . . . .	64
Chapter 4		Measuring the limits of perception-based guidance . . . . .	66
	4.1	Introduction . . . . .	67
	4.2	Background and Related Work . . . . .	69
	4.2.1	Simulating Augmented Reality in Virtual Reality . . . . .	71
	4.3	The Virtual-Augmented Reality simulator . . . . .	72
	4.3.1	Simulator’s Features . . . . .	74
	4.4	Materials and methods . . . . .	75
	4.4.1	Study Design . . . . .	76
	4.4.2	Alignment interfaces . . . . .	77
	4.4.3	Metrics and Participant Feedback . . . . .	78
	4.4.4	Task . . . . .	79
	4.4.5	Participants . . . . .	80
	4.4.6	Experimental Procedure . . . . .	80
	4.4.7	Implementation . . . . .	81
	4.5	Results . . . . .	82
	4.5.1	Quantitative Metrics . . . . .	82

4.5.2	Alignment error . . . . .	83
4.5.3	Perceptual differences . . . . .	85
4.6	Discussion . . . . .	88
4.6.1	Eye-display calibration errors highly impact manual alignment accuracy . . . . .	88
4.6.2	Eye-display calibration errors nullify alignment interface improvements . . . . .	89
4.6.3	Eye-display calibration errors are eventually perceived as virtual object drift . . . . .	90
4.6.4	Eye-display calibration errors increase user effort, frustration, and sense of failure . . . . .	91
4.6.5	Similarities and discrepancies between VR and AR modalities . . . . .	91
4.6.6	Limitations . . . . .	93
4.7	Closing Remarks . . . . .	93
Chapter 5	Towards Contextual Cues . . . . .	95
5.1	The constraints of perception-based Guidance . . . . .	95
5.2	Easing the cognitive burden of imagining the patient's body . . . . .	97
5.3	The case of situated ultrasound . . . . .	98
5.4	A taxonomy of Augmented Reality cues . . . . .	99
5.5	Contextual cues: integrating different coordinate systems . . . . .	101
5.6	Contextual cues: future work . . . . .	102
Bibliography	. . . . .	104

## LIST OF FIGURES

Figure 1.1:	Possible implementations of augmented reality displays. (Adapted from Oliver Bimber and Ramesh Raskar [18]) . . . . .	4
Figure 1.2:	Three examples of mixed-fidelity prototypes created with PintAR [38] . . . . .	7
Figure 2.1:	The first interactive prototype with two steps as seen by the HoloLens. . . . .	14
Figure 2.2:	HoloLens 1 with a clicker tapped to the side. Pressing the clicker was faster and more accurate than hand gestures. . . . .	16
Figure 2.3:	The <i>Red Arrow</i> guides the users, indicating to them where the important things are located and where they should act . . . . .	17
Figure 2.4:	Instructions for rescue breaths, as seen by the wearer. Visible are also the Circle of Attention and the Vertical Pole in the middle of it. . . . .	17
Figure 2.5:	The “ <i>Circle-of-attention</i> ” points the user to where the important information within HoloLens’ limited field-of-view . . . . .	20
Figure 2.6:	HoloCPR in Action. (a-b) AR interface shows instructions in space. (c-d) Chest compressions are seeing by the wearer. . . . .	21
Figure 2.7:	Dividing instructions into <i>What, How, and Where</i> . . . . .	23
Figure 2.8:	Response time, transition time, time spent searching for the AED, and time taken to attach AED pads across both conditions. . . . .	29
Figure 2.9:	HoloCPR: situated instructions showing the next step of the CPR checklist. Each instruction details <i>what, where, and how</i> they should be executed. . . . .	33
Figure 3.1:	Artistic rendering of ARTEMIS and its features. Left: the AR interface for novices. Right: the VR interface for experts. . . . .	35
Figure 3.2:	Our mock-up operating room and office environments. In the left picture, we see a video frame from one of the role-playing sessions. In the right picture, we see the expert’s office space with a number of paper and digital tools that we used to explore how experts would interact with the novice surgeons. . . . .	38
Figure 3.3:	ARTEMIS VR - Expert surgeon’s interface in Virtual Reality. . . . .	44
Figure 3.4:	ARTEMIS AR Interface - Novice surgeon’s interface in Augmented Reality. (a) The Expert’s avatar shows them holding a virtual pencil. (b) 3D Annotations as seen by the novice. (c) Procedure clips showing a video clip. . . . .	47
Figure 3.5:	ARTEMIS in action. Left: ARTEMIS AR, with a novice wearing HoloLens. Right: ARTEMIS VR, with an expert surgeon guiding the novice user, wearing the IMU-equipped gloves. . . . .	49
Figure 3.6:	ARTEMIS VR Input Interfaces. Expert surgeon’s wear gloves and interact with the interface through a physical pen. . . . .	51
Figure 3.7:	The expert surgeon annotates on the reconstruction of the patient body in VR (left) while the novice surgeon sees it overlaid onto the real patient in AR (right) . . . . .	52

Figure 3.8:	ARTEMIS is composed of two separate environments: ARTEMIS AR (left) and ARTEMIS VR (right). This diagram shows the origin and direction of data flowing from each environment. . . . .	53
Figure 3.9:	Diagram representing all tracked objects in the Augmented Reality space. . . . .	57
Figure 4.1:	Depiction of two common Augmented Reality setups . . . . .	70
Figure 4.2:	The virtual-augmented reality simulator: how it works. . . . .	72
Figure 4.3:	Study design: a 2x3 factorial study design with two independent variables: <b>display modality</b> and <b>alignment interface</b> . The display modality variable has three levels: <i>Augmented Reality (AR)</i> , <i>Virtual Reality (calibrated)</i> , or <i>Virtual Reality (1mm error)</i> . . . . .	76
Figure 4.4:	User study in AR using OptiTrack Motion Capture cameras for tracking the Target Object . . . . .	82
Figure 4.5:	Boxplot of alignment error, or the Euclidean distance between the virtual and physical objects in centimeters. . . . .	84
Figure 4.6:	Barplots of the mean alignment time per <i>Display modality</i> . Left: POINT. Right: HOLOGRAM . . . . .	85
Figure 4.7:	The five relevant dimensions from NASA TLX for this study. Frustration, Performance, Physical Demand, Effort, and Mental Demands. Temporal demand didn't present any statistically significant differences across variables. . . . .	86
Figure 4.8:	Boxplot of alignment error focusing on VR (1MM ERROR) and VR (CALIBRATED). . . . .	89
Figure 4.9:	An example of what eye-display calibration looked like for users manually aligning a virtual target (Hologram) to a physical target they could see while walking around the two. . . . .	90
Figure 5.1:	From cognitive load to offload: the journey of media technology from interpreted to perception-based guidance. . . . .	96
Figure 5.2:	Student (S) performing a central line placement in a phantom. Notice that the student alternates between the site of the procedure (A) and the ultrasound screen (B). S1 and S2 are instructors assessing the student's procedural skills . . . . .	99
Figure 5.3:	Situated Ultrasound Prototype (A) View from the headset. (B) Overview of the task - guiding a needle into the patient's neck vein. (C) Student looking at the situated ultrasound . . . . .	100
Figure 5.4:	Three different types of cues or visual aids. Depth, Interaction, and the proposed Contextual aids. . . . .	100
Figure 5.5:	An example of contextual cue in the situated ultrasound project. . . . .	102

## ACKNOWLEDGEMENTS

I cannot emphasize enough my gratitude towards my chair and adviser, Nadir Weibel. Nadir showed me the role of ubiquitous computing research in making the world a better place, which ignited my pursuit of a Ph.D. I owe him deep gratitude for years of guidance, mentorship, encouragement, draft reviews, late-night writing sessions, prototyping, outreach projects, and so much more. Nadir's compassionate approach to research also taught me how to view negative results more positively and turn them into opportunities.

I am also profoundly grateful to my Ph.D. committee. Scott Klemmer, for his vast arsenal of heuristics for academic research and his wizard-like ability to quickly understand and reframe any research into its most meaningful and impactful perspective. Sundar Murugappan, for his mentorship during my internship at Intuitive and throughout my last years of research. Michael Yip, for our enlightening weekly meetings throughout our collaboration on ARTEMIS. Larry Smarr, who led pragmatically the application of image guidance in intricate colon surgeries. And Erik Virree for his guidance and assistance in recent years.

To all my collaborators, thank you for your unwavering dedication and enthusiasm in conducting Augmented and Mixed Reality Research. Crafting and developing such systems for science is a relentless task - every component of software and hardware you engage with will fail in the most unforeseen and unpredictable ways. Nothing can be assumed, and when you finish with one system, you've merely accomplished the first phase of human-centered research. This might make our field of study seem like a "mixed reality," but you made the journey more fun and exciting. A special thanks to Tommy Sharkey, Weichen Liu, Jon Paden, Steven Rick, Vish Ramesh, Janet Johnson, Alla Vovk, Preetham Suresh, and Shiv Patel.

My gratitude extends to everyone at the Design Lab for countless rewarding research meetings where everyone contributed a small piece and departed with a wealth of knowledge. To Don, for his constant reminders to write more. To Bill Griswold for teaching us how to read an engineering research paper. To Jim Hollan, who always asked us, "What is your research

question?” To Scott, who never forgot to question us, “What is your Y axis?” To Eric Heckler and Elizabeth Keys, for showing me the way to compassionate critique. To Steven Dow, for the numerous hours we spent discussing AR prototyping toolkits. To Deborah Forster, for teaching me ethnographic video coding. And to all members, in no specific order, for so many insightful chats and fun moments and for simply being the best lab mates: Vineet, Ailie, Dorothy, Steven, Shristi, Mayya, Amy Fox, Shaun, Colleen Emmenegger, Tricia, and Melanie.

To all my CSE friends who shared the daily grind at Chez Bob, the Social Hours, and the marathon nights of deadlines: Dhiman, Shravan, Mario, Yan Yan, and many others I may be forgetting right now.

My internship mentor and manager, Govind Payyavula, deserves special mention. Govind is the best manager I had, and his passion for Augmented Reality in surgical robotics was inspiring and motivating. The two summers at Intuitive Surgical count among my most cherished academic experiences. Thank you, Govind, for countless hours of mentorship. To Intuitive’s applied research team for the opportunity to pursue novel research problems together: Simon DiMaio, Govind Payyavula, Sundar Murugappan, Cortney Jansen, and Dale Bergman.

I’m grateful to our collaborators at The Naval Medical Center: Captain Konrad Davis, Roland Champagne, Matthew Tadlock, and the wetlab staff. Those were special days with really long hours to test cutting edge research.

A salute to the serial innovator and collaborator Preetham Suresh, whose passion for tackling complex problems in Augmented Reality for medical procedures inspired me. Preetham, your commitment to the broader field of simulation training is both humbling and inspiring. I’m grateful for the research and development space you created in the Simulation Training Center, allowing us work at the intersection of surgical training and computer science research. Also, thanks to my collaborators there, Julian Mayer, Shiv Patel, and Brennan Marsh-armstrong.

On a personal note, I thank my grandpa Sebastiao, who played a pivotal role in our family and taught me discipline and work ethics from a very early age. May his soul rest in peace.

To my parents for their unwavering support, and my siblings, Elaine and Guilherme, who generously allowed me more time on the computer when we were all kids and just wanted to play games.

To the “bikers group” - Ariana and Steven, thank you for helping me maintain a work-life balance and supporting me during the most challenging academic times. Of all the friends I made during the PhD journey, you two belong to a special place in my heart. Honestly, I don’t even know what to write here because I think words are a very shallow medium to describe the importance you two had in my life. I hope that we continue biking together even as we grow older, get jobs at different locations, and sign up for varied adult responsibilities.

Thank you to my favorite artists and friends, Emily Grenader and Jon Paden. Emily, thank you for always thinking of me when working on remarkable sociotechnical art installations. Those were the most fun moments I had outside the Ph.D. research! Jon, thank you for showing the more conceptual and provocative side of Augmented Reality.

To my Brazilian Ph.D. friends, Emilia, Adriane, Leticia, and Adrielly.

Lastly, to my wife Su, my enduring companion and steadfast supporter of my mental well-being. I am grateful for her patience in dealing with a partner who is always absorbed in some form of writing or research, even during vacations.

Chapter 2 is, in part, a reprint of the material as it appears in the Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. Janet G. Johnson, Danilo Gasques Rodrigues, Madhuri Gubbala, and Nadir Weibel, from the Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. The dissertation author was a collaborator and contributor to this paper.

Chapter 3, in full, is a reprint the material as it appears in the the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. I was the lead author of this material. Danilo Gasques, Janet G. Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Konrad Davis, Michael Yip, and Nadir Weibel. The dissertation author was

the primary investigator and author of this material.

Chapter 4 is, in part, a reprint of the material as it appears in 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) and, in part, being prepared for publication. Danilo Gasques, Weichen Liu, and Nadir Weibel. The dissertation author was the primary investigator and author of both materials.

## VITA

2014	Bachelor of Science in Engineering, Federal University of Sao Carlos, Brazil
2016–2019	Master of Science, University of California San Diego
2019–2023	Doctor of Philosophy, University of California San Diego

## PUBLICATIONS

“Mixed reality surgical mentoring of combat casualty care related procedures in a perfused cadaver model: Initial results of a randomized feasibility study.” Matthew D. Tadlock, Erik J. Olson, Danilo Gasques, Roland Champagne, Michael J. Krzyzaniak, Shawn A. Belverud, Vijay Ravindra et al. *Surgery* 172, no. 5 (2022): 1337-1345.

“The Virtual-Augmented Reality Simulator: Evaluating OST-HMD AR calibration algorithms in VR.” Danilo Gasques, Weichen Liu, and Nadir Weibel. *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops*, pp. 702-703. IEEE, 2022.

“Facilitating remote design thinking workshops in healthcare: the case of contouring in radiation oncology.” Matin Yarmand, Chen Chen, Danilo Gasques, James D. Murphy, and Nadir Weibel. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1-5. 2021.

“ARTEMIS: A collaborative mixed-reality system for immersive surgical telementoring.” Danilo Gasques, Janet G. Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Xinming Zhang, Konrad Davis, Michael Yip, and Nadir Weibel. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1-14. 2021.

“Do you really need to know where ‘that’ is? enhancing support for referencing in collaborative mixed reality environments.” Janet G. Johnson, Danilo Gasques, Tommy Sharkey, Evan Schmitz, and Nadir Weibel. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1-14. 2021.

“Artemis: Mixed-reality environment for immersive surgical telementoring.” Nadir Weibel, Danilo Gasques, Janet Johnson, Thomas Sharkey, Zhuoqun Robin Xu, Xinming Zhang, Enrique Zavala, Michael Yip, and Konrad Davis. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1-4. 2020.

“Pintar: Sketching spatial experiences in augmented reality.” Danilo Gasques, Janet G. Johnson, Tommy Sharkey, and Nadir Weibel. *In Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*, pp. 17-20. 2019.

“What you sketch is what you get: Quick and easy augmented reality prototyping with pintar.” Danilo Gasques, Janet G. Johnson, Tommy Sharkey, and Nadir Weibel. *In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-6. 2019.

“Augment, Instruct, Execute, and Verify: Augmented Reality Wound Care Training with 3D Printed Props” Danilo Gasques, Weichen Liu, Kevin Broder, Diane Chau and Nadir Weibel. *Workshop on Interactive Systems for Healthcare*, CHI '19, Glasgow, UK.

“I Want to Be a Surgeon! Role Playing for Remote Surgery in Mixed Reality” Thomas Sharkey, Janet G. Johnson, Danilo Gasques, and Nadir Weibel. *Workshop on Interactive Systems for Healthcare*, CHI '19, Glasgow, UK.

“SIG: spatiality of augmented reality user interfaces.” Alla Vovk, Danilo Gasques Rodrigues, Fridolin Wild, and Nadir Weibel. *In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-4. 2019.

“Real-time guidance for cardiopulmonary resuscitation in mixed reality.” Danilo Gasques Rodrigues, Janet Johnson, and Nadir Weibel. *In 12th EAI International Conference on Pervasive Computing Technologies for Healthcare–Demos*, Posters, Doctoral Colloquium. 2018.

“Holocpr: Designing and evaluating a mixed reality interface for time-critical emergencies.” Janet G. Johnson, Danilo Gasques Rodrigues, Madhuri Gubbala, and Nadir Weibel. *In Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pp. 67-76. 2018.

“Exploring mixed reality in specialized surgical environments.” Danilo Gasques Rodrigues, Ankur Jain, Steven R. Rick, Liu Shangle, Preetham Suresh, and Nadir Weibel. *In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2591-2598. 2017.

ABSTRACT OF THE DISSERTATION

**Designing for misalignment in Augmented Reality displays: Towards a Theory of Contextual Aids**

by

Danilo Gasques Rodrigues

Doctor of Philosophy in Computer Science

University of California San Diego, 2023

Nadir Weibel, Chair

Augmented Reality (AR) displays show virtual information spatially in the physical world. For example, in the surgical domain, AR displays can overlay medical imaging on top of the patient to help surgeons distinguish healthy tissue from cancerous tissue. This affordance of adding visual information near the area of action has led to a gamut of AR applications with one caveat: their design assumes a flawless alignment between the virtual information and its physical context. If this alignment fails, the system might mislead users by presenting the virtual information out of context.

Fortunately, technical advancements in tracking, calibration, and optics have helped

improve and create a more seamless augmented reality experience; however, even in these improved systems, alignment failure will lead to user-task performance errors. In my dissertation, I argue that AR interface design needs to account for such misalignments and limitations of AR technology. First, I investigated the case of a situated checklist for cardiopulmonary resuscitation (HoloCPR), which helped deliver accurate guidance even in the presence of tracking errors. Then, I investigated a telementoring system for trauma procedures (ARTEMIS) which helped remote surgeons communicate with novices despite a constant registration error. Finally, I showed that slight optical deviation can drastically impact overall user-task performance and cognitive workload, even in an idealized environment with perfect tracking and calibration.

I sum the findings of this final piece, along with the systems designed and developed in this dissertation, to point in the direction of a new type of AR interface building block: contextual aids and how they may have significant implications for AR interface design for surgical procedures, paving the way for the development and integration of more advanced AR systems in surgical practice.

# Chapter 1

## Introduction

### 1.1 Making sense of physical task guidance

People often grapple with understanding written or spoken instructions for manual tasks, particularly when those instructions delineate the precise positioning and orientation of tools and instruments. Reading, contextualizing, and executing these instructions requires considerable mental effort. This is mainly because traditional instruction methods struggle to adequately articulate complex spatial relationships or dynamic processes, leaving room for misinterpretation and error.

The advent of multimedia instructions, encompassing illustrations, videos, and interactive content, has somewhat alleviated this difficulty. Visual aids such as diagrams and video tutorials can effectively demonstrate concepts that are otherwise challenging to articulate in words. However, even these methods are not without their shortcomings. Often, the instructions are presented in a manner that deviates in scale, orientation, or location from the actual application area. This discrepancy can still lead to confusion and requires some mental effort on the user's part to transpose the instructions into their specific context.

Augmented Reality (AR) technology can reduce the cognitive effort required to make

sense of instructions. As an interactive, real-time technology, AR superimposes virtual objects onto the physical world, creating an illusion that they coexist [11, 10]. This seamless integration allows for the presentation of instructions directly within the user's environment, thereby reducing the cognitive load involved in abstracting and applying the instructions to a different spatial context. I.e., AR enables users to comprehend instructions in their exact physical context without mentally transposing them from a different perspective. Users can probe and investigate the instructions from multiple angles by moving around the virtual elements. This interactivity, coupled with the spatial context, reduces ambiguities and makes AR a powerful tool for guiding physical tasks.

In this ideal vision of Augmented Reality, this direct perception of instructions has the potential to expedite comprehension and facilitate quicker, more accurate execution. A domain that can drastically benefit from this ideal AR display is the medical domain, wherein text descriptions and illustrations describe not the instructions but rather show patient internals, surgical planning, and live medical imaging [57].

## **1.2 Imagining the patient body**

“Surgery is fundamentally a (...) practice of exploring and modifying a physical object —the body.” Mentis et al. [72].

In surgery, imaging systems and displays serve as invaluable tools. They equip medical practitioners with internal visuals of the patient's body, a blueprint of the surgical plan, and the locations of essential instruments within the patient's body. However, the actual image of the patient's body is constructed mentally [72], relying on a blend of physical probing, preoperative scans, real-time imaging, and collaborative conversations within the medical team.

Visualizing the patient's body is cognitively strenuous and, as such, is susceptible to errors and oversights [5, 98]. Often, patient visualizations are portrayed from a perspective, scale, and orientation that differ from the surgeon's. Surgeons are then compelled to mentally

transform these images - through interpolation, rotation, translation, or scaling - to align with their viewpoint [2]. Moreover, in some cases, when surgeons need access to new views of pre-op imaging during a procedure, they might have to take their surgical gloves off, walk to the closest computer, assess the 3D scans, and then prepare to go back to the sterile field [86]. With such a choppy way of accessing imaging that still causes an intensive cognitive burden, it is unsurprising that medical errors currently rank as the third leading cause of death in the United States [67].

Under the domain of guidance for physical tasks, surgery, and medical procedures can significantly benefit from Augmented Reality (AR) [11, 40]. With AR, medical imaging, surgical plans, and real-time instrument positioning can be visualized from the surgeon's perspective within the context of the patient's body. AR provides an opportunity to transition from a constructed view of the patient's body to an actual preview of their internals. Then, perusing the patient's internals is no longer a cognitively taxing imaginative exercise but rather a straightforward action.

### **1.3 What is Augmented Reality?**

In the examples above, Augmented Reality is hypothesized as a champion for manual-task guidance. While one might be able to infer specific affordances of this technology through the examples given, a focused definition is always welcome. To me, one of the most objective definitions of Augmented Reality is given by Ronald T. Azuma in his seminal work "A Survey of Augmented Reality" [11]:

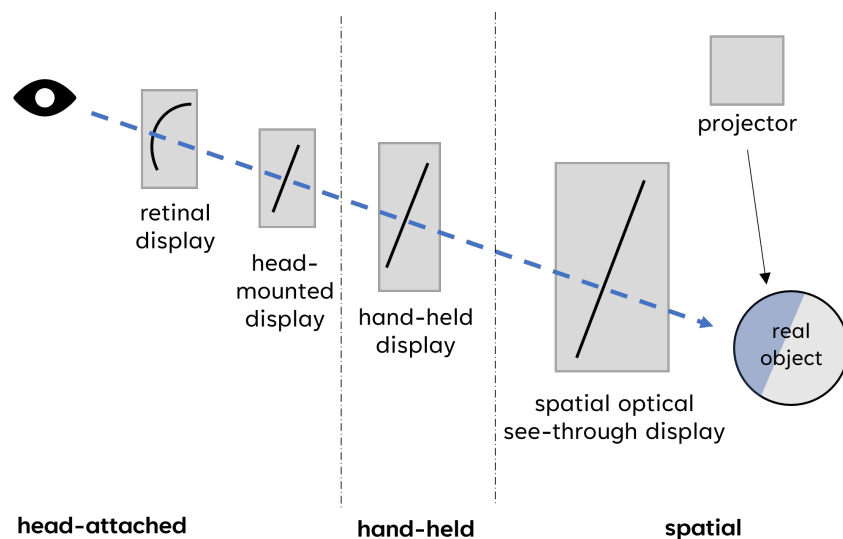
Augmented Reality are systems that have the following three characteristics:

- Combines real and virtual
- Interactive in real time
- Registered in 3D

These three characteristics are extremely important affordances of AR for the domain of manual task guidance. First, integrating virtual instructions and information with their tangible,

real-world context offers the opportunity. Second, combining both happens dynamically, adjusting what is seen as the user interacts with the physical world. Finally, the virtual content is anchored to a specific physical location, and the display adjusts its presentation to the user’s viewpoint.

According to this definition of AR, different systems might fulfill the needs of an Augmented Reality system while affording different levels of interaction, immersion, and realism. Figure 1.1 illustrates different AR display implementations, adapted from Oliver Bimber and Ramesh Raskar’s seminal book [18].



**Figure 1.1:** Possible implementations of augmented reality displays. (Adapted from Oliver Bimber and Ramesh Raskar [18])

During my Ph.D., the most commonly available AR systems were head-attached, hand-held, and projective displays. While I experimented with all these three possible designs, head-attached displays provide the most ergonomic and convenient design for manual task guidance. These devices allow users to interact with virtual and physical content hands-free and adjust to the user’s viewpoint automatically as the user moves around. Between retinal displays and head-mounted displays, retinal displays present perceptual [52] advantages over head-mounted displays; however, retinal displays are still in early development stages [18]. Off-the-shelf head-mounted displays such as Microsoft HoloLens are more widely available and enabled me to

conduct research without focusing too much on developing the enabling technology <sup>1</sup>.

## 1.4 Authoring Augmented Reality interfaces: key challenges

Augmented Reality (AR) systems traditionally required significant engineering effort. Even when the primary objective might center around studying AR's effects in a specific context, significant time was often diverted to establishing the basic building blocks of an AR system. These components include a suitable display, a tracking mechanism for the display, a rendering pipeline for virtual content, and an implementation of marker or object tracking to track other things of interest in the user's environment [11, 56, 18].

The advent of commercial AR displays, starting with handheld devices [27] and later evolving to head-mounted versions [19], allowed researchers to pivot their attention towards unique AR applications [19, 27].

Modern commodity Head-Mounted Displays (HMDs) such as the Microsoft HoloLens [75, 76] encapsulate years of research into a comprehensive standalone device [62]. These devices empower users to integrate virtual objects into unprepared environments without additional equipment, such as fiducials or motion capture technology [62, 50]. They also offer integrated head-eye calibration [77], enhancing the realism of virtual object perception. Thus, with tools like HoloLens, researchers and general users began to perceive AR less as a technology and more as a medium.

While the technology has undoubtedly matured beyond the early dreams of researchers, modern investigators, including myself, still grapple with challenges when maximizing AR's potential. These challenges span technical and user experience domains.

On the technical front, while off-the-shelf systems like HoloLens offer a significant step forward in ease of use, they still present constraints, especially when pushing the boundaries of

---

<sup>1</sup>For an in-depth description of the HMDs, I recommend reading "*Comparison of optical and video see-through, head-mounted displays*" by Jannick P Rolland, Richard L Holloway and Henry Fuchs [99].

what is achievable. Moreover, customization for niche applications such as medical applications is not always straightforward. Finally, certain limitations in tracking accuracy [65, 113] and field of view can pose significant challenges to users following instructions in AR.

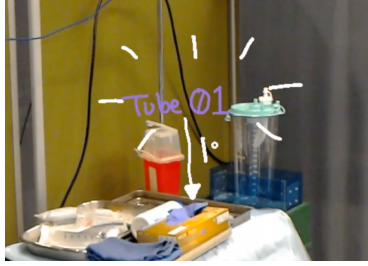
User experience presents the most complex challenge. Creating intuitive interfaces and optimizing user interactions for AR is an ongoing research area. Striking the right balance between virtual and real content so users can seamlessly switch between these two layers without feeling overwhelmed is tricky. The risk of information overload is real, potentially disorienting users, especially when they aim to follow instructions. Unlike conventional interfaces, AR inhabits a 3D space, making interactions multidimensional. Gestures, gaze, voice, and physical movements become plausible interaction modes. Deciding the optimal mode for specific tasks demands rigorous testing and iteration.

Nevertheless, during my journey, I collaborated with dedicated researchers who assisted in co-designing with our target users, establishing mock-up scenarios, and refining our proposed AR systems to address technical constraints, meet user needs, and advance the field.

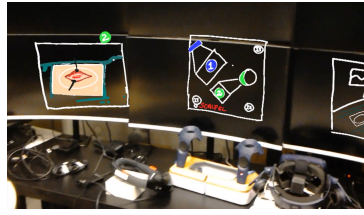
While not the primary focus of this dissertation, I developed tools that streamlined the prototyping and testing of AR applications in diverse environments. For more details on these contributions, please refer to the following publications:

- “What you sketch is what you get: Quick and easy augmented reality prototyping with pintar.” Danilo Gasques, Janet G. Johnson, Tommy Sharkey, Nadir Weibel. [38]
- “The Virtual-Augmented Reality Simulator: Evaluating OST-HMD AR calibration algorithms in VR.” Danilo Gasques, Weichen Liu, and Nadir Weibel. [39]

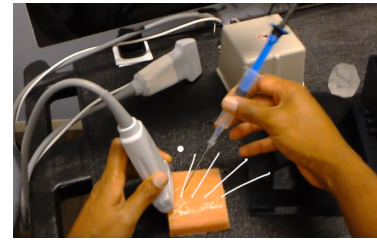
Works such “what you sketch is what you get” [38], wherein users bring digital sketches to the AR world, were vital in helping expedite the mixed-fidelity rapid-prototyping needed for ARTEMIS and other projects. The tool was also widely used in Ubiquitous Computing classes for students prototyping various cross-device systems.



(a) highlighting tool during enacted surgical telementoring



(b) overlaying application-specific content over real monitors



(c) guiding a user during a central line placement

**Figure 1.2:** Three examples of mixed-fidelity prototypes created with PintAR [38]

Along with designer-centric contributions, I also manually wrote a variety of building blocks for cross-device augmented-reality systems. These building blocks were a critical enabling technology in most AR works implemented and studied in our lab. These are available for public non-commercial use and include:

- A fast jpeg decoder for Unity (<https://github.com/WeibelLab/JPEGTurbo-Unity>). Used in [37, 54].
- A lightweight and prototyping-focused networking library co-developed with Tommy Sharkey (<https://github.com/weibellab/Comms>) and used by [37, 54, 39, 103]
- A low-latency, multi-camera video streaming and recording backend<sup>2</sup>. Used in [37, 54, 103].
- A web-based AR marker generator with support for different AR tracking systems, responsible for powering ARTEMIS [37] multi-system calibration (<https://github.com/danilogr/AR-Marker-Generator>).

---

<sup>2</sup><https://github.com/danilogr/CameraStreamer>

## 1.5 Scope of Dissertation

The primary focus of my Ph.D. work has been on Augmented Reality (AR) interfaces for guidance. This encompasses spatial checklists (HoloCPR [53]), telementoring (ARTEMIS [37], and other studies [82]), and procedural medicine (UltraTracker [103]). While my initial ambition was to delve into the technical and design building blocks that underpin these systems, the vastness of such an endeavor requires a narrower approach. Consequently, this dissertation concentrates on deriving key design insights from two select systems and investigating a fundamental source of *registration* errors in modern head-mounted displays: the calibration of the eye-display interface. From my perspective, the three most significant insights from my research are:

- A minor calibration error in an optical see-through head-mounted display—of even 1 mm significantly impacts one’s ability to follow virtual instructions accurately.
- Despite the limitations of *registration*, AR interfaces can leverage visual guidance to enhance user-task performance through *design*.
- Through AR interfaces that use contextual aids to overcome *registration* problems, users can delegate most mental computations to the AR system, allowing them to concentrate on tasks that the system is yet to solve.

## 1.6 Thesis Statement and Contributions

AR interfaces, through the implementation of contextual aids, enhance user-task performance in collaborative and computer-assisted guidance scenarios, even when perfect alignment is not possible.

### Contributions

- The development of AR interfaces that guide users toward situated instructions

- Formative research on the current approaches of surgeons to telementoring, along with unaddressed communication needs.
- Design goals for addressing the present gaps in surgical telementoring and procedural training.
- Empirical validation that shows that even minor eye-display calibration errors statistically impact alignment performance.
- The highlighting of significant differences between AR and simulated AR, thereby bridging the gap between theoretical predictions and real-world applications.
- The creation of software artifacts and system designs to facilitate the rapid prototyping of AR systems for collaborative and computer-assisted guidance.

# Chapter 2

## HoloCPR: Computer-assisted guidance in emergency scenarios

### 2.1 Introduction

During an out-of-hospital cardiac arrest, initiating cardiopulmonary resuscitation (C.P.R.) immediately after the victim collapses is paramount for their survival. Medical personnel often take 8 minutes or more to arrive on the scene. Given that the survival rate decreases by 7-10% for every minute without resuscitation, immediate C.P.R. from a bystander can significantly impact survival outcomes.

Bystanders happen to be around the victim during a sudden cardiac arrest. They are usually not trained medical professionals and must rely on past training or existing knowledge to help the victim. Although it has been shown that bystander-initiated C.P.R. can double or triple survival rates [84], only 30-40% of bystanders administer C.P.R. [85]. Fear of causing further harm or panic can hinder their willingness to act, even with previous training [44].

To encourage bystander C.P.R., first-aid kits often include manuals that provide step-by-step instructions. Additionally, mobile phones can be used to access similar guidance. However,

these approaches demand the bystander’s ability to read and interpret instructions while aiding the victim, potentially increasing cognitive load amid a stressful situation. It remains unclear whether the presence of a checklist improves C.P.R. rates. We borrow an example from hospitals to argue that easier-to-use instructions are needed.

Highly inspired by checklists in aviation [42], hospitals use checklists in scenarios where the cognitive load is significantly increased due to time-sensitive tasks. However, in medical settings, and unlike the checklists available for aircraft crew, medical checklists are not integrated with the operating equipment. Typically found on plastic sheets or separate mobile apps, their use can add time to the procedure, prompting noncompliance due to concurrent cognitive demands.

Given these challenges, we argue for using Mixed Reality checklists for guidance. We expect faster response times and fewer errors by presenting real-time, context-sensitive instructions directly to the user. Here, we use C.P.R. as an example but see the potential for more complex cases in hospitals. In the future, surgeons could access checklists via their goggles, instantly accessing information needed for their tasks. These MR checklists blend virtual data with real-world objects, positioning instructions near their point of application.

In this study, we addressed the following research question: *How does situated guidance affect users’ capability to complete tasks efficiently?* We answer this research question by conducting a between-subjects user study with users using the state-of-the-art checklist from the RedCross with a Mixed Reality checklist. To design the Mixed Reality checklist, we followed an iterative prototyping approach to understand better how people make sense of situated virtual information in a time-constrained environment.

In conclusion, we demonstrate that an interface harnessing the unique capabilities of an MR device significantly reduces unintended errors and expedites task completion, thereby enhancing procedural adherence. Our results further reveal that overlaying information on real-world objects provides context (e.g., overlaying A.E.D. pads exactly where they need to be placed) and enable users to interact with the checklist more efficiently. Information once confined to

a screen is now spatially distributed and contextually juxtaposed with physical objects. This advancement eliminates the need for users to divert their attention to read instructions or grapple with spatial comprehension, allowing them to focus on acting promptly instead.

By designing and evaluating a Mixed Reality Interface for checklists and comparing it to a baseline, this paper brings the following contributions:

- empirical results of whether the new interaction affordances of M.R. improve task completion time and quality, especially in time-sensitive scenarios
- design guidelines that can inform more complex medical crisis checklists
- observations that can inform the creation of a framework for M.R. interfaces.

## **2.2 Related Work**

In the ever-evolving landscape of human-computer interaction, Mixed Reality interfaces can potentially improve how we guide users through manual tasks. One of the early examples in this field was Knowledge-based Augmented Reality for Maintenance Assistance (KARMA) by Steven Feiner, Blair MacIntyre, and Dorée Seligmann [31], which laid the groundwork by presenting an augmented reality system for automatically spatializing instructions for printer maintenance. KARMA explored how augmented reality and some artificial intelligence could automatically generate and display step-by-step assistance for executing complex 3D tasks. Around the same time, researchers at Boeing envisioned using AR to expedite the efforts of transforming a digital airplane design into instructions available to workers at the manufacturing facilities [24]. Finally, a few years later, Whitaker and colleagues [116] further pushed the boundaries of Mixed Reality interfaces by developing a method to track and register information on real-world objects. Their ingenious approach allowed virtual annotations to be placed accurately on different components of complex systems. Effectively, they created an authoring interface for such situated

instructions.

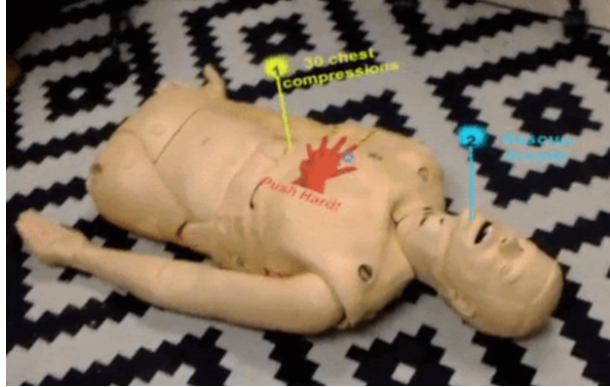
While these pioneering works bear resemblances to our own application, it is vital to recognize that they belong to the broader realm of Augmented Reality research, which has historically lacked formal user studies. To date, a staggering 8% of AR publications have proper user-based experimentation [105]. Hence, there remains a gap in understanding how well users can integrate and engage with these Mixed Reality systems.

This work focuses on two novel aspects of employing situated instructions in Mixed and Augmented Reality. First, we describe our iterative design process and decisions that ultimately led to a usable medical checklist in a time-constrained environment. Then, we conduct an experiment to measure how our situated checklist decreases the time it takes users to execute it and how it impacts their adherence to detailed instructions. We hope that user-centric design, our approach to rapid prototyping in the iterative design, and the study results will serve as the foundation for checklists in Mixed Reality.

## **2.3 Designing a situated checklist for Mixed Reality**

When using digital checklists on mobile devices, users typically have some familiarity and understanding of navigation paradigms. Content is conventionally displayed and read from top to bottom. Navigating to the next step usually takes a single tap on a digital button.

In Mixed Reality interfaces, the potential for interactions multiplies - encompassing movement, voice, gaze, and more- all requiring meticulous design consideration. Critically, in scenarios where time is of the essence, users don't have the luxury of learning the interface. Therefore, it's imperative to tailor the interface to resonate with users' most instinctive interactions, ensuring the interface intuitively responds to their immediate, natural commands.



**Figure 2.1:** The first interactive prototype with two steps as seen by the HoloLens.

### 2.3.1 Role-Playing a heart attack

To better understand how users would behave in the stressful scenario of doing CPR on a person, we prepared an environment with a CPR simulation mannequin. In this environment, we invited participants to role-play as the caretaker of a sick person that could require immediate care. They aimed to stay in that room for a few minutes while we stepped out. If the simulation mannequin needed urgent care, an alarm would soar, and the Mixed Reality device or tablet would promptly display instructions they should follow.

### 2.3.2 Pilot Study

Using a simpler checklist, we ran a pilot study ( $n=5$ ) to understand better how users would interact with situated instructions. For that, we designed a two-step checklist that covers two steps of the cardiopulmonary resuscitation checklist: chest compressions and rescue breaths. Fig. 2.1 shows the Mixed Reality prototype as seen by the HoloLens on a mannequin. We also designed a similar two-step checklist as a single tablet screen.

In this pilot study ( $n=5$ ), two users were assigned to the Mixed Reality scenario and three to the baseline scenario. While mostly anecdotal, users on the Mixed Reality device were slightly faster than the ones using the tablet. However, we also noticed that users in the Mixed

Reality group struggled to find the content and execute steps in the order they were intended to be executed.

Building on this experience, we decided to design a more complex checklist and follow an iterative prototyping approach to improve the overall user experience while making sure that users adhered to instructions.

### 2.3.3 Iterative Design Process

After the initial 2-step prototype, we created a six-step checklist that resembled the Red Cross' CPR checklist. Although the tablet interface was trivial and mostly copied Red Cross' instructions and graphics, the Mixed Reality interface required us to understand better the process of foraging, parsing, and executing situated instructions. It took us a total of four iterations before users could complete all the tasks without hiccups. This complete CPR guide requires users to achieve a couple of repeated steps on the patient and locate an Automated Electronic Defibrillator (A.E.D) in the room and use it correctly on the mannequin (See next section for more information on the complete checklist).

We structured our iterative prototyping session the following way: First, we recruited participants (n=16. 7 females, 9 males) with varied backgrounds: Computer Science (4), Cognitive Science (4), Public Health (1), Biology (1), and Other/Undeclared (6). They all had no or limited experience with MR or HoloLens, and only 5 had been trained in CPR. With these users, we engaged in a three-phase iterative design process based on (1) user testing, (2) semi-structured interviews, and (3) rapid prototyping. After executing steps (1) and (2) with a sufficient number of participants, we would then engage in step (3) to design and implement a new iteration of our prototype.

1. **User Testing:** Using the role-playing scenario described above, participants used the latest Mixed Reality prototype to resuscitate a simulation mannequin. They were instructed to wear the HoloLens and follow the steps indicated by the application. Each session lasted



**Figure 2.2:** HoloLens 1 with a clicker tapped to the side. Pressing the clicker was faster and more accurate than hand gestures.

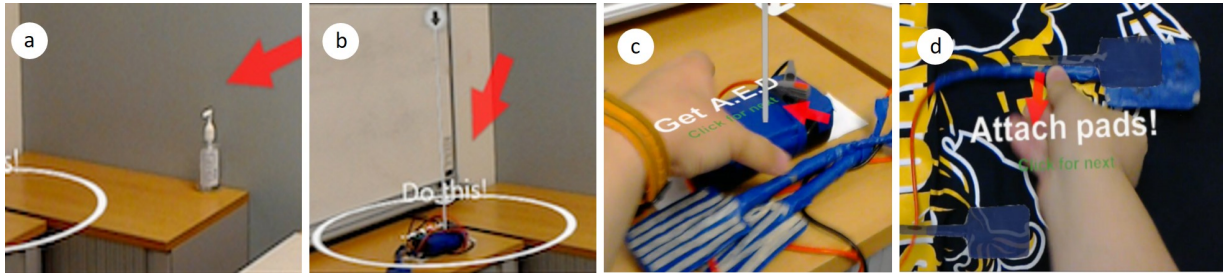
between 10 and 15 minutes. We watched these sessions remotely and took notes of any unexpected behavior. (i.e., not using the checklist as expected).

2. **Semi-structured Interviews:** After testing the prototype, we interviewed participants to understand any difficulties they might have faced in making sense of the instructions. We employed semi-structured interviews [26] to balance eliciting specific feedback on our prototype with new questions related to our observations. No
3. **Rapid Prototyping:** After parsing through interviews and observations, we would use low-fidelity prototyping techniques such as paper prototyping to quickly envision the next version of the AR interface. Then, we would implement the prototype using Unity 3D and deploy it to the HoloLens to be tested with participants in the new iteration. Development continued until we reached saturation, and no new or missing design element emerged.

## **Iterative Design results**

### **2.3.4 Design Challenges and Proposed Solutions**

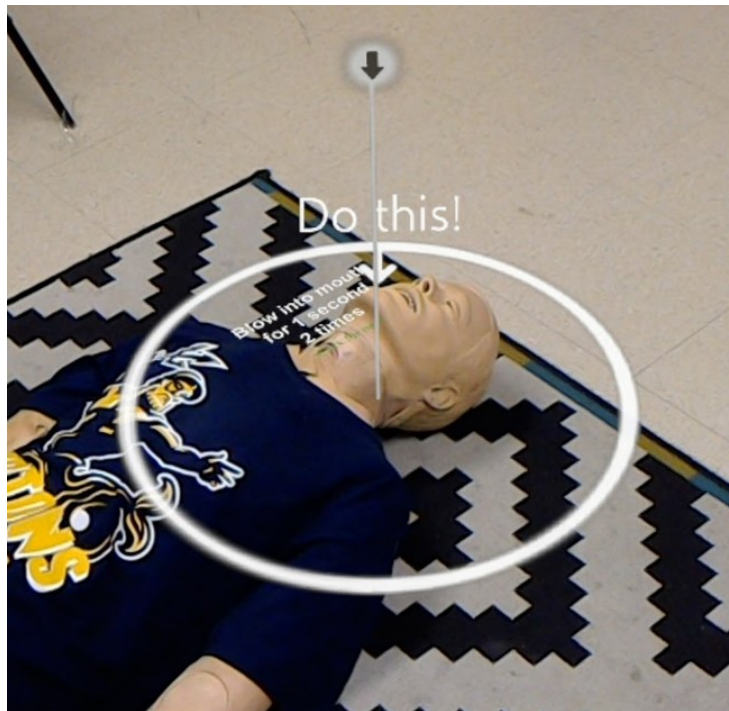
We hypothesize that mixed reality interfaces will overcome some of the problems of providing real-time guidance on traditional media. Through our design phase, we uncovered



**Figure 2.3:** The *Red Arrow* guides the users, indicating to them where the important things are located and where they should act. From left to right, the prototype shows how to first locate the AED (a, b) and then grab it and position the AED’s pad correctly on the victim’s chest (c, d).

a number of challenges that we did not anticipate but resulted in key design elements that we eventually integrated in our application, including 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.



**Figure 2.4:** Instructions for rescue breaths, as seen by the wearer. Visible are also the Circle of Attention and the Vertical Pole in the middle of it.

## Lack of familiar design patterns for information 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 [109]. On the other hand, MR does not inherently offer users a way to intuitively scan the interface for information. Besides some basic guidelines,<sup>1</sup> 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 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 2.4 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 them toward the instruction they needed to attend to. Figure 2.3 illustrates how the red arrow helps users find the AED pads elsewhere in the room and apply them correctly to the victim.

---

<sup>1</sup><http://developer.microsoft.com/windows/mixed-reality/design>

## **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. 2.5).

Although this is mostly a limitation of the current technology, we wanted to mitigate the limited FoV and compensate for 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. 2.4) 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 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 toward the center to find more instructions.

## **Close Ranges and Erratic Movement**

Current technical limitations are inherent to many MR devices cause visual aids to disappear when users get too close to objects or people in front of them. The 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.

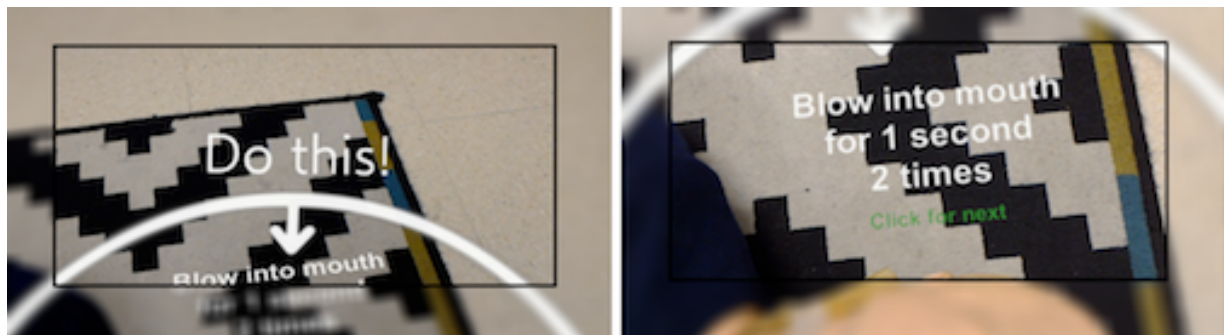
The circle of attention described above also served to help users retain their focus on the instructions, even when in close range to the digital content and when moving extensively. In addition, the constant presence of the circle of attention helped mitigate the effect of *disappearing* holograms.<sup>2</sup> 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.

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.

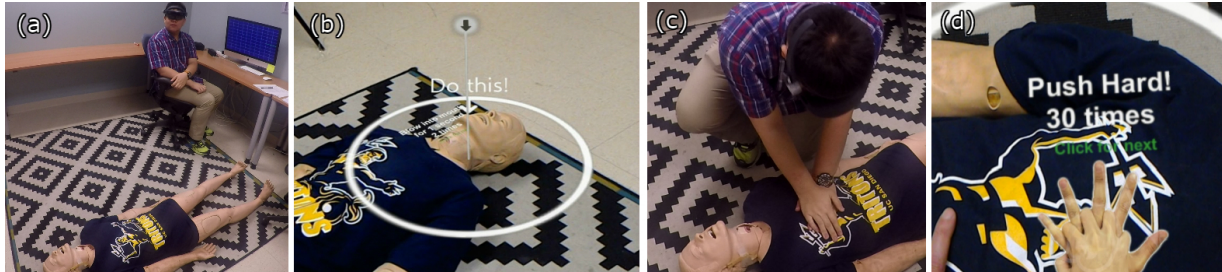
### Managing High User Expectations for Intelligent Systems

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

<sup>2</sup>E.g., when users lean in to breathe into the victim’s mouth, the head moves very close to the instructions, and digital content disappears.



**Figure 2.5:** 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 to the user, given HoloLens’ narrow field-of-view.



**Figure 2.6:** HoloCPR in Action. (a-b) As soon as the simulation manikin heart’s stops beating, the user looks at it to find spatially localized information of what he needs to do anchored where it he needs to perform it. (c-d) Wearer is performing chest compressions while seeing the instructions in place.

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. 2.2) 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.

## 2.4 HoloCPR

HoloCPR is a Mixed Reality application developed in Unity3D<sup>3</sup> for Microsoft HoloLens. It has two main parts. The first revolves around physical resuscitation, and the second aids the

<sup>3</sup><https://unity3d.com>

placement of an Automatic Electronic Defibrillator. The system was designed around the Red Cross CPR checklist <sup>4</sup> by breaking down instructions into “What”, “How”, and “Where” users should perform them (Figure 2.7). The “Where” is a crucial piece of information as it defines the real-world location HoloCPR uses to fix the instructions (Figure 2.6).

HoloCPR uses physical elements associated with the task at hand to display the instructions that should be performed where they should be performed (Figure 2.6, d). Thus, the environment gives context to instructions, allowing them to be short and deictic. As an example, Figure 2.6 (d) shows how HoloCPR communicates chest compressions with only four words and a picture. Figure 2.7 shows the alternative to the Mixed Reality interface: a paper checklist. Notice how the text is lengthened by the need to communicate the location of the procedure. Moreover, notice how the instructional picture shows the task from a third-person perspective, whereas HoloCPR (Figure 2.6 (d)) shows it from a first-person perspective.

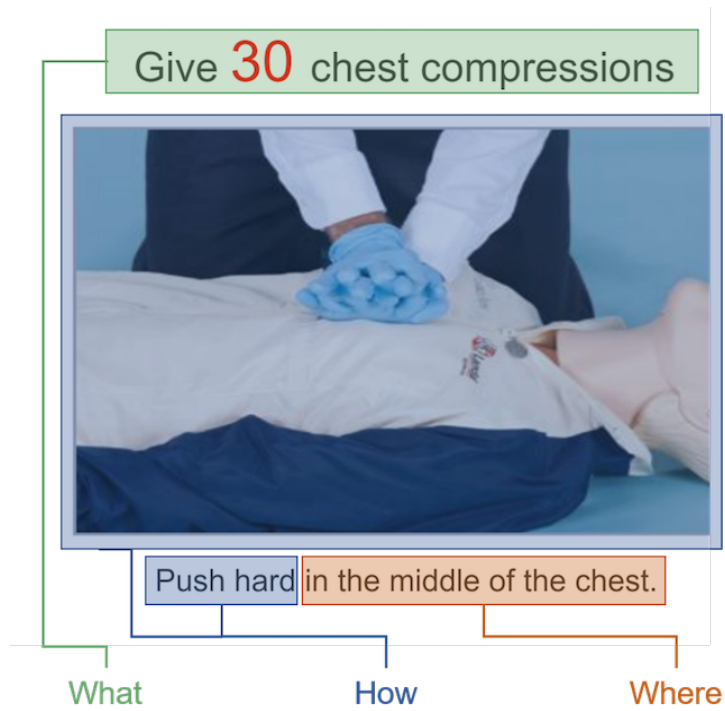
Similar to how mapping services (e.g., Google Maps) turn drivers that are new to a city into navigation experts, we envision HoloCPR to increase bystanders’ confidence in their ability to perform CPR during emergency procedures. In case of a cardiac arrest, wearing a HoloCPR headset in a first-aid kit would immediately take them to the checklist. By looking at the victim, the system would automatically indicate where to act first (see Fig 2.6, b), enabling users to address the task at hand. Similar to navigation instructions, steps are revealed as users progress in the rescue procedure.

We based HoloCPR’s resuscitation guidance on published and recognized guidelines taken from the Adult First CPR/AED Reference from the American Red Cross.<sup>5</sup> 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:

---

<sup>4</sup><https://www.redcross.org>

<sup>5</sup><http://www.redcross.org/cpr>



**Figure 2.7:** Dividing instructions into *What*, *How*, and *Where*

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

## 2.5 Evaluation

To evaluate HoloCPR, we conducted a between-subjects experiment in a realistic lab setting. Participants were tasked with executing CPR on a simulated mannequin, guided by the step-by-step instructions provided through HoloCPR. We assessed the duration they required to complete the simulation and the precision of their procedure execution. For comparison, we introduced a control condition replicating the conventional CPR process, which was guided by 2D instructions on a separate platform. The only independent variable in this setup was the medium employed for instruction delivery.

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 a 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.

### 2.5.1 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 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.

## 2.5.2 Experimental Design

*Setup* — Similar to our iterative prototyping, 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 a basic familiarity with its weight, the concept of holograms, and the clicker, we asked them to play RoboRaid<sup>6</sup> (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

---

<sup>6</sup><https://www.microsoft.com/en-us/hololens/apps/roboraidd>

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 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 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.

### **2.5.3 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 [35]. 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.

## 2.6 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 2.8 summarizes our results.

### 2.6.1 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.

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.

## 2.6.2 Performance and Transition Time

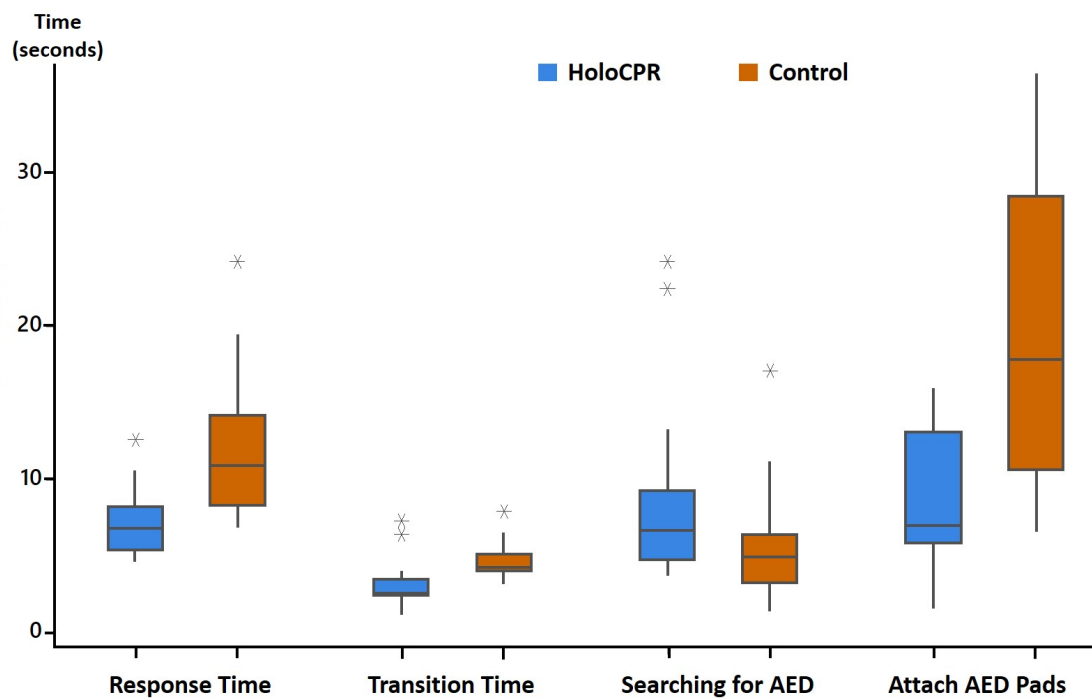
*HoloCPR users spend more time searching for AED* – While the difference is not statistically significant ( $p = 0.12$ ), we observed that the control participants generally spent less time searching for the AED (5.77 seconds on average) compared to HoloCPR users (8.61 seconds on average).

*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).

## 2.6.3 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



**Figure 2.8:** Response time, transition time, time spent searching for the AED, and time taken to attach AED pads across both conditions.

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.

## 2.6.4 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 like [P32] and [P41] noted how *“the system told me what to do - otherwise, I would have no idea”* 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”*. HoloCPR Participants were however 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 do something, it would've been nice to know ahead of time what I was being guided to".*

## **2.7 Discussion**

Overall we have been able to observe a trend toward 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 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 position hands and improves the 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.

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 instructions 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 victim. 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 the burden of reading and interpreting instructions, instilling a higher sense of confidence in users.

We also observed a higher degree of procedural adherence from participants using HoloCPR. 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 [117]. 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 influences 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 [117] which points to a need to design mixed reality interfaces to sustain users' attentional focus, without encouraging unwarranted reliability on the technology. This, in concert with a number of new and exciting possibilities, will be the focus of our future work.

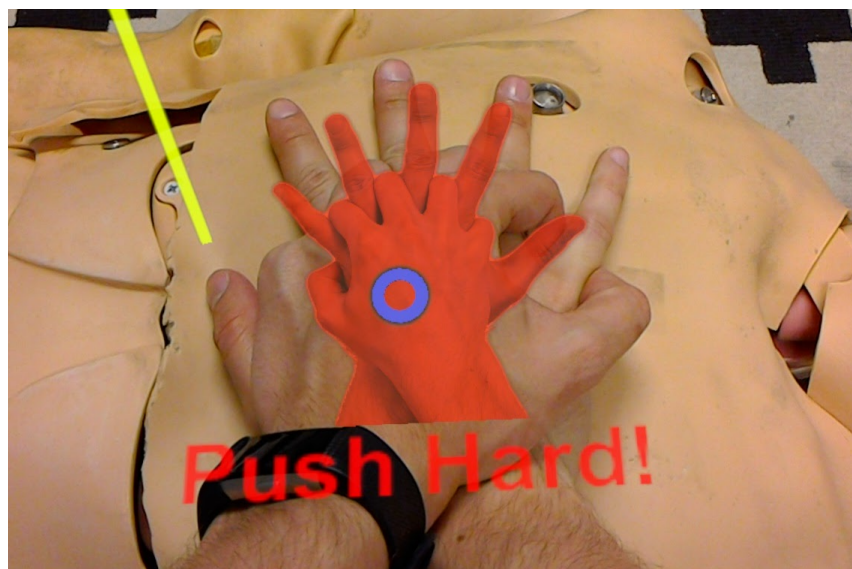
## 2.8 Closing Remarks

In HoloCPR [96], we designed an AR cardiopulmonary resuscitation (CPR) checklist. This interface showed users *what*, *where*, and *how* to perform the different steps of Red Cross' <sup>7</sup> CPR checklist. Our evaluation [53] showed that HoloCPR users are faster and adhere better to instructions than users following a comparable paper checklist.

HoloLens is often accurate enough to align AR instructions with the physical world. Occasionally, instructions were a few centimeters off, but users could still follow them. Look at Fig. 2.9 to exemplify a possible reason. It is not hard to perform chest compression on a mannequin, even if instructions are a few centimeters off the dummy's chest. There were, however, two notable user interactions with HoloCPR. One when HoloLens failed to track the environment and one when it was tracking perfectly. The first one happened during a catastrophic tracking failure. Virtual imagery skewed meters to the side, causing a user to jump under a

---

<sup>7</sup><https://www.redcross.org/>



**Figure 2.9:** HoloCPR: situated instructions showing the next step of the CPR checklist. Each instruction details *what*, *where*, and *how*, with *where* given by the location the instruction is displayed. In the example, the user has to push hard (what) with their hands (how) on the patient's chest (where).

nearby table to continue performing a chest compression even though no patients were there. The second notable interaction happened when another user deliberately placed the defibrillator pads in a different position than instructed. When we interviewed the subject and asked about their behavior, they mentioned that they could not trust the head-mounted display to show them exactly what to do.

Chapter 2 is, in part, a reprint of the material as it appears in the Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. Janet G. Johnson, Danilo Gasques Rodrigues, Madhuri Gubbala, and Nadir Weibel, from the Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. The dissertation author was a collaborator and contributor to this paper.

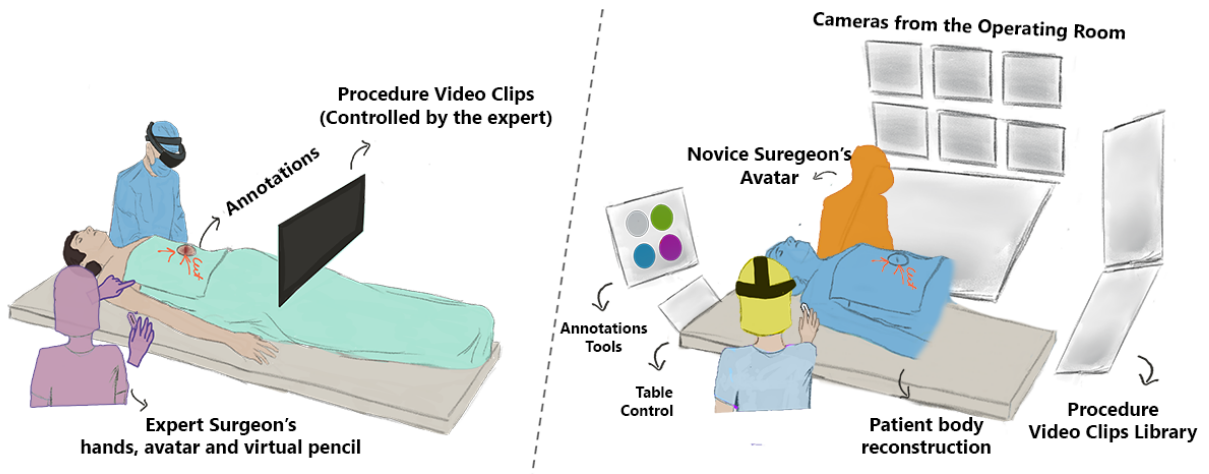
# Chapter 3

## ARTEMIS: Collaborative guidance in a trauma telementoring

### 3.1 Introduction

Whether from a car accident, gunshot wound, or traumatic injury due to natural or man-made disasters, trauma often requires timely life-saving interventions; however, the expertise required to perform these actions is not ubiquitous even within first-world medical care systems. When medical expertise or facilities are unavailable locally, surgeons rush to perform life-saving interventions that stabilize patient vitals before transferring the patient to a specialized surgical facility. When local surgeons lack the expertise to perform these life-saving interventions, they rely on remote guidance from expert surgeons through telementoring. Unfortunately, current telementoring systems limit how remote experts can guide and collaborate with local, inexperienced surgeons.

Telementoring uses information systems to provide real-time remote guidance to an inexperienced surgeon from an expert surgeon [5]. Today, telementoring systems typically require expert surgeons to instruct novice surgeons by annotating a live video of the surgical environment.



**Figure 3.1:** Artistic rendering of ARTEMIS and its features. Left: a Novice Surgeon in Augmented Reality receiving help from a remote expert. Right: a Remote Expert Surgeon in VR interacting with a 3D point-cloud of the patient, and engaging with the novice on a surgical procedure.

Remote experts must map actions they normally express through gestures and actions into limited video-based articulation. Novices operating on the patient have the extra burden of mapping instructions from a nearby screen to the operating field, with increased possibilities of making mistakes in the process [21].

Mixed Reality (MR) technology enables the creation of remote collaboration experiences where participants can interact with remote collaborators more naturally [17]. The space of remote collaboration in MR has seen many promising works focusing on training [108, 41] and, more recently, also on telementoring projects [5, 98].

This work takes a human-centered approach to designing a surgical telementoring system. Through role-playing sessions with US Navy surgeons and iterative prototyping, we gain insight into the features of MR that are useful in surgical and collaborative applications. Based on expert surgeons' and novice surgeons' communication needs, we designed ARTEMIS, a collaborative mixed-reality system for immersive surgical telementoring.

ARTEMIS was the *first* surgical telementoring system to give experts an immersive VR operating room where they can use gestures and 3D annotations on a 3D reconstruction of the

patient's body to guide novice surgeons in Augmented Reality. ARTEMIS intuitive AR experience allows local surgeons to focus on the patient and the tasks at hand, rather than having to interact with complicated AR interfaces.

This chapter introduces three key contributions to the human-computer interaction, surgical innovation, and user interface software and technology research space:

- ARTEMIS's design process, including software artifacts that facilitated our technology exploration
- The development of ARTEMIS: a real-time, mixed reality collaborative system for tele-mentoring surgical procedures
- An initial qualitative evaluation of ARTEMIS through mannequin and cadaveric subjects

While the COVID pandemic impacted a more thorough evaluation of this work, I link our initial assessment to the second evaluation led by investigators at the Navy Medical Center of San Diego.

## **3.2 Designing a Mixed Reality Telementoring System**

To design a collaborative mixed reality platform for surgery, we need to address two problems. First, we must understand how expert surgeons mentor inexperienced surgeons - their approach, their goals, and their unaddressed needs. Second, we need to understand how different MR interfaces can support their activities. Both problems are highly contextual and experiential. Moreover, the usefulness of an MR collaboration interface for a specific scenario might depend on various contextual and environmental factors.

To take into consideration the environment and the medical procedure, and to better understand user needs, we broke down our design process into two phases. (1) A role-playing

phase in a mock-up surgical environment, and (2) An iterative prototyping phase where experts performed procedures on mannequins and cadavers.

### 3.2.1 Role-playing in a mock-up operating room

To better understand how experienced surgeons mentor inexperienced surgeons as well as to receive feedback on the use of MR interactions in a trauma telementoring situation, we invited seven domain experts (4 surgeons and 3 operating room technology specialists) to participate in four role-playing [104] sessions. To contextualize the role-playing sessions, we simulated emergency procedures on a mannequin in a mock-up operating room (Fig. 3.2. *left*).

Each session was composed of two stages. First, we asked an expert surgeon to walk through an emergency procedure while they enacted it on the simulation mannequin. This *allowed us to observe* how mentors make sense of what they do – how they make decisions, how they communicate surgical steps to the novice surgeon, as well as what expectations they have from the novice surgeon. During this first stage, we asked questions to help us understand the specifics of the procedure as well as the specifics of their mentoring approach for that procedure. Then, in the second stage of the role-playing session, we invited experts to try existing MR applications and interfaces. We tested customized prototypes based on our observations as well as existing tools such as Microsoft Remote Assist [79]. This happened on an isolated part of the room where experts were unable to see the mannequin in the mock-up OR but were still able to talk to the designer who acted as a novice surgeon (Fig. 3.2. *right*). Finally, given our experts' backgrounds in battlefield care, we selected three commonly used procedures in military emergency scenarios [23, 22]: *needle decompression*, *leg fasciotomy* and *cricothyrotomy*. While we could not actually perform each procedure on our medical mannequin, we were able to walk through different stages of the procedure through printed images of a surgical video (i.e., each image showed a different stage of the procedure).

Role-playing is commonly used in the design of interactive systems as it can help users,



**Figure 3.2:** Our mock-up operating room (left) and office environments (right). In the left picture, we see a video frame from one of the role-playing sessions. In it, the expert enacts a *cricothyrotomy* on a medical mannequin while designers, developers, and other stakeholders engage in participatory design. In the right picture, we see the expert’s office space with a number of paper and digital tools that we used to explore how experts would interact with the novice surgeons.

designers and developers communicate needs and constraints. Hence, we also used these role-playing sessions as an opportunity to involve other team members as well as operating room technology specialists from the same hospital where we later deployed ARTEMIS. The most interesting results, however, were found in the sessions with four surgeons (S1-S4) .

In the remainder of this section, we report on our analysis of these role-playing sessions; our goal is to answer three key questions related to expert surgeons (tele)mentoring novice surgeons: **1) How do mentors guide mentees? 2) How much information does the mentor need from the mentee side? 3) What do mentors want to show to mentees?**

### 3.2.2 How do mentors guide mentees?

In a common, *co-located situation*, experts stand side-by-side with a novice surgeon as the novice surgeon operates on the patient (or cadaver). Novices are expected to have basic surgical skills, but often are unsure about the specifics of the procedure they are performing — “*where*

*I put the needle, where do I make the incision, how do I put the tube in” [S1].* Thus, experts chime in to explain to novices how to find the incision location, how to make the incision, and how to use other tools required by the procedure (e.g., tubes and clamps). To find the location of an incision, surgeons rely on anatomical landmarks - known body features - as well as touching and feeling the patient. According to the surgeons we interviewed, for many procedures, the hardest part of the procedure can be finding out where to make an incision. In a cricothyrotomy, for example, surgeons hold the patient’s chin with their left hand and probe the neck with their thumb to find a soft spot where the incision will be made.

After finding the location of the procedure, expert surgeons then proceed by showing novices the landmarks and marking paths where novices should make incisions with a skin marker. Finally, if the novice needs to use a specialized tool, the expert surgeon demonstrates how to use the tool by mimicking the motion and maneuvering of the instrument above the patient. Mentoring styles vary as experts may stand side-by-side with novices [S4] or across the table from them [S3]. In the unlikely scenario a novice is unable to replicate instructions, some expert surgeons move the novice to the side, position themselves in the same location as the novice, and then enact the procedure from where the mentee should perform it (*“I move that person away to show them how to do it right (and not do it)”[S3]*).

### **3.2.3 How much information does the mentor need from the mentee side?**

During role-playing sessions, we approached this question in two ways. First, as experts walked us through the different steps of a procedure, we asked them what information they used to make procedural decisions. We also asked them how they would be able to make a decision without that information. Second, in the technological exploration part of each role-playing session, we had them interact with existing AR and VR technology to help us understand how existing applications could support their efforts. In this short subsection, we summarize both findings.

First, as highlighted in the previous section, surgeons use a combination of touch and visual landmarks to locate the part of the body where they will operate. Unfortunately, we still lack technology that can accurately sense and replicate pressure, temperature, and other haptics in a sterile environment. Hence, we focused our research on visual tools. This restriction to the visual field may seem limiting, but it allowed us to further investigate alternative ways of displaying patient/procedure specific information in MR.

In the technological exploration, part of our role-playing sessions, we presented experts with various ways of seeing what the novice sees:

1. A first person view from the novice's augmented reality headset.
2. Multiple wide-angle and close-up views of the operating room, including a top-down view of the patient (See monitor setup on the left side Figure (3.2)).
3. A static 3D model of the mannequin.

For the first-person view, we used a Microsoft HoloLens 1 and its internal live-streaming application. Due to processing power limitations, the video quality lowers and becomes “blocky”<sup>1</sup> during sudden camera movements. Moreover, the resolution is considerably low (1216x684) if compared to a modern cellphone video camera (1920x1080). The other cameras consisted of Intel RealSense cameras and a Microsoft Kinect (v2) for the top-down view. For the static 3D model, we used a custom model of our mannequin in a Virtual Reality application on an HTC Vive headset.

**First-Person View:** *“All I want to see is what they’re seeing.”* [S3]. Unsurprisingly, all experts valued seeing exactly what the novice is seeing. S4 elaborates on possible reasons: For procedures such as leg fasciotomy, for example, after the first incision, novice surgeons get so close to the leg that any external camera looking from behind or from the top would be occluded

---

<sup>1</sup>A visual artifact of the video codec used to compress the video stream

by them. “As a mentor, I would want a very good view of where they are retracting” [S4] (and only a video stream from the novices’ perspective can show that).

**External cameras looking at the patient body:** When away from the mock-up OR, experts relied mostly on the first-person view. However, S4, after observing the first-person view feed for a while, said that the video stream was “choppy” and that he would like the “ability to see what is happening in the room, even around the learner.” As pointed out before, one of the technical limitations first-person view cameras is that as the novice surgeon moves around, the video stream both bounces with head movements and can decrease in quality (due to compression artifacts). When discussing possible locations for cameras, S4 mentioned the possibility of having a camera on wheels that an assisting nurse would be able to move around and re-position for per-procedure perspectives.

**3D representation of the patient body:** Most MR collaboration systems available at the time represented the novice’s space on a flat screen (e.g., Microsoft Remote Assist [79, 98]). Before developing a real-time 3D reconstruction system, we wanted to understand whether or not surgeons would be interested in interacting with a 3D representation of the body in Virtual Reality. Overwhelming feedback on the reconstruction was positive, provided it was in real time and clear enough to show more than just gross anatomy (“It would be important for the model of the patient to show the opened up leg” [S4]). Crucially, the three-dimensional representation would also allow for three-dimensional annotations.

### 3.2.4 What do mentors want to show to mentees?

Modern telementoring and remote collaboration systems augment the communication stream with deictic annotations. STAR [29, 64], for example, allows expert surgeons to annotate the novice’s video stream or physical space. This allows them to highlight specific objects as well as describe actions [36]. Recent works, however, have shown that combining two or more visual cues such as gestures and annotations can have a significant positive impact on remote

collaboration tasks [59]. In our role playing sessions, we wanted to understand which visual cues could lead to better surgical collaboration between mentors and mentees. From the expert surgeons' perspective, this section answers the following questions: What would help you best when guiding a novice surgeon? Do you need to bring in external content such as surgical videos or 3D anatomy? Should novices see virtual surgical tools? Where should novices see guidance information? This section summarizes this design exploration.

**Pointing and sketching.** *"I want to be able to do two things. I want to be able to point and say: 'this is your septum'. And then I want to be able to draw a line, like this, and say: 'make your incision here'. (...) [S3].* Building upon their approach of showing anatomical landmarks and then marking incision lines, experts wanted the system to afford a similar approach. S4 envisioned pointing with his fingers while S3 suggested the use of virtual pointers, similar to a thin pencil.

**Hands and gestures.** Hands and gestures play a key role in how experts communicate what to do next in a surgical procedure. *"now just lift that up and just gently spread like this"* ([S3] while gesturing what to do) *"I would want them to be able to see this"*. Experts not only use hands to point at a specific location on the patient's body, but they also use gestures to enact different steps of the procedures. This behavior was consistently observed even when experts were not mocking the procedures. S2, for example, was enacting the procedure with his hands while narrating out loud.

Overall, a virtual representation of hands can provide great versatility to surgeons as they use their hands to point, to demonstrate hand positioning/angling, to mimic tool behaviors (e.g. clamps and scissors), and to quickly enact different ways of performing the same procedure (e.g., *"Being able to show hands would be very helpful with the dissection as well, because for blunt dissection there are a lot of ways of doing it"*, showing with their hands different ways of doing a blunt dissection).

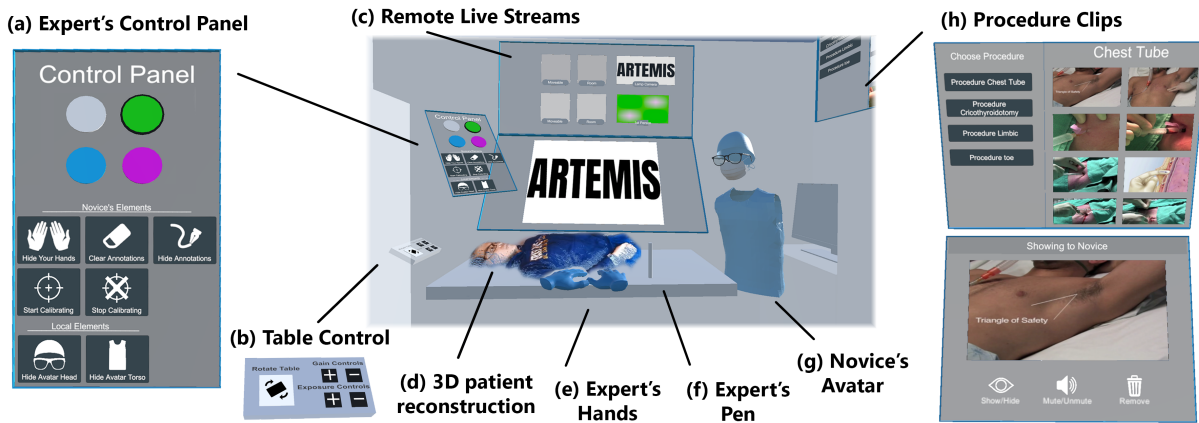
As we spent more time with the surgeons we realized how critical these gestures were,

not just because of their ability to show tool approaches and positioning techniques, but because surgery is fundamentally a physical act that cannot be easily verbalized. Many procedures rely on experts' tacit knowledge along side their strong understandings of both spatial layout and physical properties of anatomy. These understandings are difficult to convey verbally, but can be more easily communicated through physical gesturing.

**Annotations in 3D.** In the telementoring scenario, annotations allow experts to communicate the location and length of incisions. A important aspect of annotations for telementoring is the ability to communicate *depth* by drawing them in three-dimensions (“*It’s hard to show you on a non-three-dimensional thing.*” [S3, while explaining leg fasciotomy over a paper picture]). This became clearer when our experts telementored using existing MR collaboration technology such as Microsoft Remote Assist [79]. Similar to STAR [63], in Microsoft Remote Assist mentors sketch on a tablet and these annotations are then *projected* (ray cast) to the closest surface in the mentee environment. S4 was particularly concerned with that since procedures such as leg fasciotomy have “*a natural curve to it...*” and he wasn’t able to sketch that.

**Tools are not necessary, but could help in more complicated scenarios.** Most emergency procedures we role-played required specialized tools such as clamps, scalpel, metzenbaum scissors, hooks, tracheostomy tubes, and others. Experts had mixed opinions on whether or not they should have access to equivalent virtual tools; they thought that these tools could be enacted with their hands (“*The mentee can see the mentor hands right? I think that I can show that with my hands. I don’t think that the mentee need to physically see the tube*”) [S4]) but only when the procedures were not “too complex” (“*for something more complex than this, I think that I would want to have the actual instrument.*” [S3]).

**Egocentric guidance.** Easing the amount of cognitive workload on the novice was reinstated by experts several times during role-playing sessions. This impacts not only where information is displayed (“*(As a mentee) It would be beneficial to see information from my (point of) view*” [S4]) and how it is displayed (“*it would be great for the expert to use the same hand I*



**Figure 3.3:** ARTEMIS VR - Expert surgeon's interface in Virtual Reality. At the center, the main view that the expert surgeon sees. (a) Expert's control panel with colors for 3D annotations and buttons to control novice's calibration and visibility of hands. (b) Table control to adjust orientation, camera exposure and gain. (c) Remote Live Streams with two *moveable* cameras (RealSense on wheels), two fixed cameras (RealSense), one patient-focused camera (Kinect) and a first-person view from the novice's HMD (HoloLens). (d) Real-time 3D Patient Visualization. (e) Expert's Hands as seen in VR. (f) Expert's Pen. (g) Novice's Avatar (when visible). (h) Expanded View of the Procedure Clips (the lower part is only visible when the expert is watching a procedure clip with the novice).

*am using*" [S4] but also how much control novices should have over their MR interface ("*The learner is going to be fairly overwhelmed*" (*because they are doing a new procedure alone*) [S4]). Experts also agreed that it was a good idea to give novices as little control as possible: "*the idea of only getting an ON button is what most people would want*" [S4].

**Educational content can support very inexperienced surgeons.** Mixed Reality technologies creates an opportunity to bring 3D representations of human anatomy as well as other multi-media content novice's environment, something experts mentioned could be useful when novice has no familiarity with a procedure. While not their main goal while telementoring trauma surgery, experts expressed an interest in having access to a library of educational video clips that they could show novices. Interestingly, none of the experts expressed interest in showing 3D anatomy as — "*it varies a lot from person to person, and standard models won't be helpful*"[S4].

### 3.2.5 Summary

In summary, our analysis showed that when expert surgeons mentor novice surgeons, they focused on the following four expert goals (EG):

- EG1: Watch the procedure from the novice’s perspective
- EG2: Show the location of anatomical landmarks
- EG3: Mark the location, length, and depth of incisions
- EG4: Enact the use of tools

They need a system that enables them to fulfill all these four goals while also not overburden novice surgeons. Novice surgeons are already overwhelmed by the medical procedure at hand, and a new communication system shouldn’t distract them from their main goal: keeping the patient alive.

## 3.3 ARTEMIS overview

After a 12-month long iterative design and development work that included rapid AR prototyping [38] and role playing in collaboration with expert surgeons, we created the **Augmented Reality Technology-Enabled reMote Integrated Surgery (ARTEMIS)** system. ARTEMIS enables skilled surgeons and novices to work together in the same virtual space and approaches the problem of remote collaboration through a hybrid interface: expert surgeons in remote sites use Virtual Reality to access a 3D reconstruction of a patient’s body and instruct novice surgeons on complex procedures; novice surgeons in the field focus on saving the patient’s life while being guided through an intuitive Augmented Reality interface. From now on, we will refer to the expert interface as ARTEMIS VR and the novice interface as ARTEMIS AR. Also, to contextualize our

implementation choices, we will refer back to the Expert's Goals (EG) when describing features of ARTEMIS AR or VR.

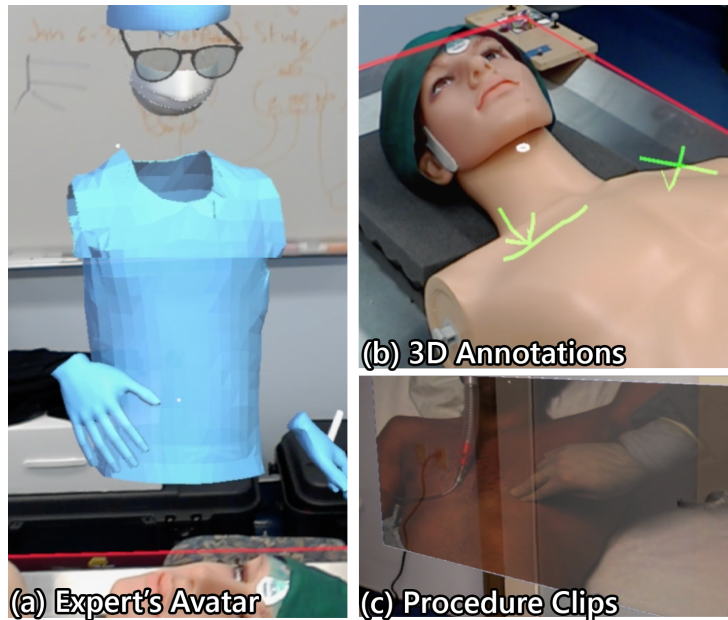
The novice bedside environment is equipped with a Microsoft HoloLens v1 [75] worn by the novice surgeon, 5x depth-cameras to capture the 3D scene (1x Microsoft Azure Kinect [74] placed on top of the bed and attached to a surgical lamp, 2x Intel RealSense cameras [49] in the corners of the room, and 2x Intel RealSense cameras on wheels movable in the OR), and an OptiTrack optical marker system [89] to track the movements of objects and people in the room (markers are attached to the HoloLens, the Kinect camera, and the surgical table).

The expert's remote environment is equipped with a video-see-through Virtual Reality headset (HTC Vive Pro [48]) provided with a wireless adapter that makes it untethered, hand+finger tracking for the expert surgeon based on an IMU-equipped gloves [114], a wireless mouse-pen used for annotations and VR-based interactions, and an OptiTrack optical marker system to track the movements of objects and people in the room (markers attached to the Vive Headset, surgical table proxy, gloves, and wireless pen).

Figure 3.5 shows ARTEMIS in action during one of our evaluation studies, and outlines the devices used for both ARTEMIS AR and ARTEMIS VR.

### **3.3.1 Novice Surgeon's Interface**

The novice surgeon's main goal is to operate on the patient. To avoid distractions, the ARTEMIS AR interface components are passive, and by design, the novice surgeon is not able to *directly* interact with the interface. All the features described below are controlled remotely by the expert; for instance, if the novices need to hide annotations or play a procedure clip, then they can request the expert surgeon to do so. This decision emerged directly from our role-playing sessions where it became clear that novices were not able to directly interact with an AR application while operating on a patient. Our hypothesis is that this resulting interface will not overwhelm novice surgeons nor distract them from their operating tasks.



**Figure 3.4:** ARTEMIS AR Interface - Novice surgeon's interface in Augmented Reality. (a) The Expert's avatar shows them holding a virtual pencil. (b) 3D Annotations as seen by the novice. (c) Procedure clips showing a video clip.

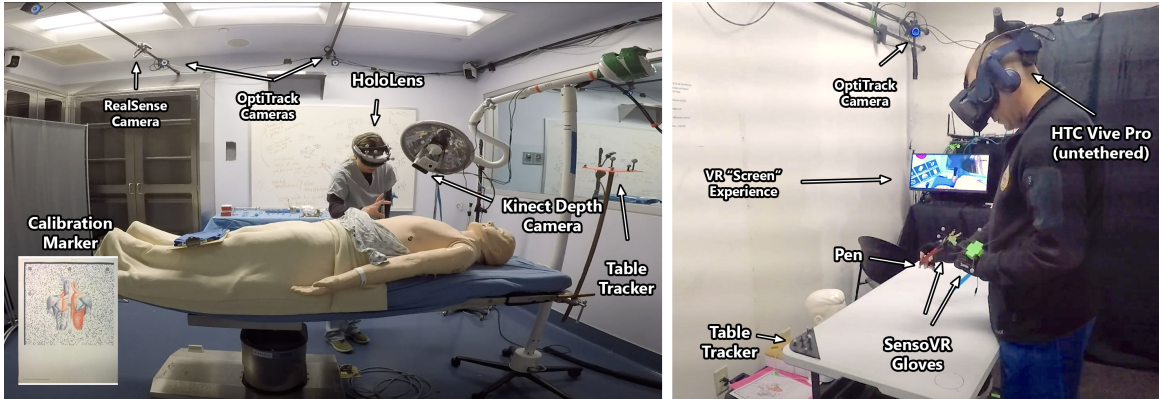
The novice surgeons are able to see three main holographic representations in front of them (Fig. 3.4): (a) the expert's avatar and hands, (b) a remote pen and 3D annotations, and (c) procedural video clips. In addition, the novice surgeon is able to provide direct views of the patient and the surgery to the remote expert, by manipulating the position of the Kinect camera and by directly attending to regions of interest with the HoloLens device's camera.

***Expert's Avatar and Hands*** – The novice surgeon can see both the expert surgeon's location and their hands (Fig. 3.4a). This enables the expert surgeon to communicate through gestures, for example, by pointing to a location on the patient's body (EG2) or by showing how to handle a surgical tool (EG4). The expert surgeon's avatar automatically disappears if the novice surgeon walks into their virtual location. This interaction allows the novice surgeon to repeat what the expert surgeon is gesturing in a more intuitive way as they are both doing it from the same point of view. In other words, the expert surgeon's hands can act as a second pair of hands that originate from the novice surgeon's body and guide the novice step-by-step.

***3D Pen and Annotations*** – 3D annotations (Fig. 3.4b) allow the expert surgeons to instruct the novice by 3D sketching over the patient’s body. Because these annotations are in 3D, they can directly communicate depth, length, and area (EG3), which are critical for surgical procedures such as incisions, tracheotomies, thoracotomies, etc. To allow for additional interaction space for the remote expert, 3D annotations can also happen in mid-air and are not limited to the patient’s body. To facilitate the novice surgeon’s understanding of where the annotations will show up, if the remote surgeon is holding a pen, also ARTEMIS AR shows a 3D model of a pen. This is the same 3D pen that expert surgeons see in their VR interface (see below), and it is shown in the hands of the expert surgeon’s avatar when in use.

***Procedure video clips*** – Procedure clips are an additional resource used to support guidance during remote telementoring. They are instructional, short video clips of different steps of specific surgical procedures that are available for the expert to show to the novice when needed (Fig. 3.4c). These video clips show up as a floating screen on top of the surgical table (Fig. 3.1) and always face the novice surgeon. They keep repeating until disabled by the expert, and typically also contain audio instructions, that can be muted by the expert surgeon if needed.

***Scanning the patient through the surgical lamp*** – ARTEMIS shows a 3D reconstruction of the patient to the expert surgeon in Virtual Reality (Fig. 3.3d), but the Microsoft Kinect depth camera that scans the patient is in the operating room (OR). To provide an intuitive interface that allow the novice surgeon to both know what the expert surgeon can see and easily change that view as needed, we designed the system to take advantage of the OR layout, as well as the interaction that we observed during our role-playing session; we, therefore, decided to attach the depth camera to the OR surgical lamp that is illuminating the patient (see Fig. 3.5, left). In this way, the novice surgeon is always aware of what the expert surgeon can see as the lamp illuminates that part of the patient.



**Figure 3.5:** ARTEMIS in action. Left: ARTEMIS AR, with a novice wearing HoloLens and in the process of performing a surgical procedure on a mannequin; on the ceiling-mounted rods are visible the OptiTrack and RealSense cameras (on the left rode), while the Kinect is attached to the surgical lamp in the middle (the white zoomable cameras on the left and right of the image are not part of the system). Right: ARTEMIS VR, with an expert surgeon guiding the novice user, wearing the IMU-equipped gloves, and holding the pen to annotate the 3D point cloud in front of him; visible are the OptiTrack cameras on the ceiling-mounted rods, and a display showing the current VR view in the background.

### 3.3.2 Expert Surgeon’s Interface

The expert surgeon’s Virtual Reality interface provides surgeons with a virtual operating room. In this operating room, the 3D reconstruction of the patient is at the center of the stage, surrounded by tools and VR camera views (from the Kinect, RealSense, and HoloLens cameras) designed to make the expert surgeons’ interaction with ARTEMIS more efficient and enable the expert to successfully explain procedural and conceptual steps of the surgery being performed. Figure 3.3 shows the entire VR interface and highlights its key components.

The expert surgeon interacts with the interface via a laser pointer, but unlike most VR experiences, this laser pointer is implemented through an actual physical pen, and not as a virtual tool that they “grab” with VR controllers. ARTEMIS VR does not use regular VR controllers to facilitate the use of gestures and tools by an expert surgeon that is most likely new to VR experiences. Thus, instead of having users learn how to map controller buttons to gestures and actions, the expert surgeons wear real gloves as if they were wearing surgical gloves (Fig. 3.6a) .

Their only surgical tool is a pen that they can hold and feel with their hands and can point, select, or draw (Fig. 3.6b).

**Control Panel** – The expert’s control panel (Fig. 3.3a) provides four types of tools: (i) annotation controls and color palette, (ii) local space controls, (iii) novice surgeon’s interface controls, and (iv) calibration controls. Through the annotation controls and color palette, the expert surgeons can change the color of the pen before making a 3D annotation. They can also erase annotations on both their side and the novice surgeon’s side. The local space controls (Fig. 3.3a, bottom) allow experts to show and hide the novice’s head or torso. The Novice surgeon’s interface controls allow experts to change the visibility of their hands and annotations on the novice surgeon’s space. Finally, the calibration controls allow experts to work with the novice surgeon to improve the alignment of annotations as seen from the novice surgeon’s side. We will elaborate on the needs and the approach for alignments later in this and the next sections.

**Table Control** – This interface serves two purposes (Fig. 3.3b); the *Rotate Table* flips the orientation of the patient so that the expert surgeon can look at surgical field from two different points of view; the *Gain* and *Exposure* controls allow the expert surgeon to control settings of the remote Kinect camera, adjusting the visibility of the patient as needed.

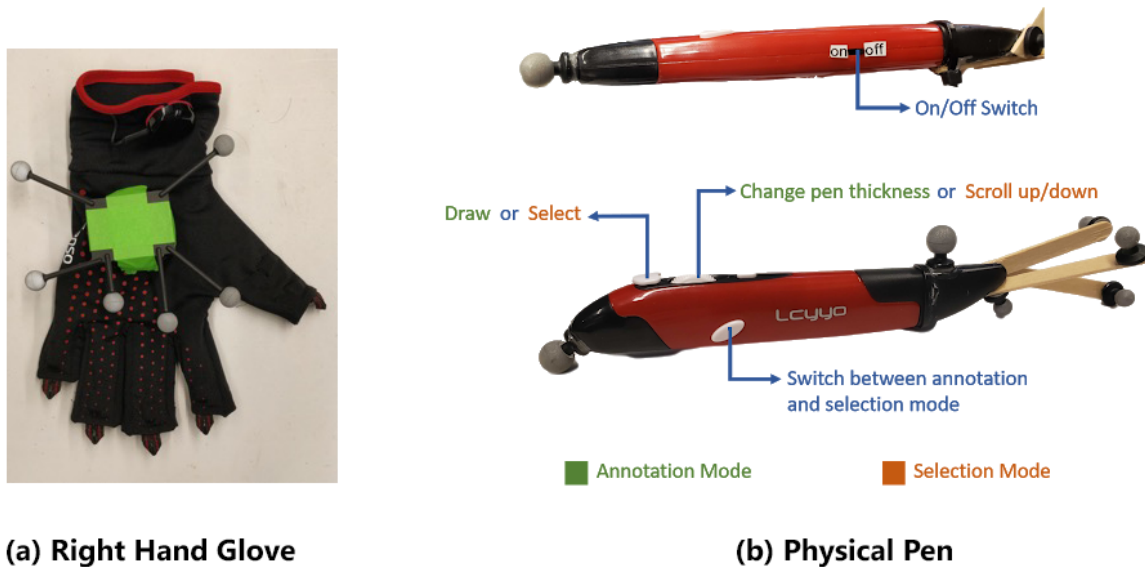
**Remote Live Streams** – This interface (Fig. 3.3c) shows six different live video streams to the expert surgeon in the top part. Two displays show cameras attached to rolling wheels that the novice surgeon can move around the operating room. Two displays show cameras located at opposite corners of the operating room. One display shows the internal camera of the head-mounted display, and the last display shows the patient as seen by the depth camera attached to the surgical lamp. The expert surgeon can use the pen as a laser pointer to select and show any of these six videos streams in the bigger display at the bottom. The location and layout of these displays allow for the user to see both the patient reconstruction as well as the video displayed in the bigger screen without the need to switch focus or move their heads.

**3D Patient Reconstruction** – This is the central interface available to the expert surgeon

to guide the novice through specific surgical procedures (Fig. 3.3d). Through a point-cloud view, the expert surgeon can see the patient in a three-dimensional rendering that keeps real-world proportions. The point cloud view is a real-world live representation of the patient and it is placed on top of an actual table in the remote expert’s environment. By looking at the point cloud the expert can see in real-time what is happening to the patient, and can interact with the patient representation by placing hands on particular parts of the body, and by annotating the body using 3D annotations. Both hand maneuvers and 3D annotations show up in real-time in the AR view of the novice.

**Novice Surgeon’s Avatar** – The novice’s avatar (Fig. 3.3f) shows the location of the novice surgeon with respect to the surgical table at all times. Experts use the avatar as a communication and interaction anchor when guiding the novice through their procedures.

**Procedure Clips Control** – This interface provides a video library containing a number



**(a) Right Hand Glove**

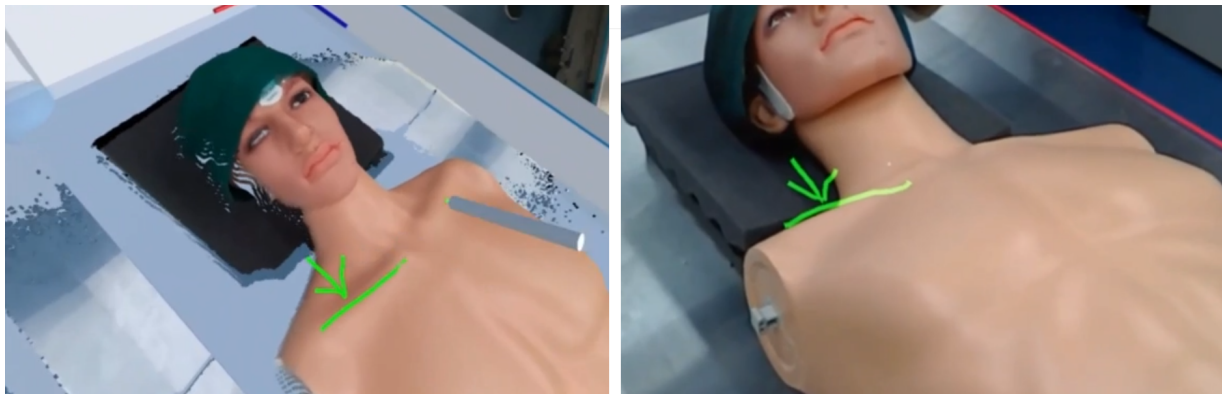
**(b) Physical Pen**

**Figure 3.6:** ARTEMIS VR Input Interfaces. Expert surgeon’s wear gloves and interact with the interface through a physical pen. (a) Right-Hand glove showing OptiTrack markers (small silver spheres) installed on the top side of the glove. (b) Physical pen, implemented using a wireless mouse, and showing OptiTrack markers; highlighted are its two modes: selection mode and annotation mode.

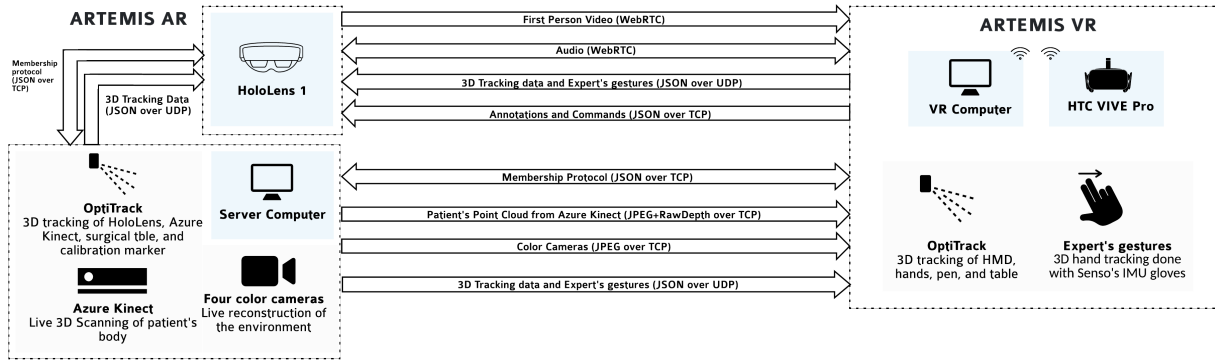
of surgical procedure video clips for different procedures (Fig. 3.3h). By selecting one of the options, a series of video clips pop up on the right side of the interface (for instance the chest tube procedure); when the expert surgeon selects one of these video clips, it displays on the larger screen at the bottom of the interface, and it plays synchronously on both the expert side in VR, and as holographic representations on the novice surgeon's side (Fig. 3.4c). The expert surgeon can show, hide, pause, mute, and remove this interface from both the expert's and the novice's side.

### 3.3.3 Calibrating 3D annotations

Because novice surgeons may *only* see the 3D annotations floating in front of them (Fig. 3.7, right), annotations displayed at the wrong location can lead them to perform an action (i.e. an incision on the patient's body) at the wrong place. To avoid this, before starting a procedure, expert surgeons can use the calibration buttons in the Control Panel (Fig. 3.3a) to calibrate the novice surgeon's head-mounted display. To calibrate, the novice surgeon brings a *Calibration Marker* (See Fig. 3.9) under the surgical lamp view. The expert surgeon then selects "Start Calibrating" in the Control Panel (Fig. 3.3a); when the novice surgeon looks at the calibration marker through the HoloLens camera, the system performs a calibration between the



**Figure 3.7:** The expert surgeon annotates on the reconstruction of the patient body in VR (left) while the novice surgeon sees it overlaid onto the real patient in AR (right)



**Figure 3.8:** ARTEMIS is composed of two separate environments: ARTEMIS AR (left) and ARTEMIS VR (right). This diagram shows the origin and direction of data flowing from each environment.

different tracking devices. The expert surgeon can confirm alignment accuracy by annotating over the calibration marker.

### 3.4 Implementation

As explained earlier in this paper, ARTEMIS consists of two separate spaces: ARTEMIS AR and ARTEMIS VR. ARTEMIS AR encompasses the novice surgeon's AR head-mounted display and the server connecting to the operating room's cameras and trackers. ARTEMIS VR encompasses all the tracking hardware used for the expert surgeon. Each computing device (HoloLens v1, Server Computer, and VR computer) runs an application developed in Unity 2018.3 [112]. Figure 3.8 summarizes the ARTEMIS' hardware and streams of data.

ARTEMIS AR uses HoloLens v1 [75] as the AR headset and a standalone computer (*Server Computer*). HoloLens sends audio and video directly to ARTEMIS VR through WebRTC [78]. It also receives tracking data and commands directly from the *VR computer*. We use a separate computer, the *Server Computer*, to encode and stream data from the many cameras installed in the operating room. The Azure Kinect depth camera [74] is used to create the patient's point cloud at a resolution of 1280x720, 30fps. The *Server Computer* also encodes and streams

frames from four Intel RealSense color cameras [49] (2 on rolling wheels, 2 in the corners of the room) at a resolution of 800x600. Finally, we use five OptiTrack Prime 13 motion capture cameras [89] to cover the entire surgical space. These cameras track HoloLens, the surgical table, the surgical lamp, and the calibration marker used to realign annotations. Tracking data from these cameras flow from OptiTrack’s Motive [87] to the *Server Computer* and then to both the HoloLens and the VR Computer through a custom protocol (Fig. 3.8, left).

ARTEMIS VR uses an untethered HTC Vive Pro [48] as the VR headset. Similar to the AR space, it also uses OptiTrack motion capture cameras to track the VR headset, the pen, the gloves, and a physical table that serves as a proxy to the surgical table (Fig. 3.6 shows the spherical passive infrared OptiTrack reflectors attached to the gloves and the pen). The pen and the gloves [114] are commodity hardware that connects to the VR computer through Bluetooth. We use OptiTrack’s active Vive tracker [88] to track the VR headset instead of Vive’s default tracking technology.

In developing ARTEMIS, we faced four key technical challenges: (1) HoloLens limitations, (2) a constant change of input and interaction, (3) rendering the point-cloud without distortions, and (4) calibrating the different coordinate systems. The remainder of this section describes the specific software and hardware strategies we used to address these challenges.

### **3.4.1 Overcoming HoloLens Limitations**

Released to the public in 2016, HoloLens v1 was the first commercial, untethered optical see-through head-mounted display to use inside-out global sensor fusion for head-tracking [62]. However, its capability to track itself in space, makes HoloLens integration with external tracking systems such as OptiTrack quite challenging, especially when HoloLens’ internal tracking system drifts over time [113]. To check that devices are properly integrated and therefore ensure smooth user experience, as part of ARTEMIS we created an interface for the expert surgeon and the novice surgeon to verify that OptiTrack-HoloLens calibration is accurate.

As an untethered device, HoloLens v1 runs on a battery that lasts around 2-3 hours during continuous use [75], but our surgeons were (correctly) worried that some procedures could last longer than that. To allow surgeons to switch to a new device quickly, each device connects to the *Server Computer* that acts as a hub and advertises that a new AR/VR display is available. This allows a newly connected HoloLens to quickly take over an existing session when a device is about to run out of battery.

As an embedded computer, HoloLens has also limited processing power. For example, our procedure clips (Fig. 3.3h) would not play while WebRTC encoded and streamed the front-facing camera. We disable the first-person view whenever the expert surgeon plays a procedure clip to resolve this problem. In these situations, we expect the expert surgeon to rely on the other cameras to gather more visual information about what the novice surgeon is doing.

### **3.4.2 Rapid Technology integration**

ARTEMIS implementation is the result of several iterations over a period of 12 months. During this time, we tested different commodity technologies to address the interaction needs we uncovered during our role-playing sessions with experts. AR and VR technologies are still maturing and off-the-shelf products have limitations that only show up through extensive testing. For example, we initially integrated Leap Motion [111] on the VR headset to track the expert hands. Still, our tests showed that Leap Motion fails with different gestures when fingers overlap each other. As a result, we integrated VR gloves as a replacement for Leap Motion and later upgraded again to a computer-vision hand-tracking system.

Often, switching from one commercial technology to another means migrating to a new Software Development Kit (SDK) with different programming languages or environment support. To facilitate our rapid technology integration, we implemented abstraction layers so that the expected data from each side were independent of the hardware and software used. For example, the point cloud is encoded as JPEG and Raw 16 bits depth. We then used a dedicated networking

library to connect and synchronize data streams across systems.

*Networking Library* – Unity has poor support for custom, high-throughput networking protocols. With all relevant data going and coming through the network, we implemented a networking library for Unity with support for Python, C++, and other platforms such as NodeJS. The Unity counterpart of the library provides a high-level interface to TCP clients and TCP servers as well as UDP sockets. It also receives and decodes network packets in an external thread to avoid impacting rendering performance. With the exception of the WebRTC audio and video streams, we implemented all the network streams described in Figure 3.8 with this library.

### **3.4.3 Rendering the Point-cloud**

We implemented a custom application to read, encode and stream color and depth frames from Kinect Azure. Color frames are encoded into JPEGs and depth frames are sent raw through a C++ interface part of our networking library. On the VR application, we use a custom shader to render the color and depth frames as a point cloud (Fig. 3.7, left). Our shader uses a pre-computed look-up table to calculate the point cloud from the depth map similar to Azure Kinect Fast Point Cloud example [73].

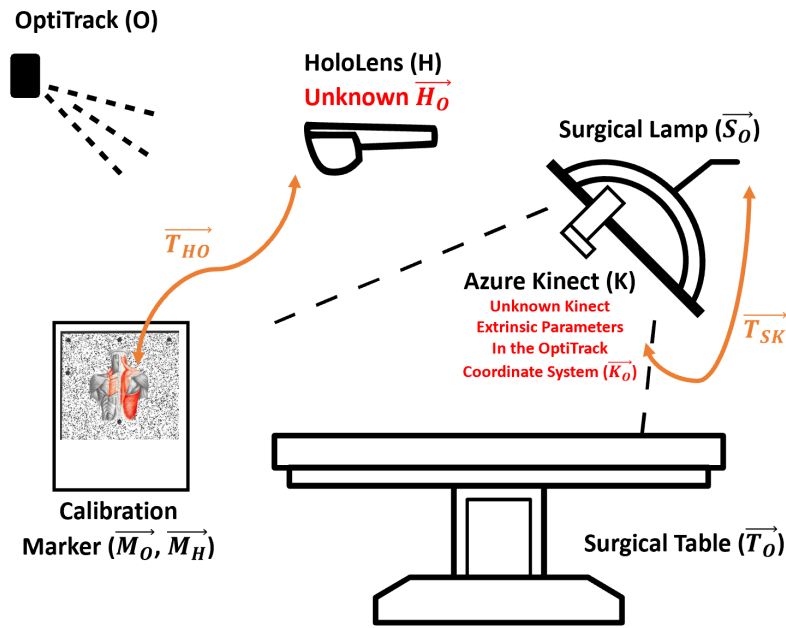
The AR space tracks both the surgical table and the surgical lamp equipped with the ARTEMIS depth camera. The novice surgeon can move the camera around the patient, and the expert surgeon will see the different parts of the patient at different locations on their virtual surgical table. Moving the surgical table and the lamp together will not move the virtual surgical table or the patient point cloud in VR.

### **3.4.4 Calibrating Different Coordinate Systems**

Showing annotations on the patient’s body requires us to know their location with respect to the HoloLens coordinate system. Unfortunately, HoloLens is unable to directly track the body

of the patient. We rely on OptiTrack to track both the patient and HoloLens so that HoloLens can translate incoming annotations from the patient’s body coordinate system to its internal coordinate system. We summarize the different coordinate systems and required transformations in Fig. 3.9. Overall, there are two main transformations needed.

The first step is to establish a coordinate system for the patient’s body. This requires a single, offline calibration between Azure Kinect and OptiTrack. We perform this calibration through a variation of Chiodini et al. [25] using one camera instead of using a stereo setup. This calibration gives us  $\vec{K}_O$ , the location of Azure Kinect with respect to the OptiTrack coordinate system ( $\vec{O}$ ). Given that Kinect is not directly tracked by OptiTrack, during calibration, we use the surgical lamp location ( $\vec{S}_O$ ) to find a transformation between itself and Kinect ( $\vec{T}_{SK}$ ).  $\vec{T}_{SK}$  allows us to know the location of Kinect at all times, even after a surgeon moves the surgical lamp.



**Figure 3.9:** Diagram representing all tracked objects in the Augmented Reality space. We represent the three independent coordinate systems HoloLens, OptiTrack, and Kinect by H, O, and K, respectively. OptiTrack tracks all physical objects of interest: the calibration marker ( $\vec{M}_O$ ), the surgical table ( $\vec{S}_O$ ), the surgical lamp ( $\vec{L}_O$ ). The goal of the calibration is to find the transformations  $\vec{T}_{HO}$  and  $\vec{T}_{SK}$ .

The second step is to find  $\vec{H}_O$ —HoloLens' location with respect to OptiTrack ( $\vec{O}$ ). Unfortunately, HoloLens' internal tracking system resets its origin every time ARTEMIS AR is started. This requires us to find  $\vec{T}_{HO}$  every time we start the system. To facilitate calculating  $\vec{M}_H$ , we created a Calibration Marker that can be tracked both by OptiTrack and HoloLens. We use Vuforia [115] on HoloLens to track the marker through its front-facing camera. With known equivalent points  $\vec{M}_H$  and  $\vec{M}_O$  we can find a transformation between HoloLens and Optitrack ( $\vec{T}_{HO}$ ).

## 3.5 System Evaluation

ARTEMIS is being prospectively evaluated within a cadaveric trauma model. Novices are randomly assigned to ARTEMIS support or the Control group (audio only consultation), and evaluation is performed retrospectively by experts blinded to the study participants group (as all participants wear a headset platform) using standardized graduate medical education methodology. Key outcome measures are surgical technique (evaluated using the OSATS tool [68]), critical decision making, and timeliness of life saving interventions.

It is outside of the scope of this paper to further discuss the large cadaver-based randomized trial or to understand if ARTEMIS could be used for training and for the provision of emergency care. We, however, performed a qualitative evaluation of the use of ARTEMIS and we report here on initial results involving 5 experts, 6 novices and 22 procedures on mannequins and cadavers.

### 3.5.1 Procedure

*As part of the initial deployment of ARTEMIS at Naval Medical Center San Diego, we enrolled 6 novices (N1-N6) and invited 5 experts (E1-E5) to work with the novices. Together they participated in pilot studies with mannequins, as well as actual studies with cadavers. Our research team engaged in direct observation, and all participants were video recorded. Direct*

*observations from cadaver studies were used in conjunction with videos from mannequin studies.* The study protocol was approved by the Naval Medical Center San Diego Institutional Review Board in compliance with all applicable federal regulations governing the protection of human subjects.

Of the five experts, two were Senior Critical Care Intensivists, and three expert were Staff Surgeons. The six novices were Surgical Technicians, Medics, and Junior Surgical Residents. One of the experts mentored two different novices.

The five experts engaged in a total of 22 procedures across all of the sessions with novices, with two novices only performing one procedure (*cricothyroidotomy*), and the other four performing five back-to-back procedures each (*cricothyroidotomy*, *dual-incision leg fasciotomy*, *femoral artery exposure*, *axillary artery exposure*, and *resuscitative thoracotomy*). Procedures on mannequins did not entail any actual incision, while cadaver studies performed the procedure as if it was on a real patient. All but one experts (E1) and all novices were trained on the system, but neither experts nor novices knew in advance the procedure to perform. Sessions covering one to five procedures spanned 15-60min. All experts and novices had the VR/AR headsets (HTC Vive Pro and HoloLens 1) calibrated for their inter-pupillary distance (IPDs)

The research team was composed by at least three researchers (sometimes four) who observed all the sessions, took notes, and analyzed the videos of the procedures to uncover usability issues, usage patterns, communication and coordination processes specific to immersive remote telementoring. Experts and novices were also informally interviewed at the end of their session to gather their feedback on the use of the system. We summarize results of this initial qualitative evaluation in the next section. *To structure the results as presented below, we conducted a qualitative synthesis through a thematic analysis that aggregated observations from videos, interviews, and researcher notes.*

### 3.5.2 Results and Discussion

In general ARTEMIS worked well, and we witnessed both novices and experts being able to communicate with increased precision, accuracy, and clarity. Novices were able to successfully complete the procedures assigned to their sessions, including those that they had never performed previously. The overall feedback from experts when asked about the system is nicely summarized by E2's comment "*To do those 5 procedures in 40 minutes, especially 2 of which he's never done before ... is pretty great*" [E2].

After reviewing the data collected during our exploratory evaluation, we organized our findings into seven major themes. The first three themes cover technology limitations and interaction limitations. The last four show us how ARTEMIS enabled experts to overcome those limitations, and how the system accommodated for different mentoring styles.

**(1) *The 3D point cloud doesn't represent fine details*** – 3D point clouds are as reliable as the depth information associated with each pixel. In situations where experts needed a closer look of the body, novices moved the surgical lamp supporting the depth camera to cover distant parts of the body (e.g. the legs during a fasciotomy), as well as to have better views (e.g. the side of the chest during a lateral puncture). Unfortunately, modern depth cameras still have a huge gap in resolution between the depth camera and the color camera. Overall, we found out that point-clouds are unable to *represent finer details [107, 90] such as veins and soft tissue such as the fascia*.

**(2) *Annotations' depth and alignment are difficult to perceive*** – One of the most difficult tasks for both experts and novices was to understand how to make annotations at the right depth (expert in VR) and at what depth were annotations done (novice in AR). Experts using VR for the first time had some difficulty making annotations at the right depth. This difficulty with the interface is partially due to depth perception in VR. VR headsets are still limited in how they model and create depth perception [91]. In our studies we had to explain to one expert that they were writing annotations far above the actual 3D body because they were looking at the body

from the top: “*here, bend low / crouch there to annotate the body*” - “*Oh*” [E4] – referring to stereoscopic understanding of depth. *While we tackled this limitation by training our expert surgeons, a future venue of exploration is the use of visual aids and sketching constraints [7].*

**(3) *Not everything is being shared, but this is not always clear*** – During our first tests, E1 tried to guide the novice surgeon by referring to one of the many video clips visible to him on the procedure clips library interface (Fig. 3.3h). It took him some time to realize that the novice was only able to see one selected video, and not the entire video library. As mentioned before, E1 was not trained on ARTEMIS as other experts, and did not realize that the AR and VR user interfaces are quite different. In addition, given that E1 was already quite experienced in collaborative VR interfaces, he was expecting similar interfaces as the ones he experienced in the past. A related problem would happen if the expert surgeon points to the position in space where the video library is located in VR, to refer to the currently playing video in AR. The location of the video player on the AR side is not necessarily the same as the location on the VR side, and therefore the pointing gestures would fail. In situations like these, it is important to not take the user interface for granted, and train even the most experienced AR/VR users. In addition, to solve unbalanced interfaces, we could intervene and “warp deixis” [102] by changing pointing gestures so that they align with what users see in their environments.

Despite these drawbacks, we found that ARTEMIS’ different communication features served as a way to overcome technical limitations.

**(4) *Expert and novice surgeons overlap each other*** – ARTEMIS allows experts to (virtually) move around the operating room, but they mostly stayed at same remote location as the novice surgeon. This resonates with works that investigated the impact of giving view independence to remote mentors. Amores et al. [3], for example, touch upon a similar observation where experts commented on the ease of guiding someone by offering a first-person perspective with their own hands. This not only allowed them to benefit from seeing the 3D reconstruction from the same perspective as the novice [59], but it also facilitated enacting the procedure from the novice’s

perspective, something that experts can't do even when working with novices side-by-side.

**(5) Expert surgeons alternate between watching the novice's first-person view and sketching on 3D reconstruction** – Most experts were initially expecting the 3D reconstruction to be as detailed as a video feed or as a real-patient. Unfortunately, as pointed out before, point cloud renders are unable to represent small details such as veins inside cadavers. Nonetheless, by providing more than a single view of the body, ARTEMIS allowed experts to overcome this limitation while still benefiting from the 3D annotation interface. After the initial incision was made, experts would alternate between watching the first-person view of the novice to gather context of what they were doing, then looking at the 3D reconstruction whenever they needed to show a maneuver with their hands or make more markings.

**(6) Verbalization and context can help overcome visual alignment issues in AR** – Similar to other AR systems such as STAR [97, 64], annotations as seen by the novices were often not perfectly aligned with the cadaver, sometimes being off by up to 1-2cm. Interestingly the expert and novice were often able to smoothly account for this — using verbal communication, hand gestures and body landmarks to resolve alignment problems. For instance, in one of the sessions where annotations were not perfectly aligned, E2 directly asked N3 for her feedback: “Does the 3D writing help spatially?” (despite its inaccuracies) [E2], “Yeah, definitely. It's pretty neat actually” [N3]. Another expert-novice dyad [E5-N6] learned that by standing over the body from a similar point-of-view, they could align annotations better. This helped them deal with the lack of depth cues in virtual spaces such as the lack of occlusion of the annotation by the patient body. Despite these difficulties, as mentioned above, users were always able to resolve possible annotation positioning problems, by talking through them and using hands movements to better explain.

**(7) A Digital Whiteboard is needed ... but not really** – In most telementoring systems, experts annotate a video interface to contextualize the instructions they want to convey - for example, by marking the location of an incision. In fact, during our role-playing sessions,

experts annotated the body to show the location of anatomical landmarks (EG2) as well as mark the location, length, and depth of incisions (EG3). Naturally, in addressing these two goals, ARTEMIS' interface is centered around 3D annotations on the patient body. In previous telementoring systems, however, experts could use the video interface to make annotations that are unrelated to the patient body. For example, they might want to explain a technical concept or describe future steps of the procedure [101]. During the exploratory study, *a couple of experts* asked for additional shared writing supports; E1, for instance, mentioned how he “*would like to have a whiteboard so that I can make pauses and explain steps*” [E1]. While this *could inspire a new feature for ARTEMIS*, we noticed how *most experts relied on annotations over the virtual body (in accordance to EG1 and EG3)*. We also noticed that *the lack of a digital whiteboard did not constrain experts in any way*. E3, for example, used the space above the patient to create 3D mid-air handwritten notes that the novices were able to read and act upon.

### **3.6 Conclusion**

In this paper we introduced ARTEMIS, a Mixed-Reality system for immersive surgical telementoring. *Through a participatory design process with expert surgeons, we explored the use of Mixed Reality technology in a collaborative surgical task and defined four design goals that systems should address to better support expert surgeons' needs*. Our system, ARTEMIS, addresses these goals to create a high-fidelity remote collaboration environment for time-critical environments. *Through a qualitative evaluation, we showed that ARTEMIS allows untrained medical personnel to respond to an emergency, and to perform complex surgeries on critical patients under direct guidance from remote experts*. *The qualitative evaluation of ARTEMIS in a real-world surgery scenario outlined a number of important aspects that will be key for the further development of immersive collaborative environments for time-critical applications*.

While telementoring for trauma care was the lead use of ARTEMIS, concepts explored

here can be generalized to remote physical collaboration that requires a high-degree of fidelity and accuracy (down to *the centimeter scale*). *Unlike previous systems that support physical task mentoring through a single, complex interface for both mentors and mentees (e.g., LOKI [108]), ARTEMIS provides specialized interfaces that allow mentors in VR to control what mentees immersed in AR can see so that mentees can focus on the task at hand.*

While much can be achieved in the future with a system like ARTEMIS, we believe that a critical role of this work will be to allow researchers to understand the *unique needs of surgical collaboration in trauma settings as well as* the impact of new mixed-reality technology for collaborative environments, specifically investigating the development of common ground, collaboration, and communication.

In the near future, along with conducting a formal evaluation comparing ARTEMIS with other telementoring systems, we plan on refining this novel technology for mobile immersive AR and VR to enable 3D real-time telementoring in two ways. First, to scale and support time-critical tasks that require co-presence of experts alongside novices. Second, by studying ARTEMIS from the perspective of the novice surgeon. *Currently, a limitation of this work is that it was designed through expert surgeons' perspective. We expect our clinical evaluation to help us uncover unknown mentees' needs that would further improve our understanding of how a system such as ARTEMIS can provide remote guidance to novice surgeons.*

### **3.7 Closing Remarks**

In ARTEMIS [37], we designed a telementoring system where expert surgeons in Virtual Reality see and annotate a 3D reconstruction of a patient. They do so to guide novice medics by showing, for example, the location, length, and depth of an incision. Novice medics, in Augmented Reality, see these virtual annotations on the patient's body and use them as a guide for their procedure (Fig. 3.7).

The alignment requirement between instructions and patients is more critical in ARTEMIS than in HoloCPR. Here, novice surgeons follow virtual lines (Fig. 3.7 (B) ) to, for example, make an incision. Theoretically, if these lines are off by a couple of centimeters, they can lead novices to cut the patient at the wrong location. However, this is not what we observed.

Despite our best technical efforts, the AR interface was often off by one centimeter. Nonetheless, users collaborating through the system could still follow AR annotations correctly, as mentors would often draw more lines to contextualize their annotations or verbalize instructions to guide novices around the misaligned interface.

After carefully tackling all possible tracking problems, what could possibly cause a registration error of that much? In the next chapter, we delve into the innerworkings of head-mounted displays to better understand other factors that can cause registration errors even when perfect tracking is available.

Chapter 3, in full, is a reprint the material as it appears in the the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. I was the lead author of this material. Danilo Gasques, Janet G. Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Konrad Davis, Michael Yip, and Nadir Weibel. The dissertation author was the primary investigator and author of this material.

## Chapter 4

# Measuring the limits of perception-based guidance

Users following physical task guidance in Augmented Reality (AR) align physical objects with virtual objects (and vice-versa). For example, in an assembly task, workers follow 3D virtual models to place physical objects in the correct location and orientation. While placement accuracy depends on users' dexterity, it is bound by where the display shows the virtual object to the user. In stereoscopic displays such as optical see-through head-mounted displays (OST-HMD), users can discern virtual objects presented at different depths; however, when comparing virtual and real objects, depth perception is warped by how well calibrated the OST-HMD is for that user. Because OST-HMDs display images on small screens between the user and the real world, they must accurately calibrate to the user's eye; otherwise, users will see virtual objects at the wrong location. This paper measures how eye-display calibration errors affect users' ability to align virtual and physical objects. We conducted a user study (n=12) where users either align a physical object with a virtual one or a virtual object with a physical one using three different displays: a perfect AR system simulated in Virtual Reality (VR), a simulated AR with optical error, and a modern AR headset (HoloLens2). Our results show user performance is directly related to the

display used. Eye-display calibration errors, when present, negatively impact user performance, effort, and frustration.

## 4.1 Introduction

Augmented Reality displays create the illusion that virtual objects co-exist with physical objects in the real world. This affordance of AR displays is commonly used to show situated information near an area of interest so that users more easily follow instructions. In domains where users follow instructions describing how to move and orient objects, situated Augmented Reality guidance has been shown to improve precision [53] and reduce errors [97], all while requiring less cognitive effort from users [20]. These benefits of AR displays for manual task guidance become even more critical in high-stakes domains where a mistake can lead to unbearable personal or financial losses, such as trauma medicine and emergency procedures.

In manual task guidance, alignment accuracy is a crucial metric as it measures how far virtual and physical objects are from one another. For years, researchers have worked on improving alignment accuracy through a myriad of approaches, including user interaction [106], visual design [32], alignment interface [4], and comparing different AR displays. The current consensus is that wearable head-mounted displays (HMD) are the most appropriate format for manual task guidance [94] as they render virtual graphics from an egocentric perspective while allowing users to operate hands-free.

Presenting virtual information from the user's perspective is a complex problem impacting alignment accuracy. To date, there are two distinctive implementations of HMDs [11, 16]: video see-through (VST) and optical see-through (OST). VST-HMDs use one or more cameras to combine virtual objects with a live video stream of the physical world, constraining users from seeing from the camera perspective. OST-HMDs, on the other hand, use semi-transparent displays between the user's eyes and the physical world. One of their most significant advantages

is that the real world is seen unaltered, in full resolution, and from the user's perspective. Modern commercial HMDs such as Microsoft HoloLens 2, Magic Leap, and Snapchat's Spectacle 3 - are optical see-through HMDs.

Despite their advantages and widespread use, OST-HMDs have a critical limitation: perfect registration - or alignment - between virtual and real objects is unattainable with current technology [45]. Despite improvements to display technology, the underlying working mechanism is the same: virtual objects are rendered and presented to the user as 2D images on a screen positioned at a fixed distance from their eyes.

Recent works have investigated these perceptual issues' impact on user performance and task workload, with promising results, as no difference was found when compared against a baseline task. Nonetheless, no work has investigated the impact of specific calibration parameters on users and user performance. This work investigates the impact of eye-display calibration errors on user perception and performance. After all, manual alignment accuracy is bound to how well the augmented reality display can render virtual information at the right location as perceived by the end user.

Unfortunately, there is a unique difficulty in experimenting with user perception in AR as we cannot confirm what users see [45] nor control essential parameters of eye-display calibration, such as where the system perceives the eye location. We create an optically-accurate simulation of Augmented Reality in Virtual Reality to overcome these difficulties. A significant advantage of using virtual reality to simulate Augmented Reality is that both the virtual and real worlds are simulated, and the simulation knows the exact location of the user's eyes.

Using this controlled environment, we measured the impact of eye-display calibration on alignment accuracy and user task load (NASA TLX [46]). In this experiment, twelve (n=12) users either align a physical object with a virtual one or a virtual object with a physical one using three different displays: a perfect AR system simulated in Virtual Reality (VR calibrated), a simulated AR with a standard eye-display calibration error (VR 1mm error), and a modern AR headset

(HoloLens 2). Our results show alignment accuracy is directly related to the display used. A small (1 mm) eye-display calibration error, when present, negatively impacts the user’s perception of performance, effort, and frustration.

We further investigate how users cope with optical errors through interviews and observations. Despite struggling with alignment tasks in the presence of optical errors, users cannot always detect or understand what is happening. Optical errors are perceived as a “swing and sway” error, a change of brightness, and a mismatch of object sizes. This latter point directly affects how we design interfaces for manual alignment - especially in the presence of optical calibration errors.

In this paper, we make the following contributions:

1. A Virtual Reality OST-HMD simulator
2. A user study comparing user alignment performance and task workload on both an ideal headset and a headset with a 1mm optical error.
3. An in-depth comparison between the results in a simulated environment to those of a modern OST-HMD (HoloLens 2)

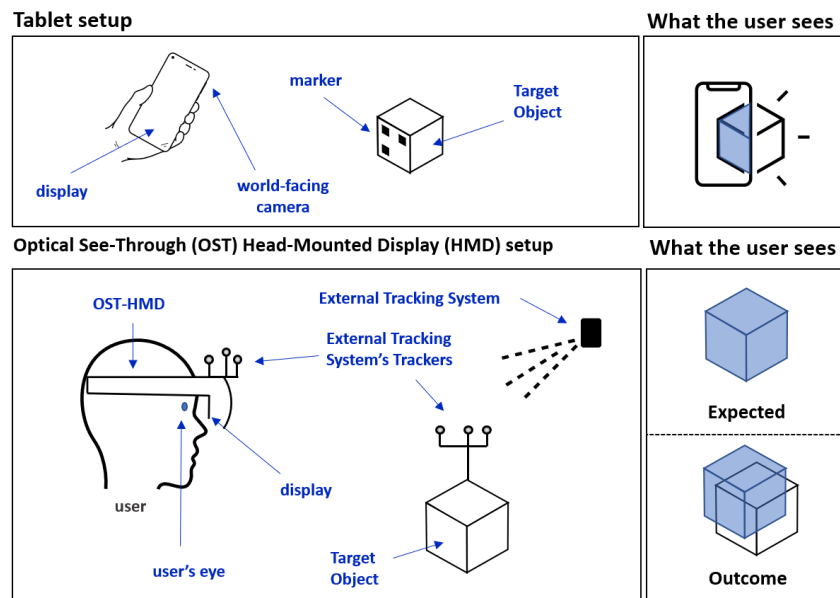
In short, this paper answers the following research question: **How do perceptual differences impact one’s ability to align a physical with a virtual object?** In the next section, we briefly review OST-HMDs, covering concepts from virtual image registration, eye-display calibration, and depth perception. Then, we situate our work in the current research landscape of alignment interfaces. We then describe our experiment, and its results and discuss them.

## 4.2 Background and Related Work

A key feature of Augmented Reality displays is that they “register real and virtual objects with each other” [10]. To the user, registration creates the illusion that virtual objects exist in

the physical world. To create this illusion, the AR display has to “see” the world from the same perspective as the user. More specifically, the AR system needs to *track* the pose of a real-world *target object* – its 3D position ( $x, y, z$ ) and orientation (*yaw, pitch, roll*) – with respect to the display. Once the system can determine the target object’s pose, it can then determine the position of a *virtual camera* able to *render* [33] 2D pixels on the physical display so that users see a virtual 3D object aligned with the physical one.

The complexity of *registration* lies in the many errors that can be introduced during *tracking*, *rendering* as well as by the *virtual camera* being used. In a VST-HMD setup, the user sees the real world mediated by a video stream from a physical, world-facing camera. The physical camera “sees” the world for the user, and the AR system uses its video stream to combine virtual and real. Thus, as long as the *virtual camera* is analogous to the physical one, virtual and physical objects will be aligned (Fig. 4.1, top).



**Figure 4.1:** Two common Augmented Reality setups. The top image shows a tablet AR example and the bottom image shows an OST-HMD example. (Top) In the tablet setup, users see the augmented world through a world-facing camera. (Bottom) in the OST-HMD setup, the *user* wearing an OST-HMD sees the *target object* through transparent displays. Because of a small error on display-eye calibration, the virtual cube (shaded) can appear poorly aligned with the real one.

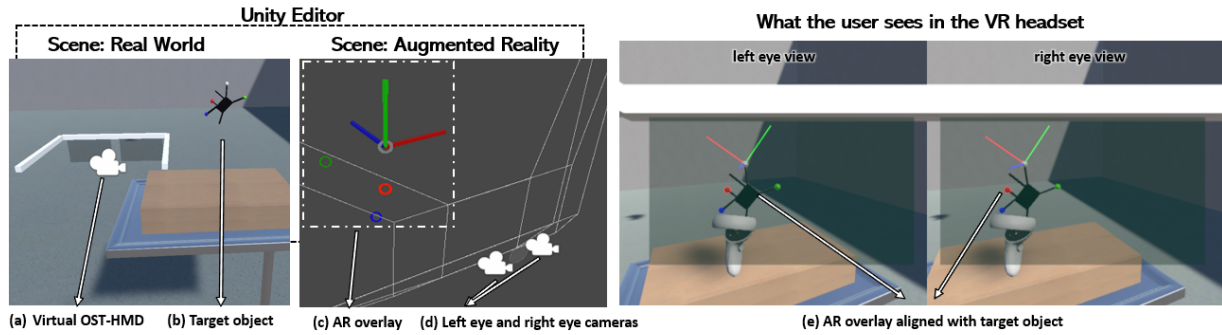
The user sees the real world directly in a more complex setup, such as OST-HMDs. The AR system then displays virtual objects on a transparent screen placed between the real world and the user. Thus, to render virtual objects aligned with the real world, the AR system must rely on accurate tracking technology and use *virtual cameras* that are analogous to a more complex real-world camera: the user's eyes.

To find the transformation between the display and the user's eyes, a calibration procedure between the user and the headset is always necessary [45]. This calibration estimates the intrinsic and extrinsic parameters of the virtual rendering camera representing the eyes. These parameters account for the eye position with respect to the OST display as well as the projection properties of the eye-display pinhole model [51, 110, 8].

Another advantage of modern OST-HMDs is the use of two displays, one per eye to rely almost on stereopsis [61] to create the perceptive illusion of depth. Stereopsis occurs due to the slightly different viewpoints each eye captures due to their horizontal separation on the head. The human brain calculates the depth of an object given its positional disparity between each eye. Thus, even though the optical display is at a fixed focal plane, it can interpret different virtual objects at different depths. This depth perception allows us to judge distances accurately, navigate our environment, and perform tasks that require hand-eye coordination. Stereopsis is essential for catching a ball, driving, and avoiding obstacles in our path.

#### **4.2.1 Simulating Augmented Reality in Virtual Reality**

Evaluating alignment workflows in any OST-HMD is particularly challenging because only the headset user can access it. A common approach is to employ a confirmation study that helps validate user perception of alignment [45]. For example, showing minor virtual points on a tracked surface and measuring how accurately users can touch them [71]. However, these validation studies are often constrained and introduce biases and errors in the evaluation [66]. Seeing exactly what the user sees is an impossible task, and even approximations, such as



**Figure 4.2:** How it works: to simulate AR in VR, our simulator uses two scenes. Like any VR application, the “Real World” is a VR scene. In it, the simulator adds a virtual OST-HMD with two displays. The Augmented Reality scene is a separate scene containing only the AR part of the simulation. The OST-HMD *pose* is sent to the Augmented Reality scene for every frame rendered. Based on the user’s inter-pupillary distance, two cameras - one per eye - follow the pose of the virtual OST-HMD to render one AR view per display. The poses of other objects can also be sent and used in the Augmented Reality scene. **Left:** Simulated real world scene – (a) Virtual OST-HMD with two displays. (b) Target object to be augmented in AR. **Middle:** Augmented Reality scene – This scene is designed to be completely isolated from the real-world environment and can be re-used with different environments. (c) Example of an AR visualization to be aligned with the target object in a real-world scene. (d) left and eye cameras responsible for rendering the AR overlay scene. **Right:** Virtual Reality user view showing a preview of both eyes – (e) rendering on both displays.

replacing the human eye with a physical camera [80], do not make alignment assessments objective because of mismatches between camera and eye intrinsics [43, 9].

With so many possible sources of error, such as variable eye-display calibration errors, and inside-out tracking drift, it is impossible to measure alignment errors that isolate one specific controlled error objectively.

### 4.3 The Virtual-Augmented Reality simulator

In the past, researchers have successfully used Virtual Reality to simulate Augmented Reality for studies that measure different forms of calibration [69, 83, 30, 95]. For example, Martin et al. [70] studied aligning a 3D object over another 3D object in Virtual Reality. Simulations in Virtual Reality are an overly optimistic simplification of a real head-mounted display. Moreover,

the alignment error is a combination of errors arising from different sources; many are ignored in a Virtual Reality study.

That said, conducting alignment or calibration studies using an actual Optical See-Through Head-Mounted Display (OST-HMD) presents substantial challenges. Implementing and experimenting with various calibration methods requires comprehensive environmental and user instrumentation. Still, even when these are available, we lack tools that can help us understand trade-offs between different calibration algorithms. Moreover, the alignment error is a combination of errors arising from other sources, so it is hard to pinpoint what calibration or aspect should be improved next: optical distortions from the display technology. Poor user-display setup (e.g., wrong IPD, poorly worn head-mounted display)? Drift or occlusion introduced by tracking system? Numerical instability from mathematical models?

To address the current complexity of running more realistic alignment studies while providing the constraints required for a type study, we introduce an AR simulator in Virtual Reality. Our simulator provides an interface to implement and test calibration setups without needing to deploy to the real AR HMD or to instrument the environment with expensive tracking technology. Moreover, unlike in real life, we can always obtain ground-truth measurements in the simulator.

Our VR OST-HMD simulator creates a virtual, augmented reality headset in a virtual environment. Similar to real life, the head-mounted display in this virtual environment is a pair of goggles that users can wear to experience AR. The virtual OST-HMD renders a different picture per eye so the user can see and interact with them in three dimensions (Fig. 4.2).

One of the main benefits of simulating both real and augmented environments is that we have control over both environments and how and when information traverses from one to the other. For example, the simulator guarantees that all tracking data from the simulated real world is used by the virtual OST-HMD simultaneously within the same frame.

### 4.3.1 Simulator's Features

Many factors influence alignment. To simulate a variety of alignment workflows, our virtual AR simulator provides researchers and developers with the following features:

(1) **Virtual Trackers** - A core building block of our system is a virtual tracking system that broadcasts the *pose* of objects of interest in the real simulated scene to the augmented reality scene. This includes the OST-HMD position and the location of its displays. Trackers can either share a global coordinate system center or belong to a shared coordinate system centered at a specific tracker. *Virtual trackers* can be customized in many ways: tracker drift (e.g., constant, per acceleration, per time-step, based on a curve), jitter (e.g., Gaussian), and latency (e.g., constant, distance-based, Gaussian). Moreover, they can also be customized through user-defined callbacks. Finally, *virtual trackers* can be referenced by the AR scene by name.

(2) **Locatable Cameras** - Locatable cameras are cameras that, for each frame captured, attach a transformation matrix that maps points from the camera coordinate system to the head-mounted display coordinate system. They are commonly available in modern HMDs such as Microsoft HoloLens. Researchers have used this camera, for example, to track and augment target objects visible to the user. We provide a locatable camera with our virtual OST-HMD that developers can use to test and implement computer-vision tracking methods. They can also control the camera's resolution, field-of-view, and its location on the OST-HMD.

(3) **Linear Equation Solver** - Generally, the goal of calibration algorithms is to find a perspective, affine or orthogonal transformation between two coordinate systems. This transformation is organized as a system of linear equations and solved through point correspondences between the coordinates systems [45]. Our simulator includes a basic generic solver that can produce perspective, affine or orthogonal transformations given a list of point correspondences.

(4) **Customizable OST-HMD** - All aspects of the OST-HMD are customizable: display position, rotation, size, resolution, and its projection matrix. This customization flexibility allows us to

create simulations for modern OST-HMDs (e.g., HoloLens and Magic Leap) and experiment with new displays. Through post-processing shaders, we can also simulate color bleeding, optical distortion, etc.

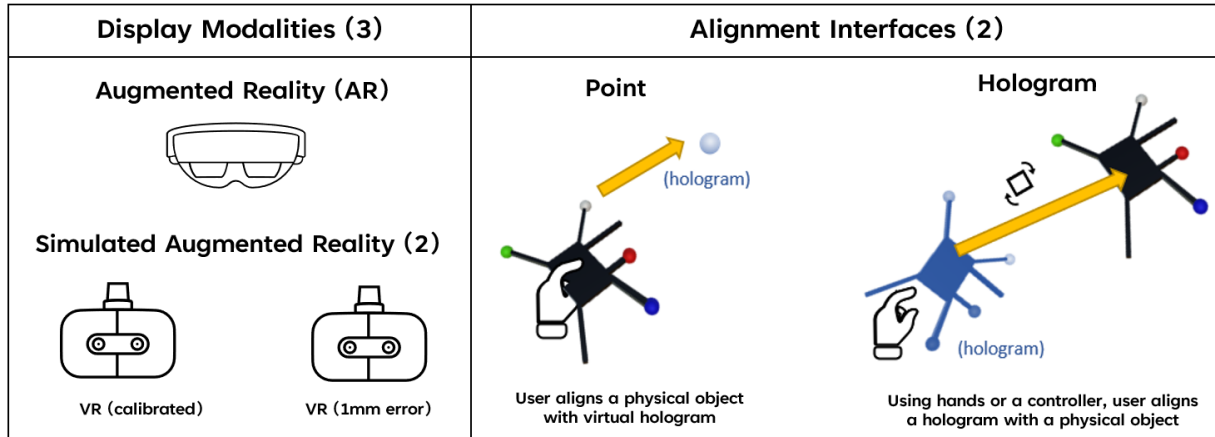
(5) *Template scenarios* - To help other researchers get started with our simulator, we provide three templates scenarios: (a) **3D-3D scenario** in which a user maps points from the OST-HMD 3D coordinate system to another 3D coordinate system (e.g., a tracked object [92]); (b) **locatable camera scenario** a template that shows how to use the virtual camera to track fiducial markers such as ArUco Markers [81]); (c) **3D-2D calibration**: a variation of the Single Point Active Alignment Method (SPAAM) [110].

## 4.4 Materials and methods

To better understand the impact that displays technologies have on users' ability to align virtual and physical objects, our study compares users' ability to align objects using three different display technologies: two simulated AR headsets and a modern AR headset (HoloLens 2). Both simulated AR headsets use an optical see-through simulation in Virtual Reality. The modern AR headset serves as a way to compare and contrasts findings in a controlled, simulated environment with findings in less controlled device.

The two simulated AR headsets represent two different OST-HMDs. One is a perfectly calibrated OST-HMD, and the other has a static eye-display calibration error. This calibration error creates visual effects that misrepresent the location of virtual objects in the physical world. To better understand their impact on user performance and task workload, we focus on a common type of eye-display calibration error: the focal point (i.e., eye) location. Hence, one of the simulated AR headsets offsets the eye location by 1 mm at a random axis (X,Y,Z).

User performance on manual tasks can be directly related to the task. In this study, we compare the two most common types of alignment interfaces in use: aligning a physical object



**Figure 4.3:** Study design: a 2x3 factorial study design with two independent variables: **display modality** and **alignment interface**. The display modality variable has three levels: *Augmented Reality (AR)*, *Virtual Reality (calibrated)*, or *Virtual Reality (1mm error)*. Both Virtual Reality modalities simulate an optical see-through head-mounted display. VR (1mm error) simulates a display with 1 mm of error on the focal point location. The alignment interface can assume two levels: *Point* or *Hologram*. We explain the rationale for these interfaces in the section *Alignment Interfaces*.

with a virtual object and aligning a virtual object with a physical object. These two types of alignment interfaces are generalizations of the two most common ways users follow guidance through Augmented Reality. For example, a surgeon sees the location of a cancer tumor as a virtual sphere to manually align a physical medical instrument to it. In a different setup, the same surgeon manually aligns a virtual MRI image over the patient to plan for the procedure.

In the rest of this section, we delve deeper on our study design, the alignment interfaces, the task performed by the participants, the metrics collected, the number of participants, and our hypotheses. We present results and discuss results in the next two sections.

### 4.4.1 Study Design

Our study follows a 2x3 factorial design with two independent variables: **display modality** and **alignment interface** (see summary in Fig. 4.3). The display modality variable has three levels: AUGMENTED REALITY (AR), a perfectly simulated Augmented Reality in Virtual Reality:

VR (CALIBRATED), and Augmented Reality with a static 1mm focal point error: VR (1MM ERROR). The two VR modalities simulate an optical see-through Augmented Reality display. The alignment interface can assume two levels: POINT and HOLOGRAM. We explain the rationale for these interfaces in the next section.

We chose a within-subjects design for both variables because it gives us an opportunity to observe and interview each participant to learn how their experience in the simulated AR interfaces compares with the experience using and interacting with a modern optical see-through head-mounted display. Previous studies with similar alignment interfaces have shown that participants perform faster in a second trial than in a first trial, but not more accurately [106]; hence, we don't expect any order effect. Nonetheless, to avoid possible order effects, we counter-balance both variables. Moreover, we required participants to repeat each calibration interface 4 times in a randomized order.

#### 4.4.2 Alignment interfaces

When following guidance from an Augmented Reality interface, the end goal of aligning a physical object with a virtual object is to follow a plan as accurately as possible. Aligning a virtual object with a physical object also creates a common frame of reference between the virtual environment and the real world. This frame of reference is essential in several tasks, such as situating surgical plans onto a patient.

This study focuses on two standard interfaces for aligning virtual and real objects. These interfaces are well-studied, and we can rely on existing research to explain our results. The two interfaces are dubbed **Point** and **Hologram** (Fig. 4.3). The first modality requires users to place a physical, hand-held object into a specified mid-air position (Fig. 4.3 Point). The second modality - Hologram - requires the opposite: users dock a virtual target over a physical target (Fig. 4.3 Hologram). More specifically:

- **Point:** Users see a small sphere in their headsets in this interface. Using their dominant

hand, they move a physical, tracked sphere to align with the virtual one. We require a total of 25 alignments.

- **Hologram:** Instead of moving a physical target, users align a virtual target by moving it over a stationary physical target. They either use a controller (in VR) or hand gestures (in AR) to move the target. This interface requires users to perform 8 alignments, thereby collecting 36 points. This interface is also known as 6 degrees-of-freedom (DoF) docking [4], and it is one of the most commonly employed manual alignment interfaces today [32, 34, 4].

### 4.4.3 Metrics and Participant Feedback

We are measuring the following metrics to understand better how each display interface influences one's ability to align virtual and physical objects.

- **Alignment error.** We measure and collect the cartesian distance between virtual and physical objects. This is a key metric as it is correlated to alignment accuracy.
- **Alignment time.** There is a trade-off between accuracy and time, with users who align fast not necessarily being as accurate as users that take longer to align. Also, longer alignment times might be correlated to a more complex procedure. We measure the time participants take per alignment to discard participants behaving hastily and detect possible discrepancies between the time taken across display and interface modalities.
- **Task Load Index.** A general guideline for alignment interfaces is that they should not burden users [4, 6]. To better understand the effect that the display modality has on alignment interfaces, we measure the participant's perceived workload, frustration, and mental and physical demand by asking participants to fill out a standard (NASA TLX [46]) questionnaire after their last time completing an alignment task for each alignment interface in each display modality.

A crucial part of this study is the qualitative component of the data collected because it helps us explain differences observed on the metrics collected above. To that extent, we asked users to speak out loud while they were performing alignments (but before confirming an alignment). At the end of each study, we also conducted a semi-structured interview, asking participants to elaborate on their notes during the study as well as asking them to describe differences and similarities between the displays and the alignment interface. We clustered these statements to answer the following question:

- **Perceptual differences.** Do users notice differences between displays? If so, what are those? How do they perceive and explain the optical error present?

#### 4.4.4 Task

Participants, either wearing an AR HMD or a VR HMD, performed a total of 24 *alignment tasks*. Each alignment task consisted of repeatedly aligning physical and virtual objects. Each alignment can be either physical-virtual or virtual-physical. The first requires participants to move the physical object to meet a virtual, mid-air target. The latter requires participants to move a virtual object (“hologram”) to meet a physical object.

The 24 alignment tasks are split across 3 *display modalities* for 8 alignment tasks for each display. For each display modality, the 8 calibrations are further divided into 4 alignments per alignment interface (POINT and HOLOGRAM) in random order. Each alignment interface requires different alignment operations, with POINT requiring 25 alignment operations and HOLOGRAM requiring only 8 alignment operations.

To confirm an alignment, participants wearing a VR headset pressed a button on any VR controller. On the other hand, participants wearing an AR headset pressed a key on a small portable keypad. After pressing the button to confirm an alignment, participants hear audio feedback, and then either a new virtual target was presented (POINT), or the physical target was moved to a new location (HOLOGRAM) by a study facilitator. The location of the targets was

randomized in front of the user but always within arm's reach. Finally, after completing all the required alignments for a task, the interface would average them to present participants with a visual summary of how well they did.

Participants were asked to perform alignments as accurately as possible. However, they were also told that the entire task was timed and that they shouldn't spend more than a few minutes on a whole alignment task.

#### **4.4.5 Participants**

We recruited 12 unpaid participants from our university (7 male, 5 female with an average age of  $30.42 \pm 5.42$ ). All participants had some prior experience wearing Augmented Reality and Virtual Reality headsets. Three participants had performed a similar alignment task to the Hologram interface in Augmented Reality as they were used to moving virtual holograms on top of physical objects. Two participants were short-sighted but used corrective lenses during the study.

#### **4.4.6 Experimental Procedure**

First, we explained the study goal and design to each participant. They were also told they could drop out of the study anytime for any reason. They were also allowed to rest between calibrations if needed. Participants then signed a consent approved by our University's Independent Review Board.

Then, we assigned participants to a *display modality*. As a counter-balanced within-subject study design, a third of the participants were first assigned to AR, the other third to VR (1mm error), and the final third to VR (calibrated). For each display modality, participants performed a total of 8 alignment tasks. To avoid order effects, each participant used each alignment interface (2) four times (4) in a counter-balanced order.

After completing each alignment task, if this were the last trial of an *alignment interface* for a *display modality*, participants would then remove their headset to fill a Nasa TLX survey form [46]. Each participant filled a total of 6 NASA TLX forms: one per alignment interface per display modality.

After completing all 8 alignment tasks for a display, participants were interviewed for 5-10 minutes to describe their experience. Then, they were given another headset, and, once again, they had to perform 8 alignment tasks, 4 per alignment interface as described above.

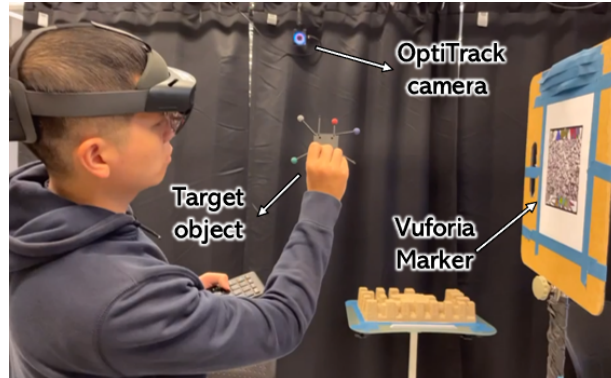
After completing all 24 alignments, 8 per display, we then interviewed participants to understand how each interface compared between display modalities. The experiment time varied as participants decided to spend more or less time aligning an object. The study took an average time of 2 hours with a max of 3h.

#### **4.4.7 Implementation**

We use an optically accurate simulation of OST-HMD in Virtual Reality based on an open-source OST-HMD simulator [39]. The experiment was implemented using Unity 2019 for each display modality.

For the VR modality, we created two scenes. The first scene (the “Real World” environment) had all assets required for a regular OpenVR/OpenXR application. We also added walls, text instructions, an External Tracking system, and a virtual target under it (Fig. ??). We used Unity assets from Virtual-Augmented Reality [39] to create a simulated OST-HMD in the VR scene. The second scene was the simulated Augmented Reality. All the logic for the user study and the calibration procedure was implemented on the AR scene. We deploy the VR modality of the study to an Oculus Quest 2.

For the AR part of the study (Fig. 4.4), most of the software was copied from the VR application’s Augmented Reality scene. Moreover, we had a physical setup with a motion capture tracking system. Real-time tracking data of the physical target object was collected



**Figure 4.4:** User study in AR using OptiTrack Motion Capture cameras for tracking the Target Object. One Vuforia marker with OptiTrack markers was used to create a common coordinate system for the OST-HMD (HoloLens 2) and OptiTrack.

from OptiTrack’s Motive software, encoded in a binary format, and sent over UDP to the AR application through a custom networking protocol. We created a custom Vuforia marker with OptiTrack Markers to obtain ground-truth measurements. For each trial of the user study, users would look at the marker to calibrate OptiTrack with the OST-HMD.

## 4.5 Results

This section lists this study’s results and their respective statistical analysis.

### 4.5.1 Quantitative Metrics

Because the POINT interface and the HOLOGRAM interface collect a different number of data points (25 and 32 respectively) and require a different number of alignment operations (25 and 8 respectively), analysis of alignment error and alignment time was standardized using the following procedure. For alignment error, we randomly sampled 24 points from the 25 and 32 points collected. For alignment time, we calculate and compare the mean alignment time per alignment operation.

Moreover, for each metric, we tested for normality using the Shapiro-Wilk test, and where normality was violated, we used a Box-Cox transform to normalize the data. With the normalized

data, we tested for within-subjects effects with a two-way repeated measures ANOVA, where the two factors were *Display Modality* (3 levels) and *Alignment Interfaces* (2 levels). For each factor, we also performed post hoc tests. We report the results for each metric in the following three sub-sections.

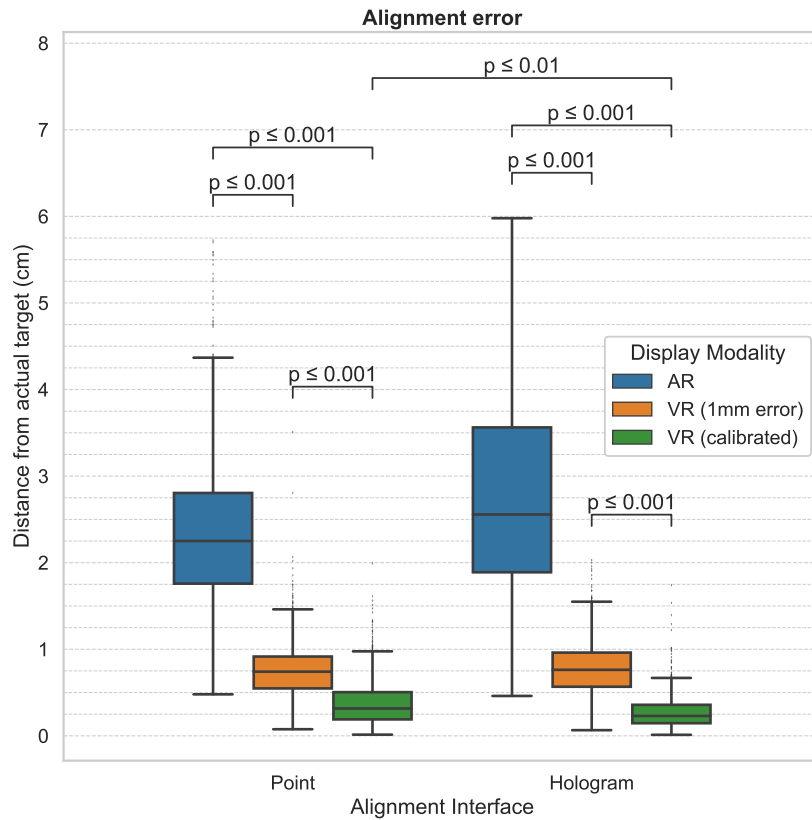
## 4.5.2 Alignment error

A total of 96 alignment points ( $24 \times 4$  repetitions) was collected per participant ( $n = 12$ ) for each *Display Modality* (3)  $\times$  *Alignment Interface* (2). Figure 4.5 summarizes the results in a box-plot chart. We report all alignment errors in centimeters.

For the POINT *alignment interface*, the *display modality* AR was statistically significantly worse ( $p < 0.001$ ) than all other modalities with an average error of 2.339 cm (min = 0.479, max = 5.723,  $\sigma = 0.861$ ). Moreover, VR (1MM ERROR) was statistically significantly worse ( $p < 0.001$ ) than VR (CALIBRATED) with an average error of 0.751 cm (min = 0.076, max = 3.507,  $\sigma = 0.311$ ). VR (CALIBRATED) yielded the lowest error with an average of 0.383 cm (min = 0.013, max = 1.991,  $\sigma = 0.260$ ).

For the HOLOGRAM *alignment interface*, the *display modality* AR was statistically significantly worse ( $p < 0.001$ ) than all other modalities with an average error of 2.774 cm (min = 0.460, max = 5.978,  $\sigma = 1.889$ ). Moreover, VR (1MM ERROR) was statistically significantly worse ( $p < 0.001$ ) than VR (CALIBRATED) with an average error of 0.777 cm (min = 0.065, max = 2.02,  $\sigma = 0.332$ ). VR (CALIBRATED) yielded the lowest error with an average of 0.272 cm (min = 0.011, max = 1.740,  $\sigma = 0.180$ ).

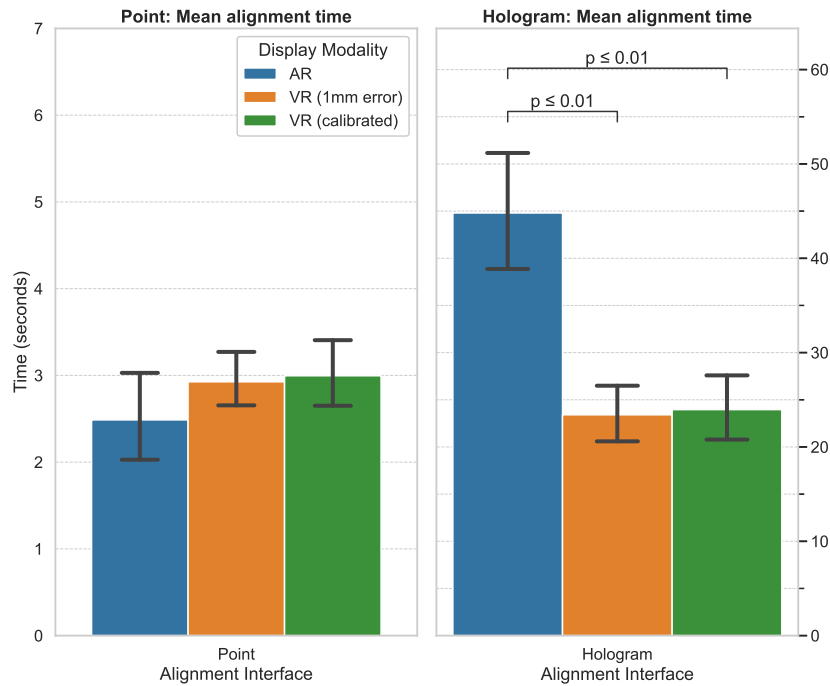
VR (CALIBRATED) was the only *display modality* with a statistically significant interaction ( $p < 0.01$ ) across *alignment interfaces*, with HOLOGRAM being 0.111 cm more accurate than POINT on average.



**Figure 4.5:** Boxplot of alignment error, or the Euclidean distance between the virtual and physical objects in centimeters.

### Alignment time

Overall, across *Alignment Interfaces*, POINT alignment was roughly ten times faster than each HOLOGRAM alignment (Fig. 4.6). Each POINT alignment took in average 2.803 seconds (min = 1.15, max = 9.85,  $\sigma$  = 1.47) and each HOLOGRAM alignment took 30.721 seconds (min = 5.67, max = 121.145,  $\sigma$  = 18.75). For the POINT interface, no statistically significant differences were found across display modalities. For the HOLOGRAM interface, alignment times for AR were statistically significantly higher ( $p < 0.001$ ) when compared to both VR (1MM ERROR) and VR (CALIBRATED).



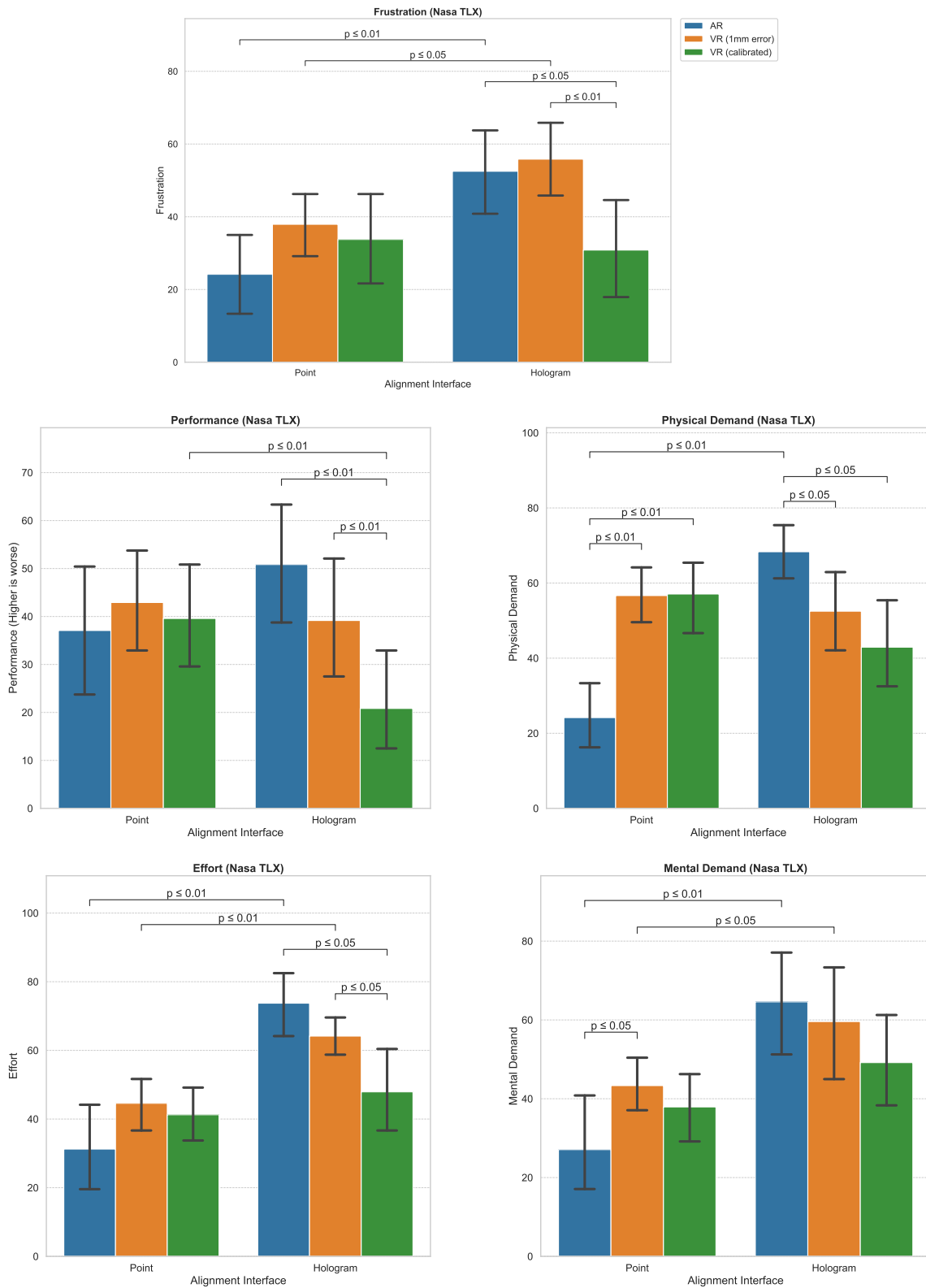
**Figure 4.6:** Barplots of the mean alignment time per *Display modality*. Left: POINT. Right: HOLOGRAM

### Task Load Index (NASA TLX)

The NASA task load index measures the subject mental workload of a task under six dimensions: Mental demand, Physical demand, Temporal demand, Effort, Performance, and Frustration. We don't include Temporal demand because it lacks significant differences across variables. For a visual summary of these results, check the bar plots in Figure 4.7.

### 4.5.3 Perceptual differences

For this qualitative metric, we used affinity diagramming to combine statements participants said out loud during the experiment or the semi-structured interview. The overall goal of this analysis was to describe what differences participants notice across displays, if any. There are three main reports:



**Figure 4.7:** The five relevant dimensions from NASA TLX for this study. Frustration, Performance, Physical Demand, Effort, and Mental Demands. Temporal demand didn't present any statistically significant differences across variables.

### **Apparent virtual object drift or sway**

The false perception that the virtual object was moving around was the most common perceptual issue described by participants on both the AR (6 out of 12) and VR (1MM ERROR) (11 out of 12); however, only for the HOLOGRAM interface because it was the only *alignment interface* that allowed users to confirm alignment between virtual and physical objects by walking around the two.

Users described this issue in two ways: (1) Hologram moving or drifting: “It feels that when I turn around, the hologram is moving” [P1, P3, P5, P9]. (2) Viewpoint-dependent error: “I look from one side, and it is in front of the virtual one. I look from the other side; it looks like it is behind” [P2, P6, P7, P8, P10, P11, P12]. The movement was so apparent that one participant wasn’t sure if it was part of the experiment: “is the goal to do on multiple sides or just this one?” [P10].

Only one participant (P4) who didn’t notice any apparent drift or sway commented that VR (CALIBRATED) “feels easier, but I am not sure why” [P4]. Along with P4, another 4 participants (5 out of 12) commented on how virtual objects under VR (CALIBRATED) was more stable: “(when I am) moving around, it doesn’t move with me.” [P6]

Finally, this apparent virtual object drift caused some participants to put more effort into aligning objects for both AR and VR (1MM ERROR), with many participants expressing frustration in the process (“I could do this indefinitely”[P6], “work is frustrating but because I can’t do what I want” [P5]). Whereas for VR (CALIBRATED), some participants (5 out 12) described it as easier, and that “it incentivizes me to put more effort because I know it will stay better” [P3].

### **Skewed object perception in AR**

Two participants [P1, P7] complained that virtual objects seen through the HoloLens weren’t the same size as their physical counterparts.

Two participants [P2, P6] mentioned that, depending on how they focused, they could see in “double” on the HoloLens.

### **Interaction interfaces: differences between VR and AR**

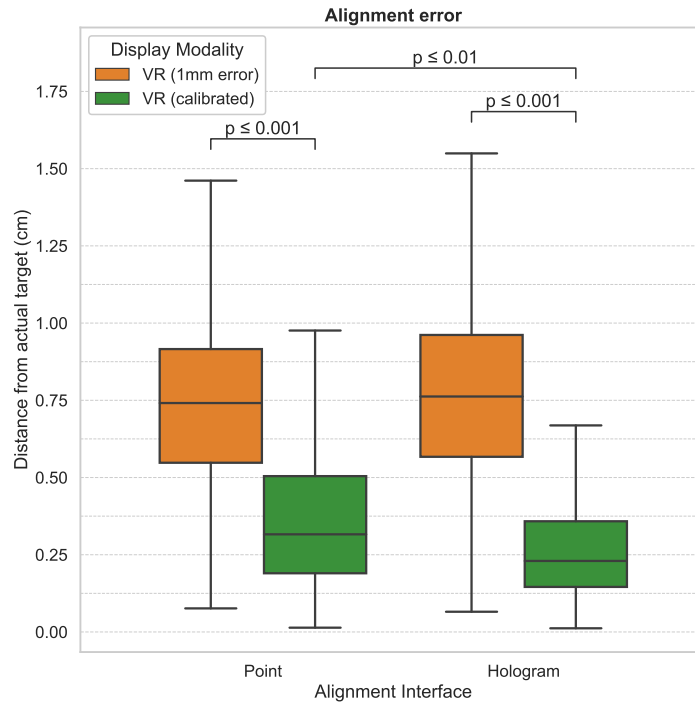
While not a goal of this study, many participants preferred the VR controller during the HOLOGRAM modality and the plastic physical object during the POINT modality. The next section discusses these preferences and their possible influence on the study results.

## **4.6 Discussion**

In this work, we used a simulated OST-HMD in VR to measure the influence of eye-display calibration on two types of alignment tasks. This section discusses these findings in more detail according to the current literature. For each subsection, we first discuss results between VR (CALIBRATED) and VR (1MM ERROR). Then, we discuss how these results in a simulated environment compare with the real-world deployment of the same experiment on a modern OST-HMD: AR.

### **4.6.1 Eye-display calibration errors highly impact manual alignment accuracy**

From the results presented in the previous section, it is clear that a small eye-display calibration error (VR (1MM ERROR)) statistically impacts alignment performance on both types of alignment interfaces studied (POINTS and HOLOGRAM) when compared to an ideal OST-HMD (VR (CALIBRATED)) (Figure 4.8). Under the experimental condition with calibration errors, participants were, on average, 0.398 *cm* and 0.505 *cm* worse than the ideal condition (VR (CALIBRATED)) for POINT and HOLOGRAM. This is an almost two-fold increase in the average error for each interaction interface.

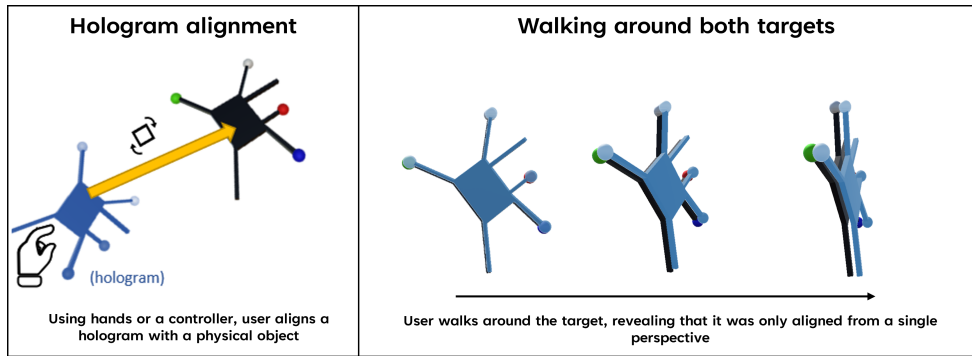


**Figure 4.8:** Boxplot of alignment error focusing on VR (1MM ERROR) and VR (CALIBRATED).

## 4.6.2 Eye-display calibration errors nullify alignment interface improvements

The two alignment interfaces in our study - POINT and HOLOGRAM - were expected to yield different alignment errors, with HOLOGRAM being more precise for two reasons. First, POINT requires participants to move a physical object mid-air to align it with a virtual object. The lack of physical support would naturally introduce inaccuracies to the manual alignment task. Second, the HOLOGRAM interface mitigated this inaccuracy and allowed participants to refine their alignment after an initial, crude virtual-physical alignment.

For the ideal display VR (CALIBRATED), the mean alignment error for HOLOGRAM was statistically significantly lower than POINT, improving alignment accuracy by 28.98% (0.111cm, see Section 4.5.2). Surprisingly, there were no statistically significant differences between POINT and HOLOGRAM under VR (1MM ERROR) (Fig. 4.8). While not statistically significant, the mean



**Figure 4.9:** In eye-display calibration errors such as those studied in this work, objects aligned to physical objects from one point of view are consistently not aligned from a different point of view. In this figure, we show an example of what eye-display calibration looked like for users manually aligning a virtual target (Hologram) to a physical target they could see while walking around the two.

alignment error for HOLOGRAM is slightly higher than POINT. We attribute that to participants being able to recognize that their display was “misbehaving” while being unable to take any corrective actions. As P10 put it, *“When I was choosing where to place the object, I was more comfortable with placing it imperfectly on one side because it wouldn’t be so bad on the other side.”*

### 4.6.3 Eye-display calibration errors are eventually perceived as virtual object drift

In the absence of tracking, optical, or other calibration errors, eye-display calibration errors can be visually recognized because virtual objects aligned to physical objects from one point of view are consistently not aligned from a different point of view (See Figure. 4.9). This, however, is not immediately clear because it requires users to walk around virtual and physical objects a couple of times before they notice that the virtual and physical objects only appear to be drifting apart; however, they always meet back and look appropriately aligned when seen again from the same point of view. While 91.6% (11/12) of the participants noticed what was

described as a virtual object “drift and sway”, not a single participant realized that the virtual object would align with the physical if they tried to look at it from the same point of view they first used to align it. Some participants even associated this apparent virtual object sway with tracking problems commonly seen in modern OST-HMD, such as HoloLens 2.

#### **4.6.4 Eye-display calibration errors increase user effort, frustration, and sense of failure**

Because participants didn’t immediately notice the apparent object “drift” reported in the previous section, all of them initially assumed they were responsible for not being able to align virtual and physical objects accurately (“*I am messing up on axis but I am not understanding how to fix it*” [P8]) with some participants eventually figuring out that it was out of their control (“*By changing perspective it always changes a bit. I am aware that it is not a 100% precise. I am not confident it will be 100% aligned*” [P5]). As seen before, while their awareness of these problems didn’t help them increase their alignment accuracy, it did, however, impact their perceived Task Load. More specifically: frustration, performance, physical demand, effort, and mental demand (Figure 4.7).

#### **4.6.5 Similarities and discrepancies between VR and AR modalities**

In subsections 4.6.1, 4.6.2, 4.6.3, and 4.6.4, we discussed and compared results between the two simulated OST-HMDs: VR (CALIBRATED) with VR (1MM ERROR). Because both of these displays are idealized OST-HMDs, in this study, we also included a modern OST-HMD (HoloLens 2). While the previous sections comparing results between VR (CALIBRATED) and VR (1MM ERROR) also applies to AR, there are a couple of significant distinctions between AR and simulated AR in general.

## **Alignment error discrepancy between AR and VR**

The AR display modality had significantly worse alignment errors for both alignment interfaces (Fig. 4.5). POINT and HOLOGRAM had mean errors of 2.339 *cm* and 2.774 *cm* for AR compared to 0.751 *cm* and 0.777 *cm* for VR (1MM ERROR). While these values are considerably higher, they are still within the bounds of similar measurements in other experiments conducted in AR [106] and have also been reported before in comparative studies in VR [70] and AR [32].

Moreover, two participants reported seeing “in double” while wearing the AR headset, most likely due to the *vergence-accommodation conflict*. When the human visual system perceives the real target and its virtual representation at the same depth but at different focal planes, it can struggle to tell the difference or similarity in depth. This conflict impairs users’ ability to perform manual tasks [47] and reduces their performance in AR [13].

## **The impact of different interaction interfaces**

One of the most notable differences between *display modalities* was the user interface used to manipulate objects. In *Virtual Augmented Reality* (VR (1MM ERROR) and VR (CALIBRATED)), participants always used a VR controller. In AR, participants used a light, plastic object (see Fig. 4.4) for the POINT, and a gesture interface for the HOLOGRAM interface. While these differences are not a focus of this study, we report them here because they help us explain some differences in the subjective data collected (NASA TLX).

For the POINT alignment interface, where matching an object mid-air was required, users complained that the VR controller was heavier to move around than the plastic object: “*the controller requires a bit more effort than using the hand.*” [P10]. While the weight of the controller didn’t impact either the Alignment Error or mean Alignment Time, it had a statistically significant ( $p < 0.001$ ) impact on the Physical Demand measurement (See Fig. 4.7) for both VR modalities when compared to the AR.

During the study, participants also commented on the differences between the hand-

tracking interface used by HoloLens 2 compared to the Oculus Quest 2 controllers. Hand tracking was seen as unstable (textit“hand detection is shaking” [P9, P11]), not precise [P5, P7], frustrating [P8], and hard to use [P10, P3]. This difficulty in using hand-tracking is directly reflected in the statistically significant ( $p < 0.001$ ) higher amount of time participants took on the HOLOGRAM modality for the AR interface when compared to both VR (1MM ERROR) and VR (CALIBRATED) (Fig. 4.6).

#### **4.6.6 Limitations**

In the simulated environment, the headset’s location and any coordinate system is always known, so it is easy to obtain ground truth values that we can compare with the results obtained in the calibration. Getting ground-truth values in the real world is difficult as it may require a combination of eye trackers, motion capture trackers, and validation studies. Unfortunately, as pointed out before, validation studies are time-consuming calibration per user. In this study, for the AR modality, we obtained ground-truth values by calibrating the OST-HMD (HoloLens) with an external motion capture system through a visual fiducial marker (based on [37]). Unfortunately, due to limitations of the OST-HMD [62], tracking drift could have affected some of our measurements, contributing to the alignment errors measured in this study.

### **4.7 Closing Remarks**

In perception-based manual guidance, users wearing an AR headset follow virtual instructions by, for example, aligning a physical tool to match that of a virtual tool. For example, surgeons use AR plans to decide where to make an incision on the patient [37] by aligning a scalpel with the virtual line. If users perceive virtual objects at a different location than the system’s intended, they are more likely to make a mistake while unaware of it. From the many reasons that could alter users’ perception of virtual objects, we isolate and measure the impact

of eye-display calibration. Eye-display calibration is a critical calibration step [9] required for OST-HMD as it finds the physical location of users' eyes. In this study, we used a simulated OST-HMD in virtual reality [39] to control important variables (e.g., calibration, tracking, optical distortion) we can't control in real life. We also compared our findings from a controlled environment with deployment on a modern, off-the-shelf OST-HMD. Our study results show that even a relatively small eye-display calibration error (1mm) is enough to have a statistically significant impact on alignment accuracy and subjective task-load.

Chapter 4 is, in part, a reprint of the material as it appears in 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) and, in part, being prepared for publication. Danilo Gasques, Weichen Liu, and Nadir Weibel. The dissertation author was the primary investigator and author of both materials.

# Chapter 5

## Towards Contextual Cues

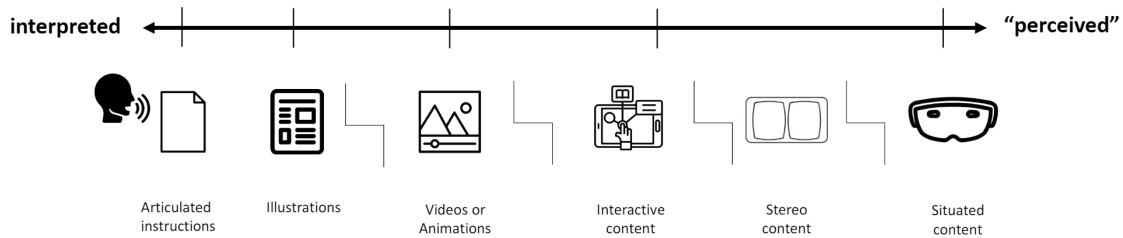
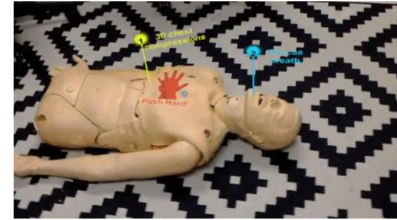
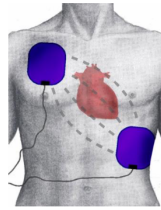
In the course of my Ph.D., I created AR interfaces for automated, collaborative, and medical imaging guidance [103]. From the beginning, my aspirations extended *beyond* the more designing, developing, and evaluating of systems and interfaces. I aimed to retrospectively examine these systems to extract their technical and design foundations. However, this task of retrospective analysis is of such scale that it could warrant an entirely separate Ph.D. and might have been impossible to tackle alone. Nonetheless, this summary should delineate the path I would have followed in my subsequent research, emphasizing the cognitive implications of situated guidance. I hope that this chapter will serve as a guide for future researchers in this field.

### 5.1 The constraints of perception-based Guidance

In perception-based guidance, the cognitive burden of interpreting instructions is transferred to the media itself (Figure 5.1). As previously discussed, Augmented Reality (AR) systems offer significant benefits over alternative media forms.

Instructions are presented in the direct context of the user's physical environment. This contextualization means that users don't have to visualize the final placement of objects; they can physically see it. This feature proves invaluable in tasks requiring the assembly or positioning of

“The first pad goes on the patient's bare upper right chest, just below the collarbone. The second pad goes on the lower left side of the chest, a few inches below the armpit, along the side of the left nipple.”



**Figure 5.1:** From cognitive load to offload: the journey of media technology from interpreted to perception-based guidance.

multiple objects. Moreover, these instructions automatically adapt to the user’s viewpoint, negating the need for users to mentally project how something would appear from their perspective; they can see it directly.

These virtual instructions respond in real-time to even the smallest movements from the user. Consequently, inspecting virtual objects requires no additional effort or action. This real-time response makes navigating through virtual instructions an intuitive and immersive experience. Nonetheless, as we saw in our journey with ARTEMIS and HoloCPR, accurate and reliable guidance through Augmented Reality was only possible after the development of a set of virtual visual aids that helped users in locating and interpreting these instructions.

Some of these visuals were indispensable in overcoming technological constraints, such as the restricted field of view presented by HoloLens. In HoloCPR, for instance, users needed to locate and understand the CPR checklist presented as spatial instructions around the area of action (e.g., near the patient’s chest or mouth). Initial versions of HoloCPR often left users struggling to find any information due to the headset’s limited field of view. To mitigate this, we introduced interactive visual cues that would appear in the user’s visual field and guide them to

move their heads toward the correct location, making the desired information visible within the limited display.

In situations where the visual guidance needed to align accurately with its physical counterpart, the limitations of the headset technology presented a considerable challenge. Despite our rigorous efforts to navigate tracking and registration limitations in ARTEMIS, we observed perceptual errors—even when tracking errors were minimal. These errors forced communication and agreement among novices and experts regarding the placement of the annotations, somehow undermining the use of AR in that domain.

Even in the absence of tracking errors, the fundamental limitations of display technology affect the final registration results. In the previous chapter, this led us down the path of measuring the absolute limits of head-mounted displays. We discovered that certain constraints of these devices would require a fundamental technological leap to overcome these limitations. As demonstrated in the preceding chapter, even minor deviations in eye-HMD calibration can profoundly affect user-task metrics related to manual guidance.

I argue that these fundamental technological improvements need not happen for us to reap the benefits of situated guidance. We found that users could effectively overcome these visual limitations throughout these systems if we introduced specific virtual contextual cues. To better develop this argument, I'll provide an example related to medical imaging.

## **5.2 Easing the cognitive burden of imagining the patient's body**

As suggested in the introduction, one of the primary challenges of contemporary medical imaging lies in its representation. Visualizing a patient's body is a cognitively taxing process, susceptible to errors and oversights [5, 98]. Typically, patient visualizations are portrayed from a perspective, scale, and orientation that significantly differ from the surgeon's viewpoint.

Consequently, surgeons are often tasked with mentally manipulating these images—through interpolation, rotation, translation, or scaling—to align them with their perspective [2].

With the advent of augmented reality displays, this cognitive burden can be substantially offloaded onto the AR system. Even when the registration isn't perfect, situated visualization is beneficial as long as a frame of reference - or contextual cues - is available by design. With that, users have an increased capacity to manage the missing parts if a portion of this mental computation is offloaded to the system. This cognitive offloading can potentially lead to a reduction in user errors.

### **5.3 The case of situated ultrasound**

In an era where precision is paramount, using ultrasound in invasive procedures has become increasingly critical. Required by the FDA in many procedures, ranging from the 5 million central line placements [60] and 1 million breast biopsies [55] conducted annually, this technology is truly a thin yet pivotal cut into the patient.

However, with it comes a unique challenge: “Keeping the needle tip in view as the needle is advanced toward the target is much more difficult” - a difficulty which, unfortunately, is not exclusive to new residents but also extends to seasoned practitioners [100]. The errors stemming from this difficulty can lead to severe complications, including nerve injury, prolonged weakness, infections, and thrombosis.

Ultrasound, while allowing us to peer within the patient, projects this information onto a monitor that is often separate from the patient, requiring clinicians to perform complex mental transformations to accurately map what they see on the screen to the patient's anatomy. Figure 5.2 shows an example.

As argued so many times in this dissertation, an augmented reality display can simplify and streamline this procedure, making it more intuitive and reducing room for error. A situated



**Figure 5.2:** Student (S) performing a central line placement in a phantom. Notice that the student alternates between the site of the procedure (A) and the ultrasound screen (B). S1 and S2 are instructors assessing the student’s procedural skills

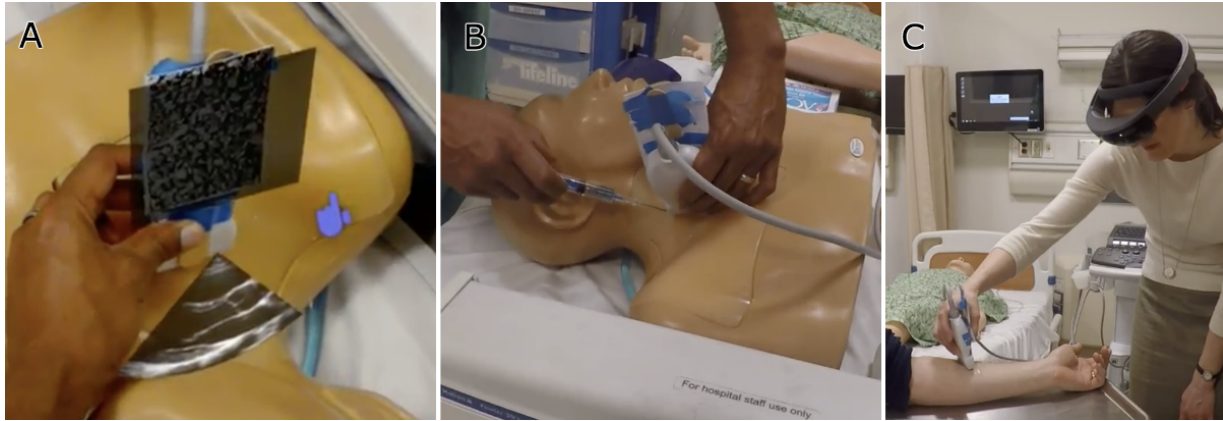
view of ultrasound has the potential to mitigate the mental transformations required to make sense of the ultrasound image in the patient’s body. Figure 5.3 shows a prototype of this idea. Unfortunately, the precision in this domain has little to no leeway for registration problems, allowing less than one or two millimeters for the overall image registration.

In our pilot studies with this prototype, medical students and practitioners championed the situated display because it made ultrasound imaging more intuitive to them; however, whenever they tried to guide a needle under the situated ultrasound image, they would very quickly realize that the overall registration of the Augmented Reality display was off, rendering it unusable.

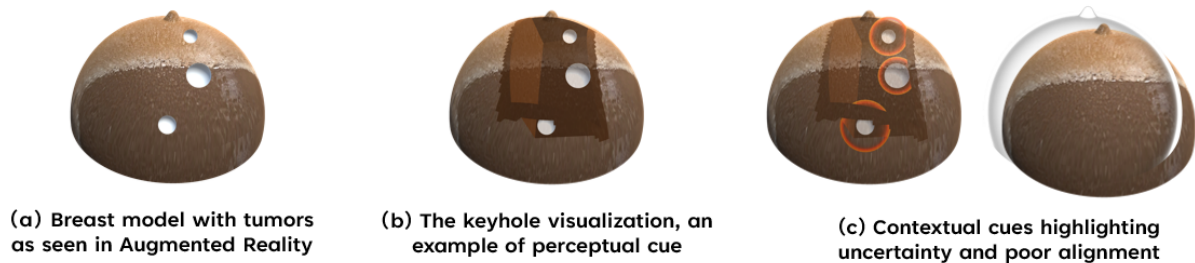
## 5.4 A taxonomy of Augmented Reality cues

Before delving into the answer to the situated ultrasound problem, I must define perceptual, interaction, and contextual cues. While I borrow the definition of perception and interaction cues from the existing literature, I propose a first definition for Contextual Cues here.

*Perceptual cues* [1] are virtual cues that can help users perceive virtual information more



**Figure 5.3:** Situated Ultrasound Prototype (A) View from the headset. (B) Overview of the task - guiding a needle into the patient's neck vein. (C) Student looking at the situated ultrasound



**Figure 5.4:** (a) In most Augmented Reality interfaces, AR visualizations are designed to depict only objects of interest [58, 14, 93]. (b) Over the years, researchers noticed that, by complementing AR visualizations with perceptual cues, users were more likely to understand what they were seeing [12, 15, 1]. (c) In my research, I pose that *contextual cues* can aid users following misaligned AR visualizations.

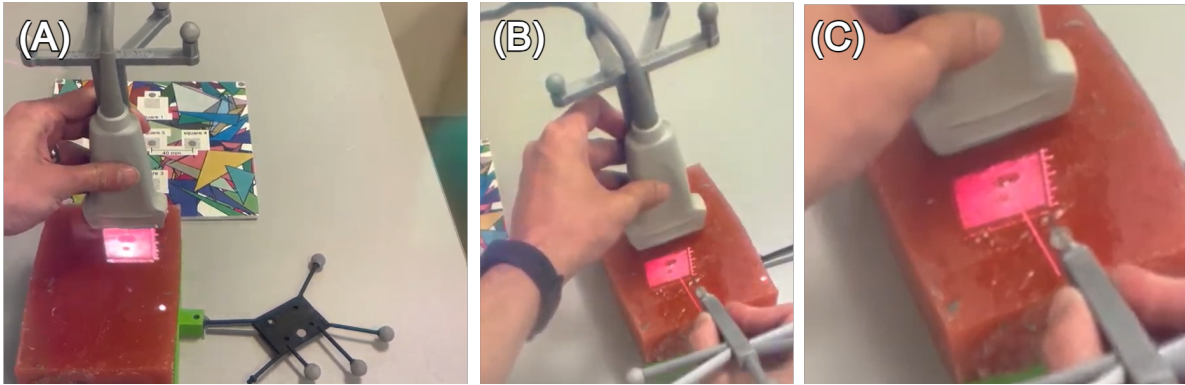
readily by reinforcing depth cues such as shadow, occlusion, motion parallax, and stereopsis. These cues are designed based on our understanding of the human visual system, so they mimic visual cues in the physical world.

*Interaction cues* are virtual cues that help, for example, guide a user to a location in the room, cue the user to look in a specific direction or outline a particular physical object. In HoloCPR, for example, we created a compass that pops in front of the user and reacts to their most minor head movement. The arrow always points in a specific direction, so users understand quickly that they should follow it. I recommend that readers check Dillman et al. [28]’s work for an extensive list of interaction cues for AR.

Here, I define *Contextual cues* as cues that function to situate or contextualize virtual information within its surrounding physical environment. They could reinforce the expected boundaries of the physical environment. For example, in Fig. 5.4 (c), the figure on the right shows a “glass” breast model in Augmented Reality on top of the physical breast phantom. This “glass” visualization contextualizes any other visualization meant to be seen registered to the breast. The case outlined in Fig. 5.4 (c), also shows that the AR interface is slightly misaligned with its physical surroundings. The most crucial aspect of contextual cues is that they hint at the different coordinate systems available and how they relate. These coordinate systems might be the real and the virtual worlds or other tracking systems within the virtual world. The following section discusses a specific example in the domain of ultrasound.

## **5.5 Contextual cues: integrating different coordinate systems**

With the right use of contextual cues, misalignment between different coordinate systems (e.g., real and virtual) is not a limiting factor anymore. In the ultrasound problem, after a thorough prototyping process, we realized that a simple contextual cue was sufficient to help users leverage the benefits of situated visualization even when registration wasn’t perfect. In this case, the most



**Figure 5.5:** An example of contextual cue in the situated ultrasound project. (a) The situated ultrasound prototype shows a live view of ultrasound. It is impossible to tell if the ultrasound and the physical target under it are accurately aligned. (b) and (c) the needle visualization as a contextual aid. (c) zooms in on (b) to show that both the ultrasound and needle display are skewed.

reliable contextual aid was another instrument in the same coordinate system as the ultrasound probe: the syringe. Figure 5.5 shows how a user can very quickly guide a needle into the center of a vessel even when the situated ultrasound is not aligned correctly.

To reiterate, by leveraging the contextual cue of the needle line and using it as a proxy to the actual needle, users can quickly guide the needle to the expected target inside the patient. Moreover, they can still leverage essential aspects of the situated ultrasound, such as its real-world scale and rough position over the patient's anatomy.

## 5.6 Contextual cues: future work

To me, this crude yet powerful example is a leading way in which contextual cues can be designed. However, the current “line” design for the needle might not necessarily be the most intuitive design for alignment or biopsy tasks. An exciting line of future research is parameterizing different types of contextual cues, their designs, and, more importantly, measuring their impact on task performance and cognitive workload.

Finally, in image-guided surgery, surgeons use a combination of images registered to

different coordinate systems. Researching the role of contextual cues in these domains wherein more than two coordinate systems are being integrated with different levels of accuracy is critical to, one day, building the perfect x-ray vision we envisioned for Augmented Reality.

# Bibliography

- [1] Kamyar Abhari. Visual perception and cognition in image-guided intervention. *Electronic Thesis and Dissertation Repository*, 2014.
- [2] Kamyar Abhari, John SH Baxter, Elvis CS Chen, Ali R Khan, Terry M Peters, Sandrine De Ribaupierre, and Roy Eagleson. Training for planning tumour resection: augmented reality and human factors. *IEEE Transactions on Biomedical Engineering*, 62(6):1466–1477, 2014.
- [3] 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.
- [4] Daniel Andersen and Voicu Popescu. Ar interfaces for mid-air 6-dof alignment: Ergonomics-aware design and evaluation. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 289–300. IEEE, 2020.
- [5] Daniel Andersen, Voicu Popescu, Maria Eugenia Cabrera, Aditya Shanghavi, Gerardo Gomez, Sherri Marley, Brian Mullis, and Juan P Wachs. Medical tementoring using an augmented reality transparent display. *Surgery*, 159(6):1646–1653, 2016.
- [6] Christopher M Andrews, Alexander B Henry, Ignacio M Soriano, Michael K Southworth, and Jonathan R Silva. Registration techniques for clinical applications of three-dimensional augmented reality devices. *IEEE journal of translational engineering in health and medicine*, 9:1–14, 2020.
- [7] Rahul Arora, Rubaiat Habib Kazi, Fraser Anderson, Tovi Grossman, Karan Singh, and George W Fitzmaurice. Experimental evaluation of sketching on surfaces in vr. In *CHI*, volume 17, pages 5643–5654, 2017.
- [8] Magnus Axholt, Matthew D Cooper, Martin A Skoglund, Stephen R Ellis, Stephen D O’Connell, and Anders Ynnerman. Parameter estimation variance of the single point active alignment method in optical see-through head mounted display calibration. In *2011 IEEE Virtual Reality Conference*, pages 27–34. IEEE, 2011.

- [9] Magnus Axholt, Martin A Skoglund, Stephen D O’Connell, Matthew D Cooper, Stephen R Ellis, and Anders Ynnerman. Accuracy of eyepoint estimation in optical see-through head-mounted displays using the single point active alignment method. In *IEEE Virtual Reality Conference 2012, Orange County (CA), USA*, 2011.
- [10] Ronald Azuma, Yohan Baillet, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE computer graphics and applications*, 21(6):34–47, December 2001.
- [11] Ronald T Azuma. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments*, 6(4):355–385, 1997.
- [12] Michael Bajura, Henry Fuchs, and Ryutarou Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *ACM SIGGRAPH Computer Graphics*, 26(2):203–210, 1992.
- [13] Anil Ufuk Batmaz, Mayra Donaji Barrera Machuca, Junwei Sun, and Wolfgang Stuerzlinger. The effect of the vergence-accommodation conflict on virtual hand pointing in immersive displays. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI ’22, New York, NY, USA, 2022. Association for Computing Machinery.
- [14] Sylvain Bernhardt, Stéphane A Nicolau, Luc Soler, and Christophe Doignon. The status of augmented reality in laparoscopic surgery as of 2016. *Medical image analysis*, 37:66–90, 2017.
- [15] Christoph Bichlmeier, Felix Wimmer, Sandro Michael Heining, and Nassir Navab. Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality. In *2007 6th IEEE and ACM international symposium on mixed and augmented reality*, pages 129–138. IEEE, 2007.
- [16] Mark Billinghurst, Adrian Clark, Gun Lee, et al. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2-3):73–272, 2015.
- [17] Mark Billinghurst and Hirokazu Kato. Collaborative augmented reality. *Communications of the ACM*, 45(7):64–70, 2002.
- [18] Oliver Bimber and Ramesh Raskar. *Spatial augmented reality: merging real and virtual worlds*. CRC press, 2005.
- [19] Ellenor J Brown, Kyoko Fujimoto, Bennett Blumenkopf, Andrea S Kim, Kimberly L Kontson, and Heather L Benz. Usability assessments for augmented reality head-mounted displays in open surgery and interventional procedures: A systematic review. *Multimodal Technologies and Interaction*, 7(5):49, 2023.

- [20] Josef Buchner, Katja Buntins, and Michael Kerres. The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning*, 38(1):285–303, 2022.
- [21] Andrius Budrionis, Knut Magne Augestad, Hiten RH Patel, and Johan Gustav Bellika. An evaluation framework for defining the contributions of telestration in surgical telementoring. *Interactive journal of medical research*, 2(2):e14, 2013.
- [22] Frank K Butler. Tactical combat casualty care: update 2009. *Journal of Trauma and Acute Care Surgery*, 69(1):S10–S13, 2010.
- [23] Frank K Butler Jr, John Hagmann, and E George Butler. Tactical combat casualty care in special operations. *Military medicine*, 161(suppl\_1):3–16, 1996.
- [24] T P Caudell and D W Mizell. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, volume ii, pages 659–669 vol.2, Kauai, HI, USA, January 1992. IEEE.
- [25] Sebastiano Chiodini, Marco Pertile, Riccardo Giubilato, Federico Salvioli, Marco Barrera, Paola Franceschetti, and Stefano Debei. Camera rig extrinsic calibration using a motion capture system. In *2018 5th IEEE International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, pages 590–595. IEEE, 2018.
- [26] Benjamin F Crabtree and William L Miller. *Doing qualitative research*. Sage Publications, 1999.
- [27] 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.
- [28] Kody R Dillman, Terrance Tin Hoi Mok, Anthony Tang, Lora Oehlberg, and Alex Mitchell. A visual interaction cue framework from video game environments for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.
- [29] Charles R Doarn and Rifat Latifi. Telementoring and teleproctoring in trauma and emergency care. *Current Trauma Reports*, 2(3):138–143, 2016.
- [30] Dylan Fafard, Ian Stavness, Martin Dechant, Regan Mandryk, Qian Zhou, and Sidney Fels. Ftvr in vr: Evaluation of 3d perception with a simulated volumetric fish-tank virtual reality display. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.
- [31] Steven Feiner, Blair Macintyre, and Dorée Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, 36(7):53–62, July 1993.

- [32] Marc Fischer, Christoph Leuze, Stephanie Perkins, Jarrett Rosenberg, Bruce Daniel, and Alejandro Martin-Gomez. Evaluation of different visualization techniques for perception-based alignment in medical ar. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 45–50. IEEE, 2020.
- [33] James D Foley, Foley Dan Van, Andries Van Dam, Steven K Feiner, John F Hughes, Edward Angel, and J Hughes. *Computer graphics: principles and practice*, volume 12110. Addison-Wesley Professional, 1996.
- [34] Javad Fotouhi, Tianyu Song, Arian Mehrfard, Giacomo Taylor, Qiaochu Wang, Fengfan Xian, Alejandro Martin-Gomez, Bernhard Fuerst, Mehran Armand, Mathias Unberath, et al. Reflective-ar display: An interaction methodology for virtual-to-real alignment in medical robotics. *IEEE Robotics and Automation Letters*, 5(2):2722–2729, 2020.
- [35] Adam Fouse, Nadir Weibel, Edwin Hutchins, and James D Hollan. Chronoviz: a system for supporting navigation of time-coded data. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 299–304. ACM, 2011.
- [36] 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.
- [37] Danilo Gasques, Janet Johnson, Thomas Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun 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)*, CHI '21, page 1–14, New York, NY, USA, 2021. Association for Computing Machinery.
- [38] Danilo Gasques, Janet Johnson, Thomas Sharkey, and Nadir Weibel. What You Sketch Is What You Get: Quick and Easy Augmented Reality Prototyping with PintAR. In *Proc. CHI 2019*, CHI EA '19, pages LBW1416:1–LBW1416:6, New York, NY, USA, 2019. ACM. event-place: Glasgow, Scotland Uk.
- [39] Danilo Gasques, Weichen Liu, and Nadir Weibel. The virtual-augmented reality simulator: Evaluating ost-hmd ar calibration algorithms in vr. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 702–703. IEEE, 2022.
- [40] Danilo Gasques Rodrigues, Ankur Jain, Steven R Rick, Liu Shanglely, Preetham Suresh, and Nadir Weibel. Exploring mixed reality in specialized surgical environments. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2591–2598, 2017.
- [41] 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.
- [42] Atul Gawande. *Checklist manifesto, the (HB)*. Penguin Books India, 2010.
- [43] Y. Genc, M. Tuceryan, and N. Navab. Practical solutions for calibration of optical see-through devices. In *Proceedings. International Symposium on Mixed and Augmented Reality*, pages 169–175, 2002.
- [44] Robert Graham, Margaret A McCoy, Andrea M Schultz, et al. The public experience with cardiac arrest. *National Academies Press (US)*, 2015.
- [45] Jens Grubert, Yuta Itoh, Kenneth Moser, and J Edward Swan. A survey of calibration methods for optical see-through head-mounted displays. *IEEE transactions on visualization and computer graphics*, 24(9):2649–2662, 2017.
- [46] 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.
- [47] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, 8(3):33–33, 2008.
- [48] HTC. Vive pro starter kit — the professional-grade vr headset. <https://www.vive.com/us/product/vive-pro-starter-kit/>, Jan 2020.
- [49] Intel. Intel realsense depth camera d435, May 2020.
- [50] Tomoki Itamiya, Toshinori Iwai, and Tsuyoshi Kaneko. The holographic human for surgical navigation using microsoft hololens. *EPiC Series in Engineering*, 1:26–30, 2018.
- [51] Yuta Itoh and Gudrun Klinker. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *2014 IEEE symposium on 3d user interfaces (3dUI)*, pages 75–82. IEEE, 2014.
- [52] Changwon Jang, Kiseung Bang, Seokil Moon, Jonghyun Kim, Seungjae Lee, and Byoung-ho Lee. Retinal 3d: augmented reality near-eye display via pupil-tracked light field projection on retina. *ACM Transactions on Graphics (TOG)*, 36(6):1–13, 2017.
- [53] 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, pages 67–76, New York, NY, USA, 2018. ACM, ACM. event-place: New York, NY, USA.

- [54] 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*, pages 1–20, 2023.
- [55] Inha Jung, Kyunghwa Han, Min Jung Kim, Hee Jung Moon, Jung Hyun Yoon, Vivian Youngjean Park, and Eun-Kyung Kim. Annual trends in ultrasonography-guided 14-gauge core needle biopsy for breast lesions. *Korean journal of radiology*, 21(3):259–267, 2020.
- [56] Hirokazu Kato and Mark Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, pages 85–94. IEEE, 1999.
- [57] Marta Kersten-Oertel, Pierre Jannin, and D Louis Collins. Dvv: a taxonomy for mixed reality visualization in image guided surgery. *IEEE transactions on visualization and computer graphics*, 18(2):332–352, 2011.
- [58] Marta Kersten-Oertel, Pierre Jannin, and D Louis Collins. The state of the art of visualization in mixed reality image guided surgery. *Computerized Medical Imaging and Graphics*, 37(2):98–112, 2013.
- [59] Seungwon Kim, Gun Lee, Mark Billinghurst, and Weidong Huang. The combination of visual communication cues in mixed reality remote collaboration. *Journal on Multimodal User Interfaces*, 14(4):321–335, 2020.
- [60] Craig Kornbau, Kathryn C Lee, Gwendolyn D Hughes, and Michael S Firstenberg. Central line complications. *International journal of critical illness and injury science*, 5(3):170, 2015.
- [61] Gregory Kramida. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE transactions on visualization and computer graphics*, 22(7):1912–1931, 2015.
- [62] Bernard C. Kress and William J. Cummings. Invited paper: Towards the ultimate mixed reality experience: Hololens display architecture choices. *SID Symposium Digest of Technical Papers*, 48(1):127–131, 2017.
- [63] 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, 2018.
- [64] 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 tementoring. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 3–8. IEEE, 2018.

- [65] Yang Liu, Haiwei Dong, Longyu Zhang, and Abdulmotaleb El Saddik. Technical evaluation of hololens for multimedia: a first look. *IEEE MultiMedia*, 25(4):8–18, 2018.
- [66] Patrick Maier, Arindam Dey, Christian AL Waechter, Christian Sandor, Marcus Tönnis, and Gudrun Klinker. An empiric evaluation of confirmation methods for optical see-through head-mounted display calibration. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 267–268. IEEE, 2011.
- [67] Martin A Makary and Michael Daniel. Medical error—the third leading cause of death in the us. *Bmj*, 353, 2016.
- [68] JA Martin, Glenn Regehr, Richard Reznick, Helen Macrae, John Murnaghan, Carol Hutchison, and M Brown. Objective structured assessment of technical skill (osats) for surgical residents. *British journal of surgery*, 84(2):273–278, 1997.
- [69] Alejandro Martin-Gomez, Ulrich Eck, Javad Fotouhi, and Nassir Navab. Looking also from another perspective: Exploring the benefits of alternative views for alignment tasks. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 811–812. IEEE, 2020.
- [70] Alejandro Martin-Gomez, Ulrich Eck, and Nassir Navab. Visualization techniques for precise alignment in vr: A comparative study. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 735–741. IEEE, 2019.
- [71] Erin McGarrity, Yakup Genc, Mihran Tuceryan, Charles Owen, and Nassir Navab. A new system for online quantitative evaluation of optical see-through augmentation. In *Proceedings IEEE and ACM International Symposium on Augmented Reality*, pages 157–166. IEEE, 2001.
- [72] Helena M Mentis. Collocated use of imaging systems in coordinated surgical practice. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–17, 2017.
- [73] Microsoft. Azure kinect - fast point cloud, Apr 2020.
- [74] Microsoft. Azure kinect dk hardware specifications, Feb 2020.
- [75] Microsoft. Hololens (1st gen) hardware. <https://docs.microsoft.com/en-us/hololens/hololens1-hardware>, Apr 2020.
- [76] Microsoft. Hololens 2 hardware. <https://docs.microsoft.com/en-us/hololens/hololens2-hardware>, Apr 2020.
- [77] Microsoft. "improve visual quality and comfort — microsoft docs". <https://docs.microsoft.com/en-us/hololens/hololens-calibration>, Aug 2020.
- [78] Microsoft. Mixedreality-webtrc, Apr 2020.
- [79] Microsoft. Overview of dynamics 365 remote assist on hololens and hololens 2, Apr 2020.

- [80] Kenneth R Moser, Magnus Axholt, and J Edward Swan. Baseline spaam calibration accuracy and precision in the absence of human postural sway error. In *2014 IEEE Virtual Reality (VR)*, pages 99–100. IEEE, 2014.
- [81] Rafael Munoz-Salinas. Aruco: a minimal library for augmented reality applications based on opencv. *Universidad de Córdoba*, 2012.
- [82] Sundar Murugappan, Danilo Gasques Rodrigues, Govinda Payyavula, and Simon Dimaio. Method and system for facilitating remote presentation or interaction, March 23 2023. US Patent App. 17/800,395.
- [83] Mahdi Nabyouni, Siroberto Scerbo, Doug A Bowman, and Tobias Höllerer. Relative effects of real-world and virtual-world latency on an augmented reality training task: an ar simulation experiment. *Frontiers in ICT*, 3:34, 2017.
- [84] 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.
- [85] Jerry Nolan, Jasmeet Soar, and Harald Eikeland. The chain of survival. *Resuscitation*, 71(3):270–271, 2006.
- [86] Kenton O’Hara, Gerardo Gonzalez, Abigail Sellen, Graeme Penney, Andreas Varnavas, Helena Mentis, Antonio Criminisi, Robert Corish, Mark Rouncefield, Neville Dastur, et al. Touchless interaction in surgery. *Communications of the ACM*, 57(1):70–77, 2014.
- [87] OptiTrack. Motive - optical motion capture software. <https://optitrack.com/products/motive/>, Apr 2020.
- [88] OptiTrack. Optitrack - active components, Apr 2020.
- [89] OptiTrack. Optitrack - hardware. <https://optitrack.com/hardware/>, Apr 2020.
- [90] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, Mingsong Dou, et al. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 741–754, 2016.
- [91] Nitish Padmanaban, Robert Konrad, Tal Stramer, Emily A Cooper, and Gordon Wetzstein. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proceedings of the National Academy of Sciences*, 114(9):2183–2188, 2017.
- [92] Long Qian, Ehsan Azimi, Peter Kazanzides, and Nassir Navab. Comprehensive tracker based display calibration for holographic optical see-through head-mounted display. *arXiv preprint arXiv:1703.05834*, 2017.

- [93] Long Qian, Jie Ying Wu, Simon P DiMaio, Nassir Navab, and Peter Kazanzides. A review of augmented reality in robotic-assisted surgery. *IEEE Transactions on Medical Robotics and Bionics*, 2(1):1–16, 2019.
- [94] Rafael Radkowski and Sravya Kanunganti. Augmented reality system calibration for assembly support with the microsoft hololens. In *ASME 2018 13th International Manufacturing Science and Engineering Conference*. American Society of Mechanical Engineers Digital Collection, 2018.
- [95] Eric Ragan, Curtis Wilkes, Doug A Bowman, and Tobias Hollerer. Simulation of augmented reality systems in purely virtual environments. In *2009 IEEE Virtual Reality Conference*, pages 287–288. IEEE, 2009.
- [96] Danilo Gasques Rodrigues, Janet Johnson, and Nadir Weibel. Real-time guidance for cardiopulmonary resuscitation in mixed reality. In *12th EAI International Conference on Pervasive Computing Technologies for Healthcare—Demos, Posters, Doctoral Colloquium*, 2018.
- [97] 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*, 167(4):724–731, 2020.
- [98] 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.
- [99] Jannick P Rolland, Richard L Holloway, and Henry Fuchs. Comparison of optical and video see-through, head-mounted displays. In *Telem manipulator and Telepresence Technologies*, volume 2351, pages 293–307. SPIE, 1995.
- [100] Brian D Sites, Brian C Spence, John D Gallagher, Christopher W Wiley, Marc L Bertrand, and George T Blike. Characterizing novice behavior associated with learning ultrasound-guided peripheral regional anesthesia. *Regional Anesthesia & Pain Medicine*, 32(2):107–115, 2007.
- [101] Dave Snowdon, Elizabeth F Churchill, and Alan J Munro. Collaborative virtual environments: Digital spaces and places for cscw: An introduction. In *Collaborative virtual environments*, pages 3–17. Springer, 2001.
- [102] Maurício Sousa, Rafael Kufner dos Anjos, Daniel Mendes, Mark Billingham, and Joaquim Jorge. Warming deixis: Distorting gestures to enhance collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.
- [103] Preetham Suresh, Danilo Gasques Rodrigues, Nadir Weibel, and Elizabeth A Anderson. Enhanced ultrasound systems and methods, June 18 2020. US Patent App. 16/643,505.

- [104] 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.
- [105] J Edward Swan and Joseph L Gabbard. Survey of user-based experimentation in augmented reality. In *Proceedings of 1st International Conference on Virtual Reality*, volume 22, pages 1–9. Citeseer, 2005.
- [106] Arthur Tang, Ji Zhou, and Charles Owen. Evaluation of calibration procedures for optical see-through head-mounted displays. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 161, 2003.
- [107] Theophilus Teo, Louise Lawrence, Gun A Lee, Mark Billingham, 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.
- [108] 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.
- [109] Jenifer Tidwell. *Designing interfaces: Patterns for effective interaction design*. ” O’Reilly Media, Inc.”, 2010.
- [110] Mihran Tuceryan, Yakup Genc, and Nassir Navab. Single-point active alignment method (spaam) for optical see-through hmd calibration for augmented reality. *Presence: Teleoperators & Virtual Environments*, 11(3):259–276, 2002.
- [111] Ultraleap. Tracking — ultraleap, Apr 2020.
- [112] Unity. Unity real-time development platform — 3d, 2d vr ar visualizations. <https://unity.com/>, Jan 2020.
- [113] Reid Vassallo, Adam Rankin, Elvis CS Chen, and Terry M Peters. Hologram stability evaluation for microsoft hololens. In *Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment*, volume 10136, page 1013614. International Society for Optics and Photonics, 2017.
- [114] Senso VR. Senso vr — interactive virtual and augmented reality, Apr 2020.
- [115] Vuforia. Vuforia developer portal. <https://www.vuforia.com>, Apr 2020.
- [116] 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.
- [117] Michelle Yeh. *Attention and trust biases in the design of augmented reality displays*. University of Illinois at Urbana-Champaign, 2000.